Joint Beamforming Optimization in RIS-Aided MIMO Wireless Systems Under Mutual Coupling and Multiple Reflections

by

Dilki Wijekoon

A Thesis submitted to The Faculty of Graduate Studies of

The University of Manitoba

in partial fulfillment of the requirements for the degree of

Master of Science

Department of Electrical and Computer Engineering

University of Manitoba

Winnipeg

April 2024

It is during our darkest moments that we must focus to see the light.

Aristotle

Abstract

In a Reconfigurable Intelligent Surface (RIS), Mutual Coupling (MC) is unavoidable as a result of the sub-wavelength structure of its unit elements. In addition, MC inherently leads to multiple reflection effects that are ignored in conventional (approximative) RIS models. In this thesis, we examine the problem of joint active and passive beamforming in a controllable multi-user reconfigurable intelligent surface (RIS)-assisted downlink and uplink wireless communication system, considering the MC among RIS elements. We formulate a problem to optimize active and passive beamforming jointly under the MMSE criterion. Given the non-convex nature of the problem, we use alternating optimization to decompose the problem into two sub-problems for downlink and uplink transmissions separately. In both transmissions, one sub-problem involves optimizing the phase-shift matrix of RIS. In downlink, the other sub-problem is the optimization of active precoding for the base station (BS), while the equivalent sub-problem in uplink is the optimization of the linear receiver matrix. We optimize the phase shift matrix under a physically-consistent model using the gradient descent algorithm for both transmissions. We use the Lagrange multiplier method to optimize active precoding in the downlink and apply the First Order Necessary Condition (FONC) to optimize the linear receiver in the uplink. Simulation results are represented for both lossless and lossy RIS scenarios under perfect and imperfect channel state information. We discuss the impact of changing the number of RIS elements and the RIS element spacing on system performance. The results show that, with optimized phase shifts and active precoding, the inherent multiple reflection effect can improve the performance of RIS-aided wireless communications systems.

In our previous analysis, we focused on optimizing joint beamforming within a physically consistent environment by incorporating MC effects. However, we did not ex-

plicitly address the optimization of MC itself. This represents a critical aspect that warrants further investigation and consideration for a comprehensive and refined analysis. As mentioned earlier, MC emerges as an inherent feature in RIS particularly with sub-wavelength inter-element spacing. Alongside, the presence of electromagnetic (EM) transmit/receive radiation patterns is also inevitable. The simultaneous presence of these two factors naturally leads to the emergence of non-local RIS structures, which can be effectively described via non-diagonal phase shift matrices. As the second analysis, we focus on optimizing MC and radiation patterns in RIS-assisted multi-user MIMO wireless communication systems. We particularly formulate a novel problem to jointly optimize active and passive beamforming as well as MC and radiation patterns in a physically consistent manner considering reflective and transmissive RIS setups separately. To characterize the physically consistent model, we deploy scattering parameters and propose a novel approach to optimize MC and radiation patterns through an offline optimization method, rather than optimizing on the fly, using two distinct solution approaches for reflective and transmissive models. We present simulation results using both parametric and geometric channel modeling approaches. Our numerical results showcase that the system performance increases with the proposed MC and radiation patterns optimization, and this improvement is achievable without the need for optimizing them on the fly, which can be rather cumbersome.

Keywords: Reconfigurable Intelligent Surface (RIS), joint active and passive beamforming, mutual coupling, multiple reflection effects, physically-consistent model, radiation patterns, non-local RIS structures, non-diagonal phase shift matrix, reflective RIS, transmissive RIS.

Acknowledgement

First, I would like to express my deepest gratitude to my supervisor, Professor Ekram Hossain, for his valuable guidance and mentorship throughout the program. I am deeply appreciative of the opportunity to learn and grow under his expert supervision.

I am immensely grateful to Professor Amine Mezghani for his exceptional insights and valuable contributions throughout the program. I am fortunate to work under his guidance, and I genuinely appreciate the knowledge and personal growth I have attained during this period.

Furthermore, I wish to extend my sincere gratitude to the committee members, Professor Pradeepa Yahampath and Professor Amine Mezghani, for their valuable feedback that has greatly contributed to the improvement of this work.

Finally, I wish to extend my heartfelt thanks to my parents for their unwavering belief in me and their constant encouragement throughout this journey, through all the ups and downs. Additionally, I express my gratitude to my friends for their support and companionship along the way.

Table of Contents

List of Figures

List of Tables

List of Abbreviations

RIS	Reconfigurable Intelligent Surface
MIMO	Multiple Input Multiple Output
mMIMO	Massive Multiple Input Multiple Output
mmWave	Millimeter-Wave
MMSE	Minimum Mean Square Error
LMMSE	Linear Minimum Mean Square Error
SNR	Signal-to-noise ratio
SINR	Signal-to-interference-plus-noise ratio
EM	Electromagnetic
BS	Base Station
FONC	First Order Necessary Condition
MC	Mutual Coupling
MISO	Multiple Input Single Output
CSI	Channel State Information
AO	Alternating Optimization

Chapter 1

Introduction

1.1 Overview

5G-advanced wireless technologies are now being developed to cope with the steadily growing demand for wireless network capacity. One technology that fulfills the capacity requirement is massive multiple-input multiple-output (mMIMO), an advanced version of conventional MIMO [1–3]. Massive MIMO utilizes antenna arrays that contain large-scale antennas to serve multiple users through the same time and frequency bands. Due to the large-scale structure of mMIMO, its implementation incurs high hardware costs [4]. In addition, managing a large-scale array of antennas and processing a massive amount of spatially multiplexed data constitute significant challenges in signal processing and resource allocation [4]. Millimeter wave (mmWave), which provides a large amount of spectrum [5–7], is another emerging technology that satisfies the capacity requirement. However, mmWave only covers a small area as a result of its short wavelength, and the path between the BS and users can be easily blocked by obstacles [5,8]. Furthermore, mmWave signals incur greater propagation losses when contrasted with lower frequency bands.

1.1.1 Reconfigurable Intelligent Surfaces (RIS) and Beamforming

One possible solution to address the above-mentioned problems associated with hardware cost and coverage is using Reconfigurable Intelligent Surface (RIS) [9–16]. RIS is a two-dimensional meta-surface consisting of a large number of passive reflective components. This passive or quasi-passive element structure of RIS leads to low manufacturing and implementation costs. Each of these reflective elements is capable of independently adjusting both the amplitude and phase angles of the incident signal through a controller [17]. This property is known as "passive beamforming" of RIS. Through passive beamforming optimization, it is possible to obtain a high signal-to-interference-plus-noise ratio (SINR) at the receivers. In general, when the number of reflective elements increases, the achievable SINR also increases [9]. Also, active beamforming, employed by the Base Station (BS), is a technique that dynamically adjusts the phase and amplitude of signals emitted from BS antennas. This approach allows the BS to focus transmitted energy towards specific users or directions, optimizing SINR and quality while minimizing interference. Therefore in RIS-aided MIMO system, simultaneously optimizing passive beamforming of RIS and active beamforming of BS leads to an even greater improvement in the SINR of the entire system by minimizing interference from unintentional users [9].

1.1.2 Reflective RIS and Transmissive RIS Configurations

Based on the energy split for reflection and transmission, RIS can be categorized as reflective RIS and transmissive RIS. In a reflective RIS configuration, both the BS and users are positioned on the same side of the RIS, known as the reflection space. Many studies on RIS have primarily focused on reflective RIS configurations [18–21]. On the other hand, some research studies have explored transmissive RIS configurations [22–27].

In this scenario, the BS and users are positioned on opposite sides of the RIS, enabling the RIS to serve users located in the transmission space.

1.1.3 Mutual Coupling and Multiple Reflections

Since the RIS is an antenna array terminated by loads, it involves scattering characteristics and has coupling effects among the elements. The tightly spaced arrangement of passive array antenna elements in RIS can lead to one reflective element absorbing signals emitted by another reflective element. This effect is called *mutual coupling* (MC) [18, 28–34]. In other words, MC denotes the Electromagnetic (EM) interaction among the reflective elements of the RIS and is one of the reasons for the non-linearity of RIS configuration with respect to tunable phase shifts [35]. When RIS elements are placed in close vicinity to each other, especially with sub-wavelength spacing, MC becomes strong and influences the EM behavior of each element. As shown in figure 1.1 each individual radiating RIS element makes some impact on the radiations of neighboring elements by interacting EM fields. Typically, the strength of the MC depends on three fundamental characteristics: *i)* scattering cross-section; *ii)* number of reflective elements; and *iii)* RIS element spacing [35]. In other words, the strength of the MC can be controlled by adjusting these characteristics. Optimizing the inter-element spacing is one approach to control the coupling. Also, large scattering cross-sections increase the MC. Thus, reducing the scattering cross-section aids in reducing the coupling. Similarly, reducing the number of RIS elements leads to a decreased MC [35]. However, the significance of using RIS diminishes when these latter two properties are reduced. Hence, it is obvious that MC effects are unavoidable when dealing with RIS-assisted communications systems. The study [36] shows a detailed derivation and mathematical interpretation of scattering and MC.

MC leads to changing the impedance of the loads (phase shifters) connected to RIS.

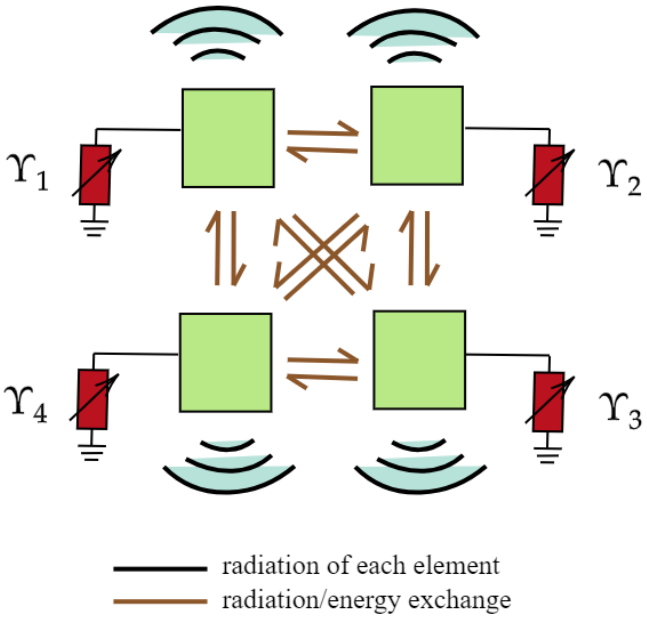


Figure 1.1: Interpretation of mutual coupling.

As the impedance of the loads changes due to MC, there will be a complex interaction between the terminated RIS loads and reflective elements. Typically, in an antenna array setup, the loads and elements are perfectly matched. Ideally, when a signal is transmitted from such an antenna array, the total signal will be absorbed or radiated by the load connected to each antenna element. However, MC leads to imperfect matching between loads and RIS elements. As a result, a part of the transmitted signal reflects back toward the RIS, and this process repeats. This is called the *multiple reflection effect*. Fig. 1.2 shows the effect of multiple reflections on RIS elements. In this instance, the RIS and the phase shifters function as two separate surfaces on which signals are reflected back and forth. Most of the previous research on RIS did not consider the role of multiple reflections, which arise as a result of MC.

Several works consider the MC impact and its consequences on RIS-assisted wireless systems [18,29–31,37–39], with the majority of them using the impedance Z-parameters

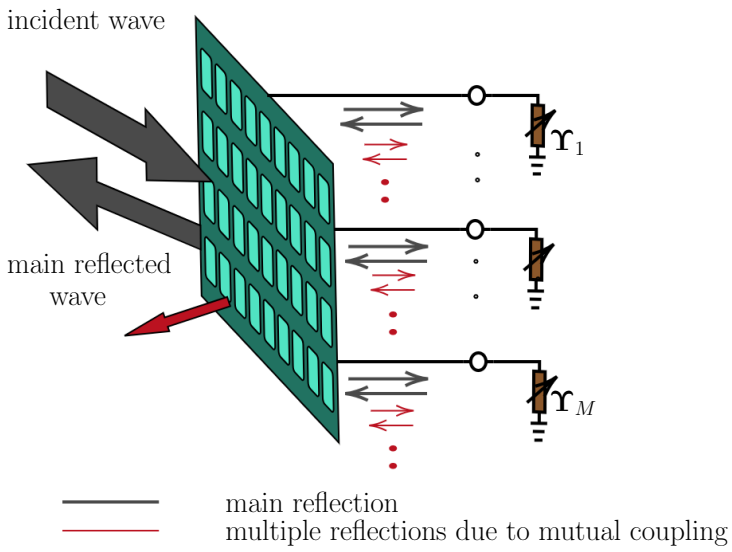


Figure 1.2: Multiple reflection effect in RIS.

for its characterization [29,37,38]. However, most of these studies rely on the minimum scattering assumption (element size $\ll$ element spacing) to find the impedance matrix, which is not a practical relevant case. On the other hand, the works [31, 39] focus on the scattering S-parameters to characterize MC. Z-parameters are related to voltages and currents while the S-parameters refer to the power ratio of incident and reflected/transmitted waves [40]. However, these two parameters have a connection with each other [40]. Irrespective of the modeling approach, all these works confirm that, when MC is present and taken under consideration in the RIS optimization, the system performance improves.

1.1.4 S-parameters and Transmit/Receive Radiation Patterns

In the context of scattering, S-parameters quantify how much of an EM incident wave is reflected, transmitted, or absorbed by the surface [36]. These parameters help to understand how the RIS scatters electromagnetic energy, influencing its effectiveness

in various applications like beamforming and interference mitigation. Hence, analyzing the S-parameters allows to determine the EM characteristics of the RIS. These EM characteristics involve MC, multiple reflection and transmit/receive radiation patterns. Transmit radiation patterns denote how much of the EM wave passes through the RIS without being scattered or absorbed. On the other hand, receive radiation patterns denote how the RIS detects scattered EM waves from different angles. Hence, transmit and receive radiation patterns also play a crucial role in RIS modeling and configuration, alongside MC and multiple reflections [36].

1.1.5 Non-local and Local RIS Structures

The phase shift matrix of RIS represents the phase shifts applied by each reflecting element across the RIS surface. Typically, this matrix is represented as a complex matrix, where each matrix element corresponds to the phase shift applied by a particular reflective element. In the context of the RIS phase shift matrix, most of the conventional models assume a diagonal phase shift matrix, which is commonly referred to as having a "local-RIS" configuration. Conversely, when the phase shift matrix is non-diagonal, it denotes a "non-local" RIS configuration. The majority of the conventional models rely on "local" RIS structures [19–21]. The recent studies [41, 42] present RIS models employing non-diagonal phase shift matrices, also known as non-local RIS structures [36], which outperform conventional RIS models. In [41], the authors coined the term "beyond diagonal" phase shift matrix and explained different RIS configurations, where direct physical connections are established among the RIS elements in order to create non-diagonal matrix structures. This study [41] investigated RIS models corresponding to fully-connected, group-connected architectures which result in non-diagonal-full matrix, and block diagonal matrix structures, respectively. By considering these RIS configurations, the study [41] revealed that the RIS models with block diagonal and

non-diagonal matrix structures outperform those with diagonal matrix structures (or local RIS) in terms of spectral efficiency. In [42], the authors investigated reducing the hardware complexity caused by full and block diagonal structures while achieving a better gain as compared to the earlier achieved results for non-diagonal cases. To obtain the corresponding non-diagonal matrix, the typical diagonal matrix was permuted with two row/column permutation matrices. Nonetheless, the proposed "non-local" RIS architectures are complex to implement due to the requirement of physical connections among RIS elements. However, these structures require extra hardware with physical connections among RIS elements, resulting in increased wiring complexity and control overhead. Fortunately, non-local RIS structures can also be achieved naturally with minimal overhead, using an accurate physically (electromagnetically) consistent model, which encounters MC between elements, multiple reflections, and transmit/receive radiation patterns.

1.2 Motivation

The majority of research on jointly optimizing active and passive beamforming in RIS-assisted wireless communications are based on multiple-input single-output (MISO) [9, 43–45] models. More recent works have considered joint beamforming based on MIMO systems [46–48]. Some studies also explore the use of multiple RIS in joint beamforming [49, 50]. On the other hand, the majority of beamforming optimization was conducted based on downlink transmission, and very few research works considered uplink transmission [51–53]. Nevertheless, all of these works assumed conventional (i.e. approximative) RIS models without considering the effect of multiple reflections. Although there are a few studies that consider MC [29, 37, 38] with minimum scattering assumptions, none of these explicitly address the optimization of MC. Therefore, inspired by the aforementioned considerations, we jointly optimize active and passive

beamforming in a RIS-aided MIMO system, incorporating mutual coupling and its consequences, as well as the EM radiation patterns. This approach not only enhances spectral efficiency but also unlocks new frontiers in wireless communication paradigms.

1.3 Related Work and Contributions

The contribution section is divided into two parts.

- First, we optimize joint beamforming in RIS-assisted MIMO systems under multiple reflections using a physically consistent RIS model. However, it is important to note that we are not optimizing multiple reflections during this process. Instead, we maintain it fixed throughout the optimization procedure.
- Then, we optimize joint beamforming under the same physically consistent model while also optimizing multiple reflections and transmit/receive radiation patterns. This approach leads to the creation of an optimized non-local RIS structure.

1.3.1 Joint Beamforming Optimization in RIS-Aided MIMO Systems Under Multiple Reflections

We propose a method to jointly optimize active and passive beamforming for both downlink and uplink transmissions in a RIS-aided wireless system, considering a physically-consistent RIS model instead of a conventional model. In this way, we take into account the effect of MC among the RIS elements. Instead of impedance matrices studied in [37, 38], we employ a more general model based on scattering matrices, which are directly related to the radiation patterns and also easy to measure. We present results that are valid for a general model of a (well-designed) RIS with a given element pattern but with no particular assumption on the impedance matrix. In other words, our

work offers a more general framework by leveraging the fundamental laws of physics. Specifically, the model is a direct consequence of the power conservation principle. We utilize a general approach that establishes connections between radiation patterns and scattering parameters, all grounded in the most fundamental law of physics. Hence, this model can be applied to any type of array antenna, not only to RIS. In summary,

- We formulate joint optimization problems for active and passive beamforming separately for downlink and uplink transmissions with the objective of enhancing the spectral efficiency performance of the system [46] based on a more accurate RIS model.

- For both downlink and uplink transmissions, the corresponding multi-variable, non-convex, and constrained optimization problem is solved based on the alternate optimization (AO) technique [54] by dividing the problem into two sub-problems. For downlink transmission, the sub-problems are referred to as active precoding optimization of BS and reflection coefficient optimization of RIS, while for uplink transmission, the corresponding sub-problems are referred to as linear receiver matrix optimization and reflection coefficient optimization. Active precoding is optimized using the Lagrange optimization technique, while linear precoding is optimized by employing the First Order Necessary Condition (FONC) for optimizing a function. In both cases, the optimization of reflection coefficients is achieved through the gradient descent method [55, 56].

- For reflection coefficient optimization, we use a physically-consistent RIS model (exact RIS model) to show the effect of MC (and hence multiple reflections). For an efficient algorithmic implementation, we use the VAMP approach [46] for algorithm initialization. The initialization process enhances the convergence speed of the overall algorithm. We compare the results with those for the conventional RIS model, which is a first-order approximation of the exact RIS model.

- Simulation results are obtained considering the lossless and lossy conditions of RIS elements with varying the number of users. Furthermore, we consider both perfect and imperfect channel state information (CSI) for the simulations to show the robustness of the proposed method. Additionally, we generate simulation results by changing the number of RIS elements and adjusting the spacing among RIS elements. The obtained results are compared with those for the conventional RIS model in [46]. The results show that the overall performance of the system is significantly improved when optimizing joint active and passive beamforming in the presence of multiple reflections.

1.3.2 Physically-Consistent Modeling and Optimization of Non-local RIS-Assisted Multi-User MIMO Communication Systems

We propose a novel approach to simultaneously optimize active and passive beamforming in RIS-Assisted multi-user Multiple-Input Multiple-Output (MIMO) systems, while also addressing the optimization of MC and radiation patterns based on S-parameters within a physically-consistent framework for reflective RIS and transmissive RIS separately, resulting in a non-local RIS structure. Our design represents an innovative approach wherein we maintain the reliability of conventional tunable phase shifters in the dynamic part while introducing a sophisticated static part based on fundamental laws of physics, resulting in a non-local structure. This proposed method ensures ease of implementation but also enhances practicality compared to designs relying solely on fully connected complex tunable phase shifters, which require more overhead. The contributions of this work are summarized as follows:

- We formulate a joint optimization problem for active and passive beamforming, along with optimizing MC and physically consistent radiation patterns for multi-

user downlink transmission based on a physically-consistent RIS model considering a reflective RIS model. We optimize MC and radiation patterns through an offline approach for a class of channels, rather than optimizing it dynamically on the fly for each channel.

- Furthermore, we also optimize joint beamforming for a transmissive RIS model within a physically consistent setting. This includes optimizing the physically consistent radiation pattern while neglecting MC, assuming a fully transmissive mode. In this case, radiation patterns are also optimized using an offline optimization technique for a particular class of channels. Despite neglecting MC, we achieve a non-local (non-diagonal) RIS structure due to the optimized radiation patterns.

- The MC and radiation patterns offline optimization introduces a complex nested structure into the formulated problems, involving both outer and inner optimization problems. The inner problem focuses on optimizing RIS phase shifters and BS active precoding, while the outer problem involves optimizing MC and radiation patterns based on scattering S-parameters.

- The method described in [18,31] addresses the inner problem of reflective RIS, and the approach from [46] is applied to solve the inner problem of transmissive RIS. A novel solution approach for outer problems considering both setups is proposed separately using the projection gradient descent method.

- We employ parametric and geometric channel modeling to illustrate simulation results to show the robustness of the proposed model. Our simulation results demonstrate that the system performance is significantly improved through joint beamforming optimization, integrating engineered Mutual Coupling (MC) and radiation patterns. This improvement is achieved even with the proposed offline

optimization process, emphasizing the effectiveness beyond real-time adjustments.

1.4 Scholastic Outputs and Achievements

This thesis includes materials previously published, accepted, and submitted in peer-reviewed journals and conferences, as outlined in the table 1.1.

1.5 Thesis Organization and Notations

We organize the rest of the thesis as follows.

- In Chapter 2, we analyze joint beamforming optimization in RIS-aided MIMO systems under mutual coupling and multiple reflections. The system model and assumptions are discussed in Section 2.1. The optimization problem formulations for downlink and uplink transmissions are presented in Section 2.2. Section 2.3 presents the proposed solution approaches, while Section 2.4 represents the simulation results.

- In Chapter 3, we optimized MC and radiation patterns along with the joint beamforming. The system model and assumptions are discussed in Section 3.1. The optimization problem formulations for reflective and transmissive RIS models are presented in Section 3.2. Section 3.3 presents the proposed solution approaches, while Section 3.4 represents the simulation results.

- In Chapter 4, we include concluding remarks and outline promising avenues for future directions. Additionally, we provide some detailed derivations in the Appendix section.

Notations: Scalar variables are denoted by lowercase letters, while vectors and matrices are represented by small and capital boldface letters (e.g., $\mathbf{s}$ and $\mathbf{S}$). The s_k is the k-th element of the vector $\mathbf{s}$. The i-th column of matrix $\mathbf{A}$ is denoted by $\boldsymbol{a}_i$ and a_{ik} stands for its (i,k)-th entry. The transpose, conjugate, and conjugate transpose are denoted by $(.)^{\mathsf{T}}$, $(.)^*$, and $(.)^{\mathsf{H}}$, respectively. Notations $\|.\|_2$ and $\|.\|_{\mathrm{F}}$ stand for the Euclidean and Frobenius norms, respectively, whereas the trace and rank of a matrix are represented with $\mathrm{Tr}(\mathbf{A})$ and $\mathrm{Rank}(\mathbf{A})$, respectively. The expectation operator is indicated by $\mathbb{E}\{.\}$. The term $\mathrm{Diag}(\boldsymbol{a})$ refers to a diagonal matrix whose diagonal elements are the components of $\boldsymbol{a}$. $\mathrm{Diag}(V)$ represents a column vector of the main diagonal elements of V. $\otimes$ is the Kronecker product, whereas $\mathbf{I}_M$ and $\mathbf{0}_M$ are the $M \times M$ identity and all-zeros matrices, respectively.

Table 1.1: Summary of publications and achievements

Publications	**Appearance**
1. **D.Wijekoon**, A. Mezghani, E.Hossain, "Phase Shifter Optimization in RIS-Aided MIMO Systems Under Multiple Reflections" in *IEEE Transactions on Wireless Communications*, DOI: 10.1109/TWC.2024.3357354.	Chapter 2
2. **D.Wijekoon**, A. Mezghani, E.Hossain, "Beamforming Optimization in RIS-Aided MIMO Systems Under Multiple-Reflection Effects" in *ICASSP 2023 - 2023 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, DOI: 10.1109/ICASSP49357.2023.10096563.	Chapter 2
3. **D.Wijekoon**, A. Mezghani, George C. Alexandropoulos, E.Hossain, "Electromagnetically-Consistent Modeling and Optimization of Mutual Coupling in RIS-Assisted Multi-User MIMO Communication Systems" accepted in *IEEE ICC 2024 Workshops -IEEE International Conference on Communications Workshops* .	Chapter 3
4. **D.Wijekoon**, A. Mezghani, George C. Alexandropoulos, E.Hossain, "Physically-Consistent Modeling and Optimization of Non-local RIS-Assisted Multi-User MIMO Communication Systems" submitted to *IEEE Transactions on Wireless Communications*.	Chapter 3

Chapter 2

Joint Beamforming Optimization in RIS-Aided MIMO Systems Under Multiple Reflections

2.1 System Model and Assumptions

2.1.1 System Model

Consider a RIS-assisted single-cell multi-user MIMO system considering downlink and uplink transmissions as shown in Figs. 2.1 and 2.2, respectively. The BS is equipped with N antennas while each K ($K < N$) users have a single antenna. The RIS consists of M ($M > K$) reflective elements. Since we consider a physically-consistent model that takes into account the coupling effect, we assume the phase shift matrix as follows [31, 36]:

$$\Phi = (\boldsymbol{\Upsilon}^{-1} - \mathbf{S})^{-1}, \tag{2.1}$$

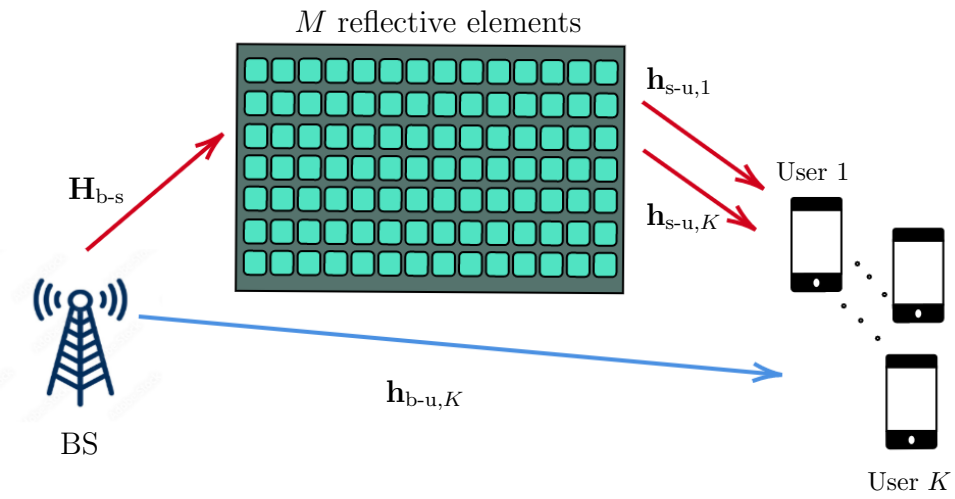


Figure 2.1: RIS-aided multiuser MIMO system for downlink transmission.

where $\boldsymbol{\Upsilon} = \text{Diag}(\boldsymbol{\upsilon}) \in \mathbb{C}^{M\times M}$ and $\boldsymbol{\upsilon} \in \mathbb{C}^{M}$ is the phase-shift vector, and $\mathbf{S}$ is the multi-port scattering matrix describing the scattering characteristics. As given in equation (2.2) below, for the physically-consistent model, the phase shift matrix can be represented by a sum of higher-order terms, which is also called Neumann series representation [36].

$$(\boldsymbol{\Upsilon}^{-1} - \mathbf{S})^{-1} = \sum_{l=0}^{\infty} (\boldsymbol{\Upsilon}\mathbf{S})^{l}\boldsymbol{\Upsilon} = \boldsymbol{\Upsilon} + \boldsymbol{\Upsilon}\mathbf{S}\boldsymbol{\Upsilon} + \cdots, \tag{2.2}$$

in which $\boldsymbol{\Upsilon}$ is referred to as the *load scattering matrix*. Typically, all conventional models ignore the multiple reflection effect represented by matrix $\mathbf{S}$. In other words, conventional models only consist of the first-order term given by (2.2) as the matrix $\mathbf{S}$ is zero. The elements of the phase shift vector $\boldsymbol{\upsilon}$ can be denoted by $\upsilon_m = e^{\mathrm{j}\theta_m}$, where $|\upsilon_m| = 1$ for $m = 1, \cdots, M$, and $\theta_m \in [0, 2\pi]$. That is, for the phase-shift vector, we restrict our selection to complex numbers that are located on a unit circle.

We use the following fundamental laws of physics to model the multi-port scattering matrix $\mathbf{S}$:

- Passivity: $\mathbf{S}\mathbf{S}^{\mathsf{H}} \leq \mathbf{I}$

- Reciprocity: $\mathbf{S} = \mathbf{S}^{\mathsf{T}}$
- Losslessness: $\mathbf{S} = \mathbf{S}^{\mathsf{T}} = (\mathbf{I} - \mathbf{B})^{\frac{1}{2}}$.

Here, $\mathbf{B}$ represents the embedded pattern coupling matrix, which complies with the passivity criteria, i.e. $\mathbf{B} \leq \mathbf{I}$. We select the cosine radiation pattern to be the optimal and compatible admissible effective area, denoted as $A_e(\theta, \varphi) = a^2 \cos\theta$, to achieve the passivity. Based on the principles of losslessness and reciprocity and leveraging the cosine radiation pattern, we can express the embedded pattern coupling matrix $\mathbf{B}$ as follows [31, 36]:

$$[\mathbf{B}]_{w+\sqrt{M}(x-1),y+\sqrt{M}(z-1)} = \int_0^{\pi/2} \int_{-\pi}^{\pi} A_e(\theta,\varphi) a(\theta,\varphi,f) a(\theta,\varphi,f)^{\mathsf{H}} \sin\theta \, d\varphi \, d\theta \tag{2.3}$$

$$= \begin{cases} \frac{\pi a^2}{\lambda^2} & ; \text{for } w = \text{y and } x = \text{z} \\ \frac{a}{\lambda} \frac{J_1\left(\frac{2\pi a}{\lambda}\sqrt{(w-y)^2+(x-z)^2}\right)}{\sqrt{(w-y)^2+(x-z)^2}} & ; \text{otherwise,} \end{cases} \tag{2.4}$$

where $J_1(.)$ is the first-order Bessel function of first-kind, and a is the distance between two consecutive reflective elements. Typically in literature, it is assumed that there is no coupling among the RIS elements, which means the matrix $\mathbf{B}$ is an identity matrix. Since we take the coupling among RIS elements into consideration, the matrix $\mathbf{B}$ is not an identity matrix anymore. We use w, x, y, and z to denote the positions of elements on the surface plane and also their indices in the matrix. Here, w, x, y, and z are counted from 1 to $\sqrt{M}$, where M represents the number of reflective elements of the surface and $\sqrt{M}$ denotes the number of rows and columns of the surface. Then we utilize the following equations to construct the matrix $\mathbf{S}$:

$$\mathbf{S} = \mathbf{U}\sqrt{\mathbf{I}_M - \Lambda}\mathbf{U}^{\mathsf{T}}, \tag{2.5}$$

$$\mathbf{B} = \mathbf{I}_M - \mathbf{S}\mathbf{S}^{\mathsf{H}} = \mathbf{U}\Lambda\mathbf{U}^{\mathsf{H}}. \tag{2.6}$$

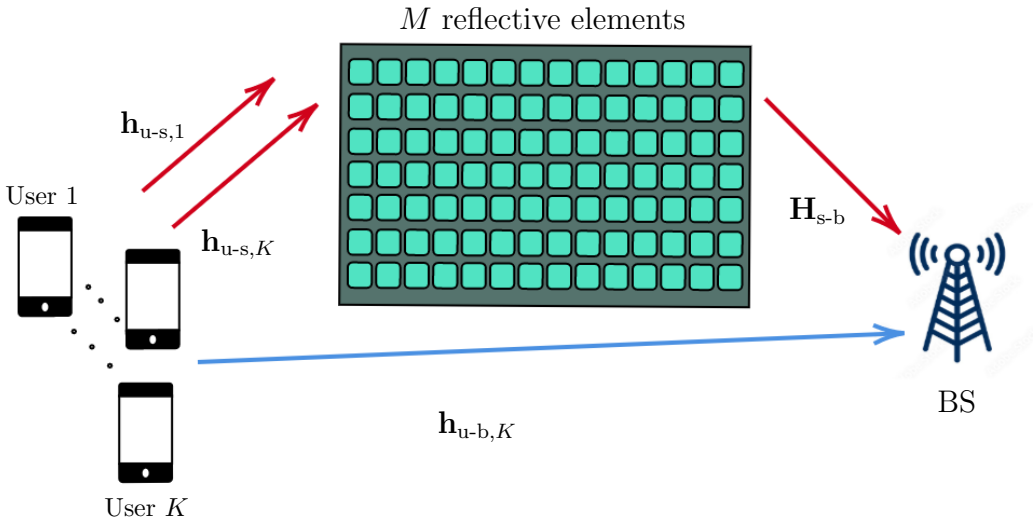


Figure 2.2: RIS-aided multiuser MIMO system for uplink transmission.

Equation (2.6) shows the singular/eigenvalue decomposition of matrix $\mathbf{B}$. Matrix Λ represents a diagonal matrix where the diagonal elements are the eigenvalues of matrix $\mathbf{B}$ and the columns of the matrix $\mathbf{U}$ represent the corresponding eigenvectors of matrix $\mathbf{B}$.

Note that the above model is a more general model based on the basic laws of physics and can be applied to any type of antenna, regardless of their specific properties.

2.1.2 Downlink Transmission

The channel for the k-th user from the BS is denoted by the channel vector $\mathbf{h}_{\text{b-u},k} \in \mathbb{C}^N$ while the channel corresponding to the RIS-k-th user link is denoted by $\mathbf{h}_{\text{s-u},k} \in \mathbb{C}^M$. The channel $\mathbf{H}_{\text{b-s}} \in \mathbb{C}^{M \times N}$ represents the BS-RIS link and satisfies the condition of $\text{Rank}(\mathbf{H}_{\text{b-s}}) \geq K$. The received signal at the k-th user is

$$\mathsf{y}_k = \alpha\Bigg(\mathbf{h}_{\text{s-u},k}^{\mathsf{H}}(\boldsymbol{\Upsilon}^{-1} - \mathbf{S})^{-1}\mathbf{H}_{\text{b-s}}\sum_{k'=1}^{K}\mathbf{f}_{k'}\mathsf{s}_{k'} \;+\; \mathbf{h}_{\text{b-u},k}^{\mathsf{H}}\sum_{k'=1}^{K}\mathbf{f}_{k'}\mathsf{s}_{k'} + \;\mathsf{w}\Bigg); \quad k = 1, \cdots, K. \tag{2.7}$$

Here, $\alpha \in \mathbb{R}$ refers to a scaling factor related to the receiver and is assumed to be common for all users. The terms $\mathsf{s}_k \sim \mathcal{CN}(s; 0, 1)$ and $\mathsf{w} \sim \mathcal{CN}(w; 0, \sigma_{\text{w}}^2)$ denote the transmit symbol and the Additive White Gaussian Noise (AWGN), respectively. The precoding vectors employed for power allocation are represented by $\mathbf{f}_k \in \mathbb{C}^N$ for $k = 1, \cdots, K$. The terms $\mathbf{s} = [\mathsf{s}_1, \mathsf{s}_2, \cdots, \mathsf{s}_K]^{\mathsf{T}}$ and $\mathbf{F} = [\mathbf{f}_1, \mathbf{f}_2, \cdots, \mathbf{f}_K]$ denotes the transmit symbol vector and precoding matrix respectively, which satisfy the condition $\mathbb{E}\left\{\|\mathbf{Fs}\|^2\right\} = P$. Here, P denotes the total transmit power. We can express the total received signal vector as in equation (2.8), where $\mathbf{y} = [\mathsf{y}_1, \mathsf{y}_2, \cdots, \mathsf{y}_K]^{\mathsf{T}}$ denotes the received signal vector.

$$\mathbf{y} = \alpha\left(\underbrace{\mathbf{H}_{\text{s-u}}^{\mathsf{H}}(\boldsymbol{\Upsilon}^{-1} - \mathbf{S})^{-1}\mathbf{H}_{\text{b-s}}\mathbf{Fs}}_{\text{BS-RIS-Users}} + \underbrace{\mathbf{H}_{\text{b-u}}^{\mathsf{H}}\mathbf{Fs}}_{\text{BS-Users}} + \;\mathbf{w}\right). \tag{2.8}$$

The matrices $\mathbf{H}_{\text{b-u}} = [\mathbf{h}_{\text{b-u},1}, \mathbf{h}_{\text{b-u},2}, \cdots, \mathbf{h}_{\text{b-u},K}]$ and $\mathbf{H}_{\text{s-u}} = [\mathbf{h}_{\text{s-u},1}, \mathbf{h}_{\text{s-u},2}, \cdots, \mathbf{h}_{\text{s-u},K}]$ denote channel matrices corresponding to BS-user and RIS-user links, respectively. The corresponding total effective channel matrix is given by

$$\mathbf{H}^{\mathsf{H}} = \mathbf{H}_{\text{s-u}}^{\mathsf{H}}(\boldsymbol{\Upsilon}^{-1} - \mathbf{S})^{-1}\mathbf{H}_{\text{b-s}} + \mathbf{H}_{\text{b-u}}^{\mathsf{H}}. \tag{2.9}$$

2.1.3 Uplink Transmission

The matrix $\mathbf{H}_{\text{s-b}} \in \mathbb{C}^{N\times M}$ gives the channel matrix corresponding to RIS-BS link. Moreover, the vectors $\mathbf{h}_{k,\text{u-b}} \in \mathbb{C}^N$ and $\mathbf{h}_{k,\text{u-s}} \in \mathbb{C}^M$ correspond to the channel vector of the k-th user-BS link and the channel vector of the k-th user-RIS link, respectively. The received signal at the n-th antenna in the BS is

$$\mathsf{y}_n = \mathbf{H}_{\text{s-b}}(\boldsymbol{\Upsilon}^{-1} - \mathbf{S})^{-1}\sum_{k'=1}^{K}\mathbf{h}_{k',\text{u-s}}\mathsf{s}_{k'} \;+\sum_{k'=1}^{K}\mathbf{h}_{k',\text{u-b}}\mathsf{s}_{k'} + \;\mathsf{w}; \quad n = 1, \cdots, N. \tag{2.10}$$

Similar to the downlink case, the transmit signal and the AWGN are indicated by $\mathsf{s}_k \sim \mathcal{CN}(s;0,1)$ and $\mathsf{w} \sim \mathcal{CN}(w;0,\sigma_{\mathrm{w}}^2)$. The total received signal of the uplink transmission can be expressed as

$$\mathbf{y} = \underbrace{\mathbf{H}_{\text{s-b}}(\boldsymbol{\Upsilon}^{-1} - \mathbf{S})^{-1}\mathbf{H}_{\text{u-s}}\mathbf{s}}_{\text{Users-RIS-BS}} + \underbrace{\mathbf{H}_{\text{u-b}}\mathbf{s}}_{\text{Users-BS}} + \mathbf{w}. \tag{2.11}$$

Therefore, the total channel matrix for the uplink transmission is

$$\mathbf{H} = \mathbf{H}_{\text{s-b}}(\boldsymbol{\Upsilon}^{-1} - \mathbf{S})^{-1}\mathbf{H}_{\text{u-s}} + \mathbf{H}_{\text{u-b}}. \tag{2.12}$$

2.2 Problem Formulation

This section provides the optimization problem formulations for downlink and uplink transmissions separately.

2.2.1 Problem Formulation for Downlink Transmission

To optimize the spectral efficiency, the optimization problem is formulated based on the Minimum Mean Squared Error (MMSE) criterion. The spectral efficiency can be maximized by minimizing the receive symbol error [57], which is equivalent to minimizing the error between the received symbols and input symbols. Based on the MMSE criterion, a lower bound on spectral efficiency of k th user is given as follows [46]:

$$C_k^{\text{MMSE}} = \log_2\left(\frac{1}{\text{MMSE}_k}\right). \tag{2.13}$$

Then, the total spectral efficiency for all K users can be given by

$$\widehat{C} \;=\; \sum_{k=1}^{K} \log_2\left(\frac{1}{\text{MMSE}_k}\right) \;=\; \log_2\left(\prod_{k=1}^{K} \frac{1}{\text{MMSE}_k}\right). \tag{2.14}$$

The total mean squared error (MSE) of the received symbols for K users can be expressed as

$$\sum_{k=1}^{K} \mathbb{E}_{\mathsf{y}_k,\mathsf{s}_k}\left\{|\mathsf{y}_k - \mathsf{s}_k|^2\right\} \;=\; \mathbb{E}_{\mathbf{y},\mathbf{s}}\left\{\|\mathbf{y}-\mathbf{s}\|_2^2\right\}, \tag{2.15}$$

where $\mathbb{E}_{\mathsf{y}_k,\mathsf{s}_k}\left\{|\mathsf{y}_k - \mathsf{s}_k|^2\right\}$ is the MSE of the received symbol for k-th user. Accordingly, the optimization problem for downlink active precoding and passive beamforming is given by

$$\begin{aligned}
\underset{\alpha,\mathbf{F},\boldsymbol{\Upsilon}}{\arg\min} \quad & \mathbb{E}_{\mathbf{y},\mathbf{s}}\left\{\|\mathbf{y}-\mathbf{s}\|_2^2\right\}, && \text{(2.16a)}\\
\text{subject to} \quad & \mathbb{E}_{\mathbf{s}}\left\{\|\mathbf{F}\mathbf{s}\|_2^2\right\} = P, && \text{(2.16b)}\\
& \upsilon_{im} = 0, \quad i \neq m, && \text{(2.16c)}\\
& |\upsilon_{ii}| = 1, \quad i = 1, 2, \cdots, M. && \text{(2.16d)}
\end{aligned}$$

The constraint (2.16b) corresponds to power allocation, while constraints (2.16c), (2.16d) are related to the phase shift matrix. Note that, under the latter two constraints, the *load scattering matrix* $\boldsymbol{\Upsilon}$ becomes a diagonal matrix where the magnitudes of the diagonal elements are equal to one. The term $\mathbb{E}_{\mathbf{y},\mathbf{s}}\{\|\mathbf{y}-\mathbf{s}\|_2^2\}$ can be transformed into another expectation function in terms of the random vectors $\mathbf{s}$ and $\mathbf{w}$ as given in equation (2.17) below:

$$\mathbb{E}_{\mathbf{w},\mathbf{s}}\Big\{\mathrm{Tr}\big((\alpha\mathbf{H}^{\mathsf{H}}\mathbf{F}\mathbf{s} + \alpha\mathbf{w} - \mathbf{s})(\alpha\mathbf{s}^{\mathsf{H}}\mathbf{F}^{\mathsf{H}}\mathbf{H} + \alpha\mathbf{w}^{\mathsf{H}} - \mathbf{s}^{\mathsf{H}})\big)\Big\}. \tag{2.17}$$

The detailed derivation of this conversion is provided in the **Appendix**. Thereby, for the proposed model, the non-convex optimization problem for the downlink transmission is given as follows:

$$\underset{\alpha,\mathbf{F},\boldsymbol{\Upsilon}}{\arg\min} \quad \left\|\alpha\mathbf{H}_{\text{s-u}}^{\mathsf{H}}(\boldsymbol{\Upsilon}^{-1}-\mathbf{S})^{-1}\mathbf{H}_{\text{b-s}}\mathbf{F}+\alpha\mathbf{H}_{\text{b-u}}^{\mathsf{H}}\mathbf{F}-\mathbf{I}_K\right\|_{\text{F}}^2+K\alpha^2\sigma_{\text{w}}^2, \tag{2.18a}$$

$$\text{subject to} \quad \|\mathbf{F}\|_{\text{F}}^2=P, \tag{2.18b}$$

$$\upsilon_{im}=0, \quad i\neq m, \tag{2.18c}$$

$$|\upsilon_{ii}|=1, \quad i=1,2,\cdots,M. \tag{2.18d}$$

The matrix $\mathbf{I}_K$ is an identity matrix of dimension $K\times K$ and σ_{w}^2 is the noise variance. Hence, the problem is a multi-variable constrained optimization problem.

2.2.2 Problem Formulation for Uplink Transmission

In this case, we exploit the concept of Linear-MMSE (LMMSE) estimation. The LMMSE is a linear transformation that minimizes the MSE, and it is given by

$$\hat{\mathbf{s}}=\mathbf{G}\mathbf{y}, \tag{2.19}$$

where $\hat{\mathbf{s}}$ is the input symbol estimator vector, $\mathbf{G}$ is the linear receiver matrix, and $\mathbf{y}$ is the output signal vector as given in (2.11). Note that the MMSE estimator becomes an LMMSE estimator when the input and the noise vectors are Gaussian. Since the input signal vector and the noise vectors are assumed to be Gaussian, we apply the LMMSE concept here.

To formulate the optimization problem, we again consider the relationship between the spectral efficiency and the MMSE given in equations (2.13) and (2.14). The target is to minimize the error between the input symbol vector $\mathbf{s}$ and the input symbol estimator vector $\hat{\mathbf{s}}$. The MSE corresponding to the input symbols for K users is given by

$$\sum_{k=1}^{K}\mathbb{E}_{\mathsf{s}_k,\hat{\mathsf{s}_k}}\left\{|\mathsf{s}_k-\hat{\mathsf{s}_k}|^2\right\} \;=\; \mathbb{E}_{\mathbf{s},\hat{\mathbf{s}}}\left\{\|\mathbf{s}-\hat{\mathbf{s}}\|_2^2\right\}, \tag{2.20}$$

where the term $\mathbb{E}_{\mathsf{s}_k,\hat{\mathsf{s}_k}}\left\{|\mathsf{s}_k-\hat{\mathsf{s}_k}|^2\right\}$ represents the MSE of the transmit symbol for the k-th user. Therefore, the optimization problem for uplink transmission can be stated

as follows:

$$\underset{\mathbf{G},\boldsymbol{\Upsilon}}{\arg\min} \quad \mathbb{E}_{\mathbf{s},\hat{\mathbf{s}}}\left\{\|\mathbf{s}-\hat{\mathbf{s}}\|_2^2\right\}, \tag{2.21a}$$

$$\text{subject to} \quad \upsilon_{im}=0, \quad i\neq m, \tag{2.21b}$$

$$|\upsilon_{ii}|=1, \quad i=1,2,\cdots,M. \tag{2.21c}$$

The derivations are given in the **Appendix**. The optimization problem for the uplink transmission can be restated as

$$\underset{\mathbf{G},\boldsymbol{\Upsilon}}{\arg\min} \quad p_k\left\|\mathbf{I}_K-\mathbf{G}\mathbf{H}_{\text{s-b}}(\boldsymbol{\Upsilon}^{-1}-\mathbf{S})^{-1}\mathbf{H}_{\text{u-s}}-\mathbf{G}\mathbf{H}_{\text{u-b}}\right\|_{\text{F}}^2+\sigma_{\text{w}}^2\|\mathbf{G}\|_{\text{F}}^2, \tag{2.22a}$$

$$\text{subject to} \quad \upsilon_{im}=0, \quad i\neq m, \tag{2.22b}$$

$$|\upsilon_{ii}|=1, \quad i=1,2,\cdots,M, \tag{2.22c}$$

where p_k is the transmission power of the k-th user. This is also a multi-variable optimization problem with two constraints. Note that, unlike the downlink case, the problem does not involve a separate power allocation constraint. Instead, the power allocation is included in the objective function itself.

2.3 Proposed Solution Approaches

Both of the formulated problems are multi-variable, non-convex, and constrained optimization problems. We employ the alternating optimization approach to decompose each formulated problem into two sub-problems. Thereby, we optimize the precoding matrix $\mathbf{F}$ and the *load scattering matrix* $\boldsymbol{\Upsilon}$ corresponding to downlink transmission, and the linear receiver matrix $\mathbf{G}$ and the *load scattering matrix* $\boldsymbol{\Upsilon}$ corresponding to uplink transmission.

2.3.1 Alternating Optimization

An iterative alternating optimization algorithm optimizes the objective functions by dividing the main problem into sub-problems [46, 58–60]. Here, we apply alternating optimization for both downlink and uplink transmissions.

Downlink Transmission

First, we optimize the precoding matrix $\mathbf{F}$ along with the scaling parameter α and then optimize the load scattering matrix $\boldsymbol{\Upsilon}$. The decomposed problems are given by [31]:

1.

$$\underset{\alpha,\mathbf{F}}{\arg\min} \quad \mathbb{E}_{\mathbf{y},\mathbf{s}} \left\{ \|\mathbf{y}-\mathbf{s}\|_2^2 \right\}, \tag{2.23a}$$

$$\text{subject to} \quad \mathbb{E}_{\mathbf{s}} \|\mathbf{F}\mathbf{s}\|_2^2 = P. \tag{2.23b}$$

2.

$$\underset{\boldsymbol{\Upsilon}}{\arg\min} \quad \mathbb{E}_{\mathbf{y},\mathbf{s}} \left\{ \|\mathbf{y}-\mathbf{s}\|_2^2 \right\}, \tag{2.24a}$$

$$\text{subject to} \quad \upsilon_{im} = 0, \quad i \neq m, \tag{2.24b}$$

$$|\upsilon_{ii}| = 1, \quad i = 1, 2, \cdots, M. \tag{2.24c}$$

Thereby, the sub-problem of optimizing scaling parameter α and precoding matrix $\mathbf{F}$ corresponding to the downlink transmission is given by:

$$\underset{\alpha,\mathbf{F}}{\arg\min} \quad \left\| \alpha \Big(\mathbf{H}_{\text{s-u}}^{\mathsf{H}} (\boldsymbol{\Upsilon}^{-1} - \mathbf{S})^{-1} \mathbf{H}_{\text{b-s}} + \mathbf{H}_{\text{b-u}}^{\mathsf{H}} \Big) \mathbf{F} - \mathbf{I}_K \right\|_{\text{F}}^2 + K\alpha^2\sigma_{\text{w}}^2, \tag{2.25a}$$

$$\text{subject to} \quad \|\mathbf{F}\|_{\text{F}}^2 = P. \tag{2.25b}$$

The sub-problem for optimizing the load scattering matrix $\boldsymbol{\Upsilon}$ is given by:

$$\arg\min_{\boldsymbol{\Upsilon}} \left\| \alpha\Big(\mathbf{H}_{\text{s-u}}^{\mathsf{H}} (\boldsymbol{\Upsilon}^{-1} - \mathbf{S})^{-1} \mathbf{H}_{\text{b-s}} + \mathbf{H}_{\text{b-u}}^{\mathsf{H}} \Big) \mathbf{F} - \mathbf{I}_K \right\|_{\text{F}}^2, \tag{2.26a}$$

$$\text{subject to} \quad v_{im} = 0 \quad i \neq m, \tag{2.26b}$$

$$|v_{ii}| = 1 \quad i = 1, 2, \cdots, M. \tag{2.26c}$$

Uplink Transmission

First, we optimize the linear receiver matrix $\mathbf{G}$ and then optimize the load scattering matrix $\boldsymbol{\Upsilon}$. After applying the alternating approach, we obtain the sub-problems problem as:

1.

$$\arg\min_{\mathbf{G}} \quad \mathbb{E}_{\mathbf{s},\hat{\mathbf{s}}} \left\{ \|\mathbf{s} - \hat{\mathbf{s}}\|_2^2 \right\}. \tag{2.27a}$$

2.

$$\arg\min_{\boldsymbol{\Upsilon}} \quad \mathbb{E}_{\mathbf{s},\hat{\mathbf{s}}} \left\{ \|\mathbf{s} - \hat{\mathbf{s}}\|_2^2 \right\}, \tag{2.28a}$$

$$\text{subject to} \quad v_{im} = 0, \quad i \neq m, \tag{2.28b}$$

$$|v_{ii}| = 1, \quad i = 1, 2, \cdots, M. \tag{2.28c}$$

Thereby, the sub-problem of optimizing the linear receiver matrix $\mathbf{G}$ is:

$$\arg\min_{\mathbf{G}} \quad p_k \left\| \mathbf{I}_K - \mathbf{G}\Big(\mathbf{H}_{\text{s-b}} (\boldsymbol{\Upsilon}^{-1} - \mathbf{S})^{-1} \mathbf{H}_{\text{u-s}} + \mathbf{H}_{\text{u-b}} \Big) \right\|_{\text{F}}^2 + \sigma_{\text{w}}^2 \|\mathbf{G}\|_{\text{F}}^2 . \tag{2.29}$$

The sub-problem of optimizing the load scattering matrix $\boldsymbol{\Upsilon}$ is:

$$\arg\min_{\boldsymbol{\Upsilon}} \quad p_k \left\| \mathbf{I}_K - \mathbf{G}\Big(\mathbf{H}_{\text{s-b}} (\boldsymbol{\Upsilon}^{-1} - \mathbf{S})^{-1} \mathbf{H}_{\text{u-s}} + \mathbf{H}_{\text{u-b}} \Big) \right\|_{\text{F}}^2, \tag{2.30a}$$

$$\text{subject to} \quad v_{im} = 0, \quad i \neq m, \tag{2.30b}$$

$$|v_{ii}| = 1, \quad i = 1, 2, \cdots, M. \tag{2.30c}$$

2.3.2 Precoding Matrix Optimization for Downlink Transmission

The problem of jointly optimizing $\mathbf{F}$ and α in (2.23) is solved by using the Lagrange multiplier method. The Lagrangian function corresponding to the active precoder optimization [31, 46] is given by

$$\mathcal{L}(\lambda, \alpha, \mathbf{F}) = \left\|\alpha \mathbf{H}^{\mathsf{H}}\mathbf{F} - \mathbf{I}_K\right\|_{\mathrm{F}}^2 + K\alpha^2\sigma_{\mathrm{w}}^2 + \lambda(\mathrm{Tr}(\mathbf{F}\mathbf{F}^{\mathsf{H}}) - P), \tag{2.31}$$

where $\lambda \in \mathbb{R}$ is the Lagrange multiplier. Setting the partial derivatives of the Lagrangian function to zero, the optimum closed form solutions for $\mathbf{F}$ and α are obtained as follows [31, 46, 61]:

$$\alpha^{\mathrm{opt}} = \sqrt{\frac{1}{P}}\sqrt{\mathrm{Tr}\left(\left[\mathbf{H}\mathbf{H}^{\mathsf{H}} + \frac{K\sigma_{\mathrm{w}}^2\mathbf{I}_N}{P}\right]^{-2}\mathbf{H}\mathbf{H}^{\mathsf{H}}\right)} \tag{2.32}$$

$$\begin{aligned}\mathbf{F}^{\mathrm{opt}} =& \frac{\sqrt{P}\left[\mathbf{H}\mathbf{H}^{\mathsf{H}} + \frac{K\sigma_{\mathrm{w}}^2\mathbf{I}_N}{P}\right]^{-1}\mathbf{H}}{\sqrt{\mathrm{Tr}\left(\left[\mathbf{H}\mathbf{H}^{\mathsf{H}} + \frac{K\sigma_{\mathrm{w}}^2\mathbf{I}_N}{P}\right]^{-2}\mathbf{H}\mathbf{H}^{\mathsf{H}}\right)}} \\ =& \alpha^{\mathrm{opt}^{-1}}\left[\mathbf{H}\mathbf{H}^{\mathsf{H}} + \frac{K\sigma_{\mathrm{w}}^2\mathbf{I}_N}{P}\right]^{-1}\mathbf{H},\end{aligned} \tag{2.33}$$

where $\mathbf{H}$ is the channel matrix given by equation (2.9).

2.3.3 Linear Receiver Optimization for Uplink Transmission

The linear receiver optimization problem is an unconstrained optimization problem that can be directly solved by taking the gradient of the function. The objective function is as follows (2.34):

$$f(\mathbf{G}) = p_k \left\|\mathbf{I}_K - \mathbf{G}\mathbf{H}\right\|_{\mathrm{F}}^2 + \sigma_{\mathrm{w}}^2 \left\|\mathbf{G}\right\|_{\mathrm{F}}^2. \tag{2.34}$$

Then we used the following transformation to simplify the objective function (2.35):

$$f(\mathbf{G}) = p_k \mathrm{Tr}\big((\mathbf{I}_K - \mathbf{G}\mathbf{H})(\mathbf{I}_K - \mathbf{G}\mathbf{H})^{\mathsf{H}}\big) + \sigma_{\mathrm{w}}^2 \mathrm{Tr}\big(\mathbf{G}\mathbf{G}^{\mathsf{H}}\big). \tag{2.35}$$

Taking the derivative of (2.35) and utilizing the First Order Necessary Condition (FONC) we have,

$$f'(\mathbf{G}) \;=\; -2p_k(\mathbf{I}_K - \mathbf{G}\mathbf{H})\mathbf{H}^{\mathsf{H}} + 2\sigma_{\mathrm{w}}^2\mathbf{G} \;=\; 0. \tag{2.36}$$

Hence, the optimum solution of the linear receiver matrix is

$$\mathbf{G}^{\mathrm{opt}} = \mathbf{H}^{\mathsf{H}}\left[\mathbf{H}\mathbf{H}^{\mathsf{H}} + \frac{\sigma_{\mathrm{w}}^2\mathbf{I}_N}{p_k}\right]^{-1}, \tag{2.37}$$

where the matrix $\mathbf{H}$ is given in (2.12) and $\mathbf{I}_N$ is an identity matrix of size $N \times N$.

2.3.4 Phase-Shift Matrix Optimization

We use the Gradient Descent (GD) method [55, 56, 62] to optimize the phase shift matrix.

Downlink transmission

The objective function corresponding to optimizing the phase-shift matrix associated with downlink transmission is given by

$$\boldsymbol{\Upsilon}_{t+1} = \arg\min_{\boldsymbol{\Upsilon}} \left\| \alpha_t \left(\mathbf{H}_{\text{s-u}}^{\mathsf{H}} (\boldsymbol{\Upsilon}_t^{-1} - \mathbf{S})^{-1} \mathbf{H}_{\text{b-s}} + \mathbf{H}_{\text{b-u}}^{\mathsf{H}} \right) \mathbf{F}_t - \mathbf{I}_K \right\|_{\mathrm{F}}^2. \tag{2.38}$$

To simplify the calculation of the gradient, the equation (2.38) can be transformed into the following format.

$$\boldsymbol{\Upsilon}_{t+1} = \min_{\boldsymbol{\Upsilon}} \mathrm{Tr}\left(\left(\alpha_t (\mathbf{H}_{\text{s-u}}^{\mathsf{H}} (\boldsymbol{\Upsilon}_t^{-1} - \mathbf{S})^{-1} \mathbf{H}_{\text{b-s}} + \mathbf{H}_{\text{b-u}}^{\mathsf{H}}) \mathbf{F}_t - \mathbf{I}_K \right) \left(\alpha_t (\mathbf{H}_{\text{s-u}}^{\mathsf{H}} (\boldsymbol{\Upsilon}_t^{-1} - \mathbf{S})^{-1} \mathbf{H}_{\text{b-s}} + \mathbf{H}_{\text{b-u}}^{\mathsf{H}}) \mathbf{F}_t - \mathbf{I}_K \right)^{\mathsf{H}} \right). \tag{2.39}$$

The gradient of the objective function is given by

$$\begin{aligned} &\nabla \boldsymbol{\Upsilon}_{t+1} = \\ &2\boldsymbol{\Upsilon}_t^{-\mathsf{T}} (\boldsymbol{\Upsilon}_t^{-1} - \mathbf{S})^{-\mathsf{T}} (\alpha_t \mathbf{H}_{\text{s-u}}^{\mathsf{H}})^{\mathsf{T}} \left(\alpha_t (\mathbf{H}_{\text{s-u}}^{\mathsf{H}} (\boldsymbol{\Upsilon}_t^{-1} - \mathbf{S})^{-1} \mathbf{H}_{\text{b-s}} + \mathbf{H}_{\text{b-u}}^{\mathsf{H}}) \mathbf{F}_t - \mathbf{I}_K \right)^{\mathsf{H}} \\ &(\mathbf{H}_{\text{b-s}} \mathbf{F}_t)^{\mathsf{T}} (\boldsymbol{\Upsilon}_t^{-1} - \mathbf{S})^{-\mathsf{T}} \boldsymbol{\Upsilon}_t^{-\mathsf{T}}. \end{aligned} \tag{2.40}$$

Uplink Transmission

The objective function of optimizing the phase-shift matrix corresponding to uplink transmission is

$$\boldsymbol{\Upsilon}_{t+1} = \arg\min_{\boldsymbol{\Upsilon}} p_k \left\| \mathbf{I}_K - \mathbf{G}_t \left(\mathbf{H}_{\text{s-b}} (\boldsymbol{\Upsilon}_t^{-1} - \mathbf{S})^{-1} \mathbf{H}_{\text{u-s}} + \mathbf{H}_{\text{u-b}} \right) \right\|_{\text{F}}^2, \tag{2.41}$$

which can be transformed as:

$$\boldsymbol{\Upsilon}_{t+1} = \min_{\boldsymbol{\Upsilon}} \text{Tr} \Bigg(\big(\mathbf{I}_K - \mathbf{G}_t (\mathbf{H}_{\text{s-b}} (\boldsymbol{\Upsilon}_t^{-1} - \mathbf{S})^{-1} \mathbf{H}_{\text{u-s}} + \mathbf{H}_{\text{u-b}}) \big) \big(\mathbf{I}_K - \mathbf{G}_t (\mathbf{H}_{\text{s-b}} (\boldsymbol{\Upsilon}_t^{-1} - \mathbf{S})^{-1} \mathbf{H}_{\text{u-s}} + \mathbf{H}_{\text{u-b}}) \big)^{\mathsf{H}} \Bigg). \tag{2.42}$$

Then, the gradient of the objective function can be given by

$$\nabla \boldsymbol{\Upsilon}_{t+1} = 2 p_k \boldsymbol{\Upsilon}_t^{-\mathsf{T}} (\boldsymbol{\Upsilon}_t^{-1} - \mathbf{S})^{-\mathsf{T}} (\mathbf{G} \mathbf{H}_{\text{s-b}}^{\mathsf{H}})^{\mathsf{T}} \big(\mathbf{I}_K - \mathbf{G}_t (\mathbf{H}_{\text{s-b}} (\boldsymbol{\Upsilon}_t^{-1} - \mathbf{S})^{-1} \mathbf{H}_{\text{u-s}} + \mathbf{H}_{\text{u-b}}) \big)^{\mathsf{H}} (\mathbf{H}_{\text{u-s}})^{\mathsf{T}} (\boldsymbol{\Upsilon}_t^{-1} - \mathbf{S})^{-\mathsf{T}} \boldsymbol{\Upsilon}_t^{-\mathsf{T}}. \tag{2.43}$$

Once the gradient matrix is obtained, it is converted into a gradient vector to satisfy the constraints specified in equations (2.24b) and (2.28b) related to each transmission as follows:

$$\nabla \boldsymbol{\Upsilon}_{t+1} = \text{Diag}\big(\nabla \boldsymbol{\Upsilon}_{t+1} \big). \tag{2.44}$$

We consider the Wirtinger derivative here. Since the domain is complex, the conjugate of the gradient vector is taken as the search direction, and the new phase-shift vector is updated as follows:

$$\boldsymbol{\upsilon}_{t+1} = \boldsymbol{\upsilon}_t - \mu \nabla \boldsymbol{\Upsilon}_t^*. \tag{2.45}$$

Here, the term μ represents the step size. In this study, we consider a varying step size instead of a fixed step size. To determine the step size at each step related to each search point, the Secant method [63] is utilized. Then we update the phase-shift vector

to satisfy the unimodular constraints as follows:

$$\boldsymbol{\upsilon}_{t+1} < \frac{\boldsymbol{\upsilon}_{t+1}}{|\boldsymbol{\upsilon}_{t+1}|}, \tag{2.46}$$

where $|\boldsymbol{\upsilon}_{t+1}|$ is the absolute value of the vector $\boldsymbol{\upsilon}_{t+1}$. Additionally, we analyze the performance by adding a phase-shift loss to RIS elements as follows:

$$\boldsymbol{\upsilon}_{t+1} < \frac{\boldsymbol{\upsilon}_{t+1}}{\sqrt{\boldsymbol{\upsilon}_{loss}}}, \quad \sqrt{\boldsymbol{\upsilon}_{loss}} > 1, 2, \cdots, \tag{2.47}$$

where the term $\sqrt{\boldsymbol{\upsilon}_{loss}}$ denotes the phase-shift loss. When $\boldsymbol{\upsilon}_{loss} = 1$, it signifies no loss. On the other hand, $\boldsymbol{\upsilon}_{loss} = 2, 3 \cdots$ implies losses corresponding to 3dB, 6dB, $\cdots$, and so on. The process is repeated for each iteration.

The overall algorithm representations for downlink and uplink transmissions are given in **Algorithm 1** and **Algorithm 2**, respectively. The algorithms converge when the stopping criterion of $|E_t - E_{t-1}| < \epsilon E_{t-1}$ is satisfied. Here $\epsilon \in \mathbb{R}_+$ is a precision tolerance, and for the downlink transmission the term E_t is given by:

$$E_t \quad \triangleq \quad \left\| \widehat{\alpha}_t \left(\mathbf{H}_{\text{s-u}}^{\mathsf{H}} (\widehat{\boldsymbol{\Upsilon}}_t^{-1} - \mathbf{S})^{-1} \mathbf{H}_{\text{b-s}} + \mathbf{H}_{\text{b-u}}^{\mathsf{H}} \right) \widehat{\mathbf{F}}_t - \mathbf{I}_K \right\|_{\text{F}}^2 \; + \; K \widehat{\alpha}_t^2 \sigma_{\text{w}}^2. \tag{2.48}$$

For the uplink transmission:

$$E_t \quad \triangleq \quad p_k \left\| \mathbf{I}_K - \widehat{\mathbf{G}}_t \mathbf{H}_{\text{s-b}} (\widehat{\boldsymbol{\Upsilon}}_t^{-1} - \mathbf{S})^{-1} \mathbf{H}_{\text{u-s}} - \widehat{\mathbf{G}}_t \mathbf{H}_{\text{u-b}} \right\|_{\text{F}}^2 \; + \; \sigma_{\text{w}}^2 \left\| \widehat{\mathbf{G}}_t \right\|_{\text{F}}^2. \tag{2.49}$$

2.3.5 Initialization Process

We employ the VAMP algorithm for initializing the gradient descent algorithm. The study in [46] demonstrated that the VAMP algorithm outperforms other existing optimization algorithms. This algorithm is specifically designed for quadratic loss optimization problems [46], where the optimization problem is in the form of a least squares problem. With the conventional RIS model, which uses $\Phi = \boldsymbol{\Upsilon}$, the optimization problem can be transformed into a least squares optimization problem. With the physically consistent model, which utilizes

Algorithm 1 Joint Optimization of Phase Shift and Precoding Matrices for Downlink Transmission

Inputs: $\mathbf{H}_{\text{b-s}}$, $\mathbf{H}_{\text{b-u}}$, $\mathbf{H}_{\text{s-u}}$, $\mathbf{S}$

At $t \leftarrow 0$, initialize $\widehat{\boldsymbol{v}}_0$

Compute $\widehat{\mathbf{H}}_0 = \left(\mathbf{H}_{\text{s-u}}^{\mathsf{H}}(\text{Diag}(\widehat{\boldsymbol{v}}_0)^{-1} - \mathbf{S})^{-1}\mathbf{H}_{\text{b-s}} + \mathbf{H}_{\text{b-u}}^{\mathsf{H}}\right)^{\mathsf{H}}$

Compute $\widehat{\alpha}_0$

Compute $\widehat{\mathbf{F}}_0$

Then set $t = 1$

Initialize first 1:T_1 iterations via the VAMP algorithm

$t = t + 1$

repeat

    // Gradient Descent Algorithm //

    Set $\mathbf{A} = \widehat{\alpha}_{t-1}\mathbf{H}_{\text{s-u}}^{\mathsf{H}}$, $\mathbf{B} = \left(\mathbf{H}_{\text{b-s}}\widehat{\mathbf{F}}_{t-1}\right)^{\mathsf{T}}$ and $\mathbf{D} = \widehat{\alpha}_{t-1}\mathbf{H}_{\text{b-u}}^{\mathsf{H}}\widehat{\mathbf{F}}_{t-1}$

    Compute $\widehat{\boldsymbol{\Upsilon}}_{t-1} = \text{Diag}(\widehat{\boldsymbol{v}}_{t-1})$

    Compute

$$\nabla\widehat{\boldsymbol{\Upsilon}}_{t-1} = 2\boldsymbol{\Upsilon}_{t-1}^{-\mathsf{T}}(\boldsymbol{\Upsilon}_{t-1}^{-1} - \mathbf{S})^{-\mathsf{T}}(\mathbf{A})^{\mathsf{T}}(\mathbf{A}(\boldsymbol{\Upsilon}_{t-1}^{-1} - \mathbf{S})^{-1}\mathbf{B} + \mathbf{D} - \mathbf{I}_K)^{\mathsf{H}}(\mathbf{B})^{\mathsf{T}}(\boldsymbol{\Upsilon}_{t-1}^{-1} - \mathbf{S})^{-\mathsf{T}}\boldsymbol{\Upsilon}_{t-1}^{-\mathsf{T}}$$

    Compute $\nabla\widehat{\boldsymbol{\Upsilon}}_{t-1} = \text{Diag}\big(\nabla\widehat{\boldsymbol{\Upsilon}}_{t-1}\big)$

    Compute $\widehat{\boldsymbol{v}}_t = \widehat{\boldsymbol{v}}_{t-1} - \mu\nabla\widehat{\boldsymbol{\Upsilon}}_{t-1}^{*}$

    Compute $\widehat{\boldsymbol{v}}_t < \frac{\widehat{\boldsymbol{v}}_t}{|\widehat{\boldsymbol{v}}_t|}$

    Compute $\widehat{\boldsymbol{v}}_t < \frac{\widehat{\boldsymbol{v}}_t}{\sqrt{\boldsymbol{v}_{loss}}}$

    Compute $\widehat{\mathbf{H}}_t = \left(\mathbf{H}_{\text{s-u}}^{\mathsf{H}}(\text{Diag}(\widehat{\boldsymbol{v}}_t)^{-1} - \mathbf{S})^{-1}\mathbf{H}_{\text{b-s}} + \mathbf{H}_{\text{b-u}}^{\mathsf{H}}\right)^{\mathsf{H}}$

    Compute $\widehat{\alpha}_t$

    Compute $\widehat{\mathbf{F}}_t$

until $|E_t - E_{t-1}| < \epsilon E_{t-1}$

return $\widehat{\boldsymbol{v}}_t,\ \widehat{\mathbf{F}}_t,\ \widehat{\alpha}_t$.

Algorithm 2 Joint Optimization of Phase Shift and Linear Receiver Matrices of Uplink Transmission

Inputs: $\mathbf{H}_{\text{s-b}}$, $\mathbf{H}_{\text{u-b}}$, $\mathbf{H}_{\text{u-s}}$, $\mathbf{S}$

At $t \leftarrow 0$, initialize $\widehat{\boldsymbol{\upsilon}}_0$

Compute $\widehat{\mathbf{H}}_0 = \mathbf{H}_{\text{s-b}}(\text{Diag}(\widehat{\boldsymbol{\upsilon}}_0)^{-1} - \mathbf{S})^{-1}\mathbf{H}_{\text{u-s}} + \mathbf{H}_{\text{u-b}}$

Compute $\widehat{\mathbf{G}}_0$

Then set $t = 1$

Initialize first 1:T_1 iterations via the VAMP algorithm

$t = t + 1$

repeat

  // Gradient Descent Algorithm //

  Set $\mathbf{A} = \mathbf{H}_{\text{u-s}}$, $\mathbf{B} = \widehat{\mathbf{G}}_{t-1}\mathbf{H}_{\text{s-b}}$ and $\mathbf{D} = \widehat{\mathbf{G}}_{t-1}\mathbf{H}_{\text{u-b}}$

  Compute $\widehat{\boldsymbol{\Upsilon}}_{t-1} = \text{Diag}(\widehat{\boldsymbol{\upsilon}}_{t-1})$

  Compute

  $\nabla\widehat{\boldsymbol{\Upsilon}}_{t-1} = 2p_k\boldsymbol{\Upsilon}_{t-1}^{-\mathsf{T}}(\boldsymbol{\Upsilon}_{t-1}^{-1} - \mathbf{S})^{-\mathsf{T}}(\mathbf{B})^{\mathsf{T}}(\mathbf{I}_K - \mathbf{B}(\boldsymbol{\Upsilon}_{t-1}^{-1} - \mathbf{S})^{-1}\mathbf{A} - \mathbf{D})^{\mathsf{H}}(\mathbf{A})^{\mathsf{T}}(\boldsymbol{\Upsilon}_{t-1}^{-1} - \mathbf{S})^{-\mathsf{T}}\boldsymbol{\Upsilon}_{t-1}^{-\mathsf{T}}$

  Compute $\nabla\widehat{\boldsymbol{\Upsilon}}_{t-1} = \text{Diag}\big(\nabla\widehat{\boldsymbol{\Upsilon}}_{t-1}\big)$

  Compute $\widehat{\boldsymbol{\upsilon}}_t = \widehat{\boldsymbol{\upsilon}}_{t-1} - \mu\nabla\widehat{\boldsymbol{\Upsilon}}_{t-1}^{*}$

  Compute $\widehat{\boldsymbol{\upsilon}}_t < \frac{\widehat{\boldsymbol{\upsilon}}_t}{|\widehat{\boldsymbol{\upsilon}}_t|}$

  Compute $\widehat{\boldsymbol{\upsilon}}_t < \frac{\widehat{\boldsymbol{\upsilon}}_t}{\sqrt{\boldsymbol{\upsilon}_{loss}}}$

  Compute $\widehat{\mathbf{H}}_t = (\mathbf{H}_{\text{s-b}}(\text{Diag}(\widehat{\boldsymbol{\upsilon}}_t)^{-1} - \mathbf{S})^{-1}\mathbf{H}_{\text{u-s}} + \mathbf{H}_{\text{u-b}})$

  Compute $\widehat{\mathbf{G}}_t$

until $|E_t - E_{t-1}| < \epsilon E_{t-1}$

return $\widehat{\boldsymbol{\upsilon}}_t$, $\widehat{\mathbf{G}}_t$.

$\Phi = (\boldsymbol{\Upsilon}^{-1} - \mathbf{S})^{-1}$, the optimization problem cannot be transformed into a least squares problem due to the presence of two inverse matrix terms. Due to this reason, we do not apply the VAMP algorithm directly to the optimization of the proposed model. Instead, we use the VAMP algorithm only for initializing the solutions of the optimization problem using the conventional RIS model as a first order approximation.

For downlink transmission, we use active precoding optimization and the VAMP algorithm [46] for initializing the phase shift matrix optimization process. For the uplink, we follow a similar initialization procedure for the phase shift matrix optimization but use linear receiver optimization. Note that throughout the initialization, we utilize the conventional RIS model and we execute the VAMP algorithm for a fixed (i.e., 40) number of iterations.

2.3.6 Analysis of Computational Complexity

Consider the overall optimization algorithm (for both downlink and uplink transmissions). The complexity of initializing the $\widehat{\mathbf{H}}_0$ is of $\mathcal{O}(M^3)$ as it involves a matrix inverse computation. Initializing $\widehat{\alpha}_0$ and $\widehat{\mathbf{F}}_0$ each takes $\mathcal{O}(N^3)$ operations. The VAMP algorithm, which is used for initialization, has a complexity of $\mathcal{O}(KN(M+N))$ [46]. The complexity of the gradient descent method is $\mathcal{O}(M^3)$ since this involves matrix inverse computations. Updating $\widehat{\mathbf{H}}_t$, $\widehat{\alpha}_t$, and $\widehat{\mathbf{F}}_t$ in subsequent steps requires complexities of $\mathcal{O}(M^3)$, $\mathcal{O}(N^3)$, and $\mathcal{O}(N^3)$, respectively. Therefore, the overall complexity of the algorithm is $3\mathcal{O}(M^3) + 4\mathcal{O}(N^3) + \mathcal{O}(KN(M+N))$. Note that this represents the worst-case complexity. The dominant factor in this analysis is the calculation of the inverse term. It is possible to reduce the complexity by using for instance Neuman series approximation techniques to simplify the inverse calculation.

2.4 Simulation Results

The parametric channel model [31, 46, 57] is considered for the simulations. Under this model, the channel response of RIS and BS is defined as:

$$\mathbf{H}_{\text{b-s}} \;=\; \sqrt{L(d_{\text{RIS}})}\sum_{q=1}^{Q_{\text{b-s}}} \mathsf{c}_q \mathbf{a}_{\text{RIS}}(\varphi_q,\psi_q)\mathbf{a}_{\text{BS}}(\phi_q)^{\mathsf{T}}. \tag{2.50}$$

Here, the vectors $\mathbf{a}_{\text{RIS}}(\varphi,\psi)$ and $\mathbf{a}_{\text{BS}}(\phi)$ correspond to the array response of RIS and the array response of BS, respectively. The terms $L(d_{\text{RIS}})$ and $Q_{\text{b-s}}$ denote the distance-dependent path-loss factor and the number of channel paths of RIS-BS, respectively. The term $\mathbf{a}_{\text{BS}}(\phi)$ can be given by $\mathbf{a}_{\text{BS}}(\phi) = [1, e^{2\pi \mathrm{j}\frac{d_{\text{b}}}{\lambda}\cos\phi}, \cdots, e^{2\pi \mathrm{j}\frac{d_{\text{b}}}{\lambda}(N-1)\cos\phi}]^{\mathsf{T}}$, and c_q, ϕ, λ, d_{b} stand for the channel path gain, angle of departure (AOD), the wavelength, and the spacing between two consecutive antennas at the BS, respectively. The array response vector for the RIS [46, 64] including the cosine-shaped element pattern is denoted by:

$$\begin{aligned}\mathbf{a}_{\text{RIS}}(\varphi,\psi) = \sqrt{|\cos\varphi|}\begin{bmatrix} 1 \\ e^{2\pi \mathrm{j}\frac{d_{\text{s}}}{\lambda}\sin\varphi\sin\psi} \\ \vdots \\ e^{2\pi \mathrm{j}\frac{d_{\text{s}}}{\lambda}(\sqrt{M}-1)\sin\varphi\sin\psi}\end{bmatrix} \otimes \\ \begin{bmatrix} 1 \\ e^{2\pi \mathrm{j}\frac{d_{\text{s}}}{\lambda}\sin\varphi\cos\psi} \\ \vdots \\ e^{2\pi \mathrm{j}\frac{d_{\text{s}}}{\lambda}(\sqrt{M}-1)\sin\varphi\cos\psi}\end{bmatrix}.\end{aligned} \tag{2.51}$$

The angles φ, ψ represent the angle of elevation, the angle of azimuth, and d_{s} is the distance between two consecutive RIS elements. The angles ψ_q and ϕ_q are uniformly distributed over the interval $[0, 2\pi]$, while the angles φ_q are uniformly distributed over the interval $[0, \pi]$. The channel responses corresponding to the BS-k-th user [46] can be expressed as

$$\mathbf{h}_{\text{b-u},k} \;=\; \sqrt{L(d_k)}\sum_{q=1}^{Q_{\text{b-u}}} \mathsf{c}_{k,q}\mathbf{a}_{\text{BS}}(\phi_{k,q}); \quad k = 1, \cdots, K. \tag{2.52}$$

Table 2.1: Simulation parameters and values

Simulation parameter	**Notation**
Reference distance	$d_0 = 1$ m
RIS-BS distance	$d_{\text{RIS}} = 500$ m
BS-User distance	$d = 500$ m
RIS-User distance	$d' \in [10, 50]$ m
Channel path numbers of RIS-BS	$Q_{\text{b-s}} = 8$
Channel path numbers of BS-user	$Q_{\text{b-u}} = 2$
Channel path numbers of RIS-user	$Q_{\text{s-u}} = 2$
Channel path gain	$c_q \sim \mathcal{CN}(0, 1)$
RIS-BS, RIS-user, path-loss exponent	$\eta = 2.5$
BS-user path-loss exponent	$\eta = 3.7$
Path-loss at the reference distance	$C_0 = -30$ dB
Noise variance	$\sigma_{\text{w}}^2 = -100$ dBm
RIS elements	$M = 256$
BS antennas	$N = 32$
Users	$K = 1, 4, 8$
Transmit power	$P = 0 - 50$ dBm

In this case, the angles $\phi_{k,q}$ are uniformly distributed in the range $[0, 2\pi]$. Here, $\mathsf{c}_{k,q}$ denotes the channel path gains and $Q_{\text{b-u}}$ denotes the channel path numbers of BS-user. The channel vectors represented by RIS-k-th users link [46] are given by

$$\mathbf{h}_{\text{s-u},k} \;=\; \sqrt{L(d'_k)} \sum_{q=1}^{Q_{\text{s-u}}} \mathsf{c}_{k,q} \mathbf{a}_{\text{IRS}}(\varphi_{k,q}, \psi_{k,q}); \quad k = 1, \cdots, K. \tag{2.53}$$

Here $Q_{\text{s-u}}$ denotes the channel path numbers of RIS-user. Furthermore, the distance-dependent path-loss factor denoted by the term $L(d)$, is given by

$$L(d) = C_0(d/d_0)^{\eta}, \tag{2.54}$$

where d_0 is the reference distance and η is the path-loss exponent. Table I shows the basic simulation parameters [46] used for both transmission scenarios. For the simulations, we use MATLAB R2022a on a laptop with Intel Core i7, and Windows 11.

2.4.1 Performance Evaluation

The performance is evaluated based on sum rate ($\widehat{C}$) of both transmissions. The sum-rate for downlink transmission is

$$\widehat{C} = \sum_{k=1}^{K} \log_2 \left(1 + \frac{\left|\mathbf{h}_k^{\mathsf{H}} \mathbf{f}_k\right|^2}{\sigma_{\text{w}}^2 + \sum_{i \neq k} \left|\mathbf{h}_k^{\mathsf{H}} \mathbf{f}_i\right|^2} \right), \tag{2.55}$$

where $\mathbf{h}_k^H = \mathbf{h}_{\text{s-u},k}^{\mathsf{H}} (\boldsymbol{\Upsilon}^{-1} - \mathbf{S})^{-1} \mathbf{H}_{\text{b-s}} + \mathbf{h}_{\text{b-u},k}^{\mathsf{H}}$. And for the uplink transmission

$$\widehat{C} = \sum_{k=1}^{K} \log_2 \left(1 + \frac{|\mathbf{g}_k \mathbf{h}_k|^2 p_k}{\sigma_{\text{w}}^2 \mathbf{g}_k^{\mathsf{T}} \mathbf{g}_k + \sum_{i \neq k} |\mathbf{g}_i \mathbf{h}_k|^2 p_i} \right), \tag{2.56}$$

where $\mathbf{h}_k = \mathbf{H}_{\text{s-b}} (\boldsymbol{\Upsilon}^{-1} - \mathbf{S})^{-1} \mathbf{h}_{k,\text{u-s}} + \mathbf{h}_{k,\text{u-b}}$. We analyze system performance *a)* without phase-shift loss, and *b)* by adding some phase-shift loss to RIS elements. We take inter-antenna spacing at BS (d_{b}) and RIS (d_{s}) as $\lambda/2$. We assume that RIS-BS and RIS-Users paths are in line of sight (LOS) while BS-User direct paths are in non-line of sight (NLOS). For both uplink and downlink transmissions, the results are analyzed considering the following four cases [31]:

i. *Case I*: Conventional RIS model as the system model where both active precoding and phase-shift optimizations are done based on the VAMP algorithm.

ii. *Case II*: Physically-consistent RIS model, but neither active precoding matrix ($\mathbf{F}$) nor phase-shift matrix is optimized. This implies that we employ the physically-consistent model as the system model, but we do not incorporate the physically-consistent model directly into the optimization process. Instead, we perform optimization using the conventional model neglecting the coupling effects.

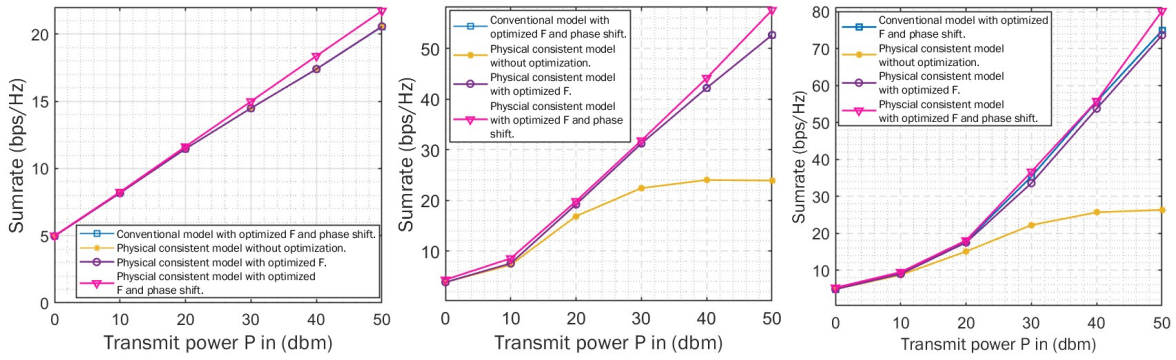
iii. *Case III*: Physically-consistent RIS model where only the active precoding matrix ($\mathbf{F}$) is optimized through the VAMP algorithm.

iv. *Case IV*: Physically-consistent RIS model where the active precoding matrix ($\mathbf{F}$), the linear receiver matrix ($\mathbf{G}$), and the phase-shift matrix are optimized using the proposed method.

Note that, we refer to the model used in [46] as the "conventional model", which does not account for multiple reflections. Specifically, this model is a **hypothetical** one that does not exist in reality. Therefore, as the main baseline for comparison, we consider the case "Physically-consistent RIS model, but neither active precoding matrix ($\mathbf{F}$) nor phase-shift matrix is optimized" denoted by Case II (which is represented by the yellow colored lines in the simulation figures).

2.4.2 Downlink performance without phase-shift loss in RIS elements

Generally, as the transmit power increases, the EM fields generated by the RIS become stronger, resulting in a more pronounced EM interaction among RIS elements. Consequently, mutual coupling also increases. Therefore, the performance of the proposed model along with joint optimization should be enhanced when increasing the transmit power. The variation in sum-rate with transmit power is shown in Fig. 2.3 for a single



(a) $K = 1$ and $M = 256$. (b) $K = 4$ and $M = 256$. (c) $K = 8$ and $M = 256$.

Figure 2.3: Sum-rate versus transmit power with perfect CSI for lossless condition: (a) single user, (b) four users, (c) eight users.

user, four users, and eight users scenarios considering *perfect CSI*. For all these cases, the sum-rate corresponding to optimized $\mathbf{F}$ and optimized phase-shift with the physically-consistent model is always higher than the other three cases. For the single-user case, we can observe that the three lines corresponding to the first three cases are almost overlapping and the highest sum-rate value is less than that observed in the cases involving four and eight users. The increase in sum-rate is lower for the four-user case when compared to the eight-user case highlighting the higher impact of mutual coupling in the multiuser setting. For Case II, the sum-rate saturates with transmit power in both multi-user scenarios. This saturation is primarily due to the residual interference effects when multiple reflections are not taken into account in the optimization. When the RIS elements reflect signals towards multiple users, inter-user interference arises due to multiple reflections and non-optimized beamforming. Moreover, as the number of users increases, interference also increases. The interference arising from multiple reflections can be effectively mitigated by precisely optimizing both active and passive beamforming jointly. On the other hand, as expressed in (2.2), there is no multiple-reflection effect in conventional models given by the first-order approximation. Therefore, the

multi-user systems with conventional RIS models will not experience this interference.

We conduct additional simulations by modifying some specific simulation parameters. In this case, we consider path-loss at the reference distance $C_0 = -20$ dB, noise variance $\sigma_{\mathrm{w}}^2 = -50$ dBm, instead of the former parameters with perfect CSI. Figure 2.4 is plotted based on latter simulation parameters. It is important to highlight that the SNR decreases with the latter set of parameters. According to Figure 2.4, the gap between the proposed model and the conventional model (hypothetical model) is higher compared to Fig. 2.3(c) with the former parameters due to the logarithmic behavior of the capacity with respect to SNR. However, the interference effect denoted by the yellow curve becomes less dominant with low SNR. On the other hand, when the SNR is high (with former parameters), the gap between the proposed model and the hypothetical model is less, but the interference effect becomes very critical with high SNR, and ignoring the multiple reflections effect is detrimental. There is a considerable amount of

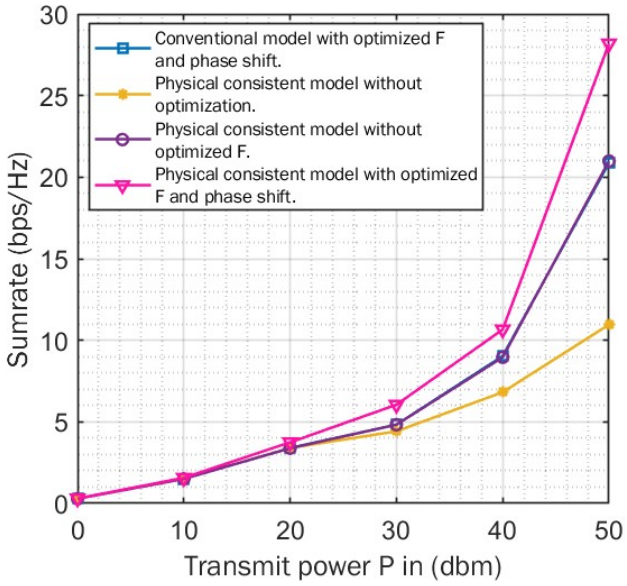
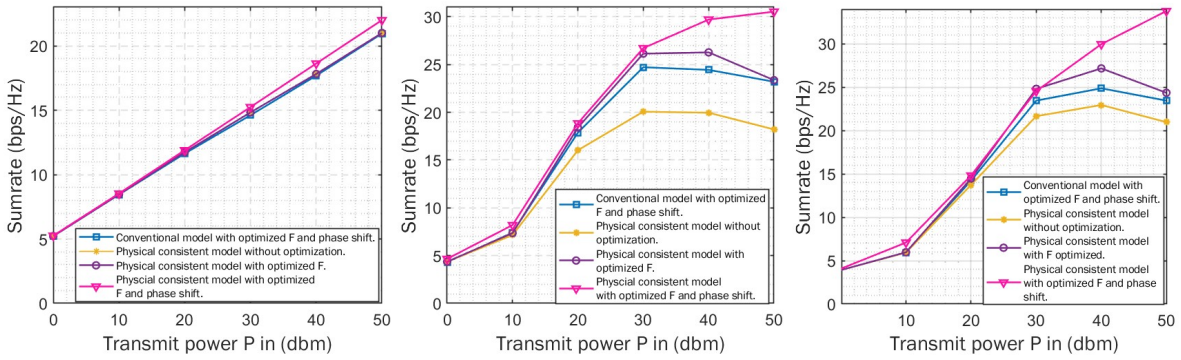


Figure 2.4: Sum-rate versus transmit power with perfect CSI for $K = 8$, $M = 256$: $C_0 = -20$ dB, $\sigma_{\mathrm{w}}^2 = -50$ dBm.

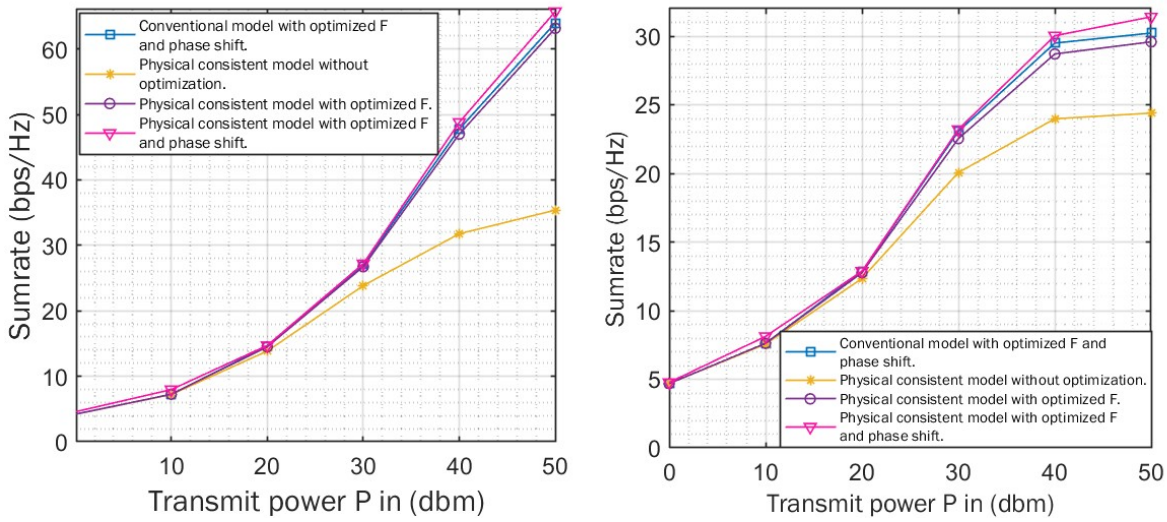


(a) $K = 1$ and $M = 256$. (b) $K = 4$ and $M = 256$. (c) $K = 8$ and $M = 256$.

Figure 2.5: Sum-rate versus transmit power with imperfect CSI for lossless condition: (a) single user, (b) four users, (c) eight users.

improvement in the sum-rate compared with our main baseline when the SNR is high. Similarly, even at low SNR levels, there is a considerable amount of improvement even compared to the hypothetical (unrealistic) performance.

In the context of *imperfect CSI*, as shown in Fig. 2.5, the trend for the single-user case is similar to that of the perfect CSI case. However, it is important to note that the sum-rate values are smaller with imperfect CSI compared to the perfect CSI scenario. Furthermore, in the multi-user scenario, the overall sum-rate values are lower compared to the equivalent multi-user scenario with perfect CSI. The interference from multiple reflections, which plays a substantial role in perfect CSI scenarios, becomes less dominant in imperfect CSI situations as there is already interference present due to imperfectness. This is the reason for observing a smaller gap between the proposed model and the main baseline given by Case II, compared to perfect CSI. However, compared to a perfect CSI scenario, the gap between the proposed model and the conventional model becomes more pronounced with imperfect CSI. In the presence of imperfect CSI, there are residual errors that can be treated as noise. Consequently, there is an additional noise component attributed to imperfect CSI, in addition to the inherent



(a) $K = 8$ and $M = 256$.

(b) $K = 8$ and $M = 256$.

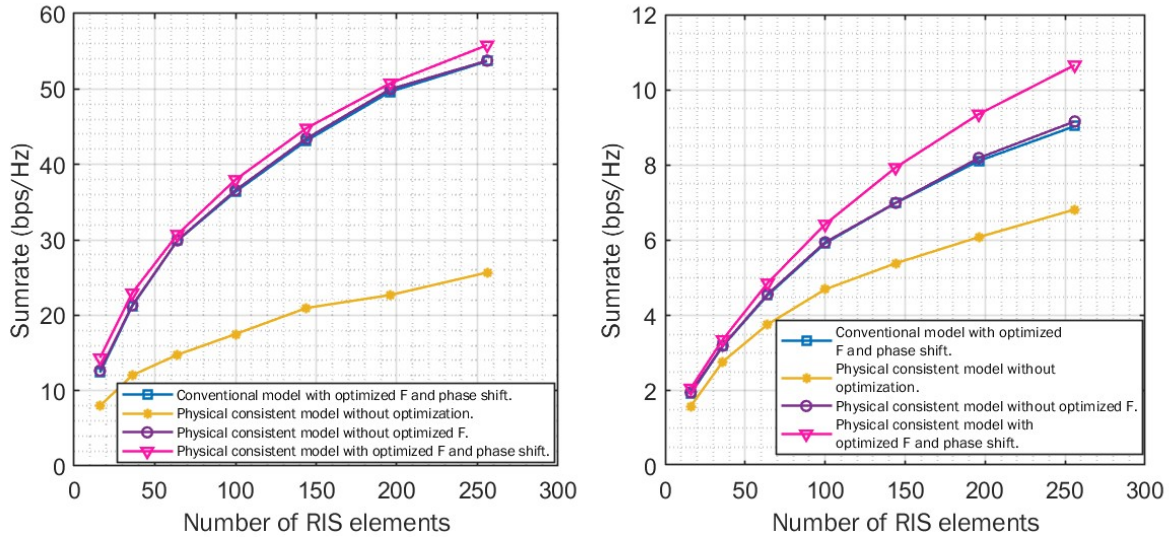
Figure 2.6: Sum-rate versus transmit power with 3dB phase-shift loss: (a) perfect CSI, (b) imperfect CSI.

noise from the signal. This results in a lower SNR. As explained earlier, when the SNR is low, we observe a considerable amount of the gap between the proposed model and the conventional model. Nevertheless, it is obvious that the physically-consistent RIS model with optimized **F** and phase shifts outperforms the other three cases even when the CSI is imperfect.

2.4.3 Downlink performance with phase-shift loss in RIS elements

Figures 2.6(a) and 2.6(b) illustrate the impact of 3dB phase-shift loss on RIS for eight users under both perfect and imperfect CSI conditions. Similar to the no-loss case, the sum-rate for the physically-consistent RIS model with optimized **F** and phase-shift outperforms the other three cases. It can be observed from Case II that multiple reflection effects are less effective when the phase shifters are lossy since they reduce

the dominance of interference caused by multiple reflections. In other words, when the RIS is lossy, multiple reflections become weak. Therefore, the sum-rate improvement due to joint optimization does not show any significant improvement compared to the lossless case. Moreover, the sum-rate improvement is even less pronounced when the CSI is imperfect. Nevertheless, the lossy phase shift scenario also verifies that the overall performance of the proposed method is better than the other three cases.



(a) $C_0 = -30$ dB, $\sigma_w^2 = -100$ dBm. (b) $C_0 = -20$ dB, $\sigma_w^2 = -50$ dBm.

Figure 2.7: Sum-rate versus number of RIS elements for $K = 8$, $P = 40$ dbm, $\lambda/2$: (a) previous simulation parameters, (b) latter simulation parameters.

2.4.4 Performance evaluation with varying number of RIS elements

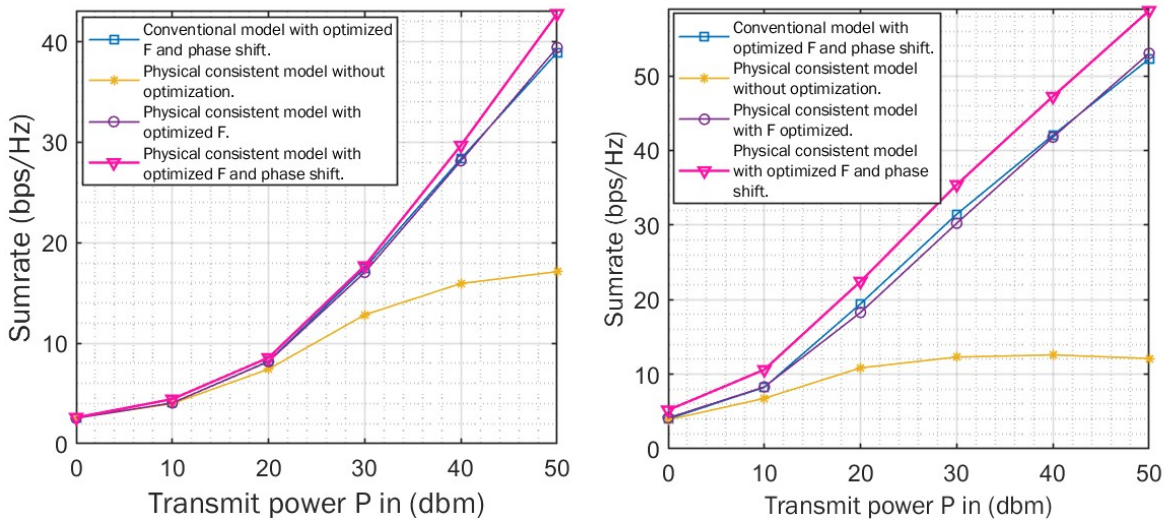
The number of RIS elements is a critical factor that influences mutual coupling. Generally, as the number of RIS elements increases, the mutual coupling effect also increases, leading to a rise in the multiple reflection effect. When there are more elements, they

are typically placed closer to each other, resulting in a strong EM interaction among the elements. This strong interaction increases the mutual coupling effect. Therefore, the proposed model should exhibit some performance improvement along with joint optimization when the number of RIS elements increases. To verify this, we conduct some simulations based on downlink transmission by adjusting the number of RIS elements. These simulations are conducted using both previous and latter simulation parameters as shown in Fig. 2.7. It can be observed that the sum-rate associated with the proposed scheme (with joint optimization) consistently surpasses that of the other three cases, as expected. Furthermore, the interference effect due to multiple reflections has more impact with the increasing number of RIS elements. Most importantly, with the former parameters, at high SNR levels, this interference effect becomes more critical as shown in Fig. 2.7(a). In this case, we can see a significant amount of improvement compared to the main baseline (Case II). On the other hand, at low SNR, the interference effect is less critical compared to high SNR, but a significant gap is observed between the proposed model and the conventional model (hypothetical model).

2.4.5 Performance evaluation with different element spacing

We illustrate how the sum-rate changes with RIS element spacing under different input power conditions. We use antenna spacings of $\lambda/16$, $\lambda/8$, $\lambda/4$, and $\lambda/2$. The corresponding reflective elements are set as $M = 4096$, 1024, 256, and 64, respectively. Note that, increasing the spacing between RIS elements reduces the number of reflective elements that can fit within the specified surface area of the RIS.

Figure 2.8 shows the variation of sum-rate with power for $\lambda/2$ and $\lambda/4$ element spacings. Except for Case II, the sum-rate shows an improvement as the antenna spacing decreases across the other three scenarios. Reduced spacing allows for more reflective elements, and the increased mutual coupling among elements leads to a higher



(a) $K = 4$ and $M = 64$. (b) $K = 4$ and $M = 256$.

Figure 2.8: Sum-rate versus transmit power for four users considering perfect CSI with the change of antenna spacing: (a) $\lambda/2$, (b) $\lambda/4$.

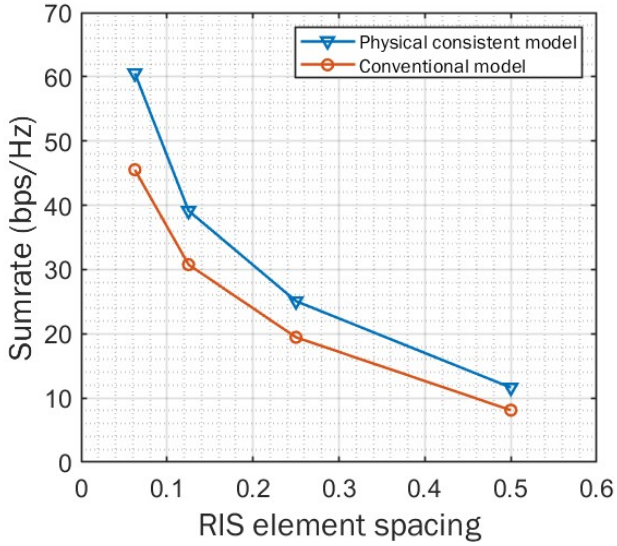
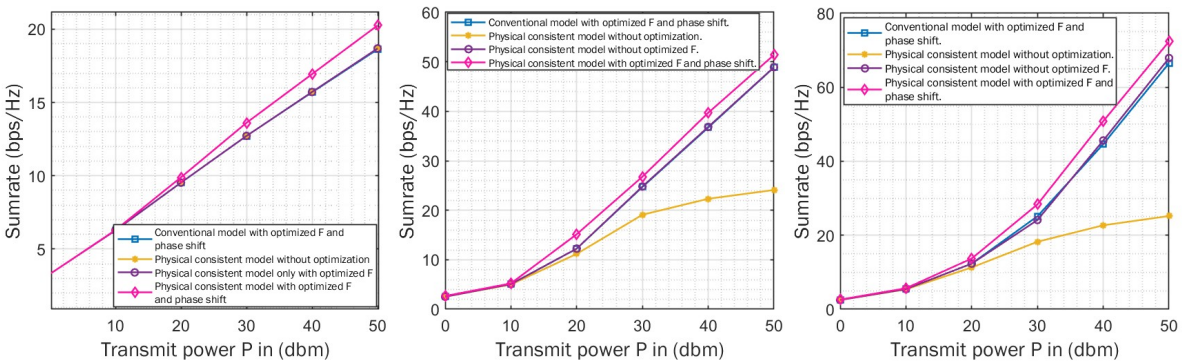


Figure 2.9: Sum-rate versus antenna spacing of RIS for element spacing $\lambda/16$, $\lambda/8$, $\lambda/4$, and $\lambda/2$: $K = 4$, $P = 50$dBm, $C_0 = -20$ dB, $\sigma_{\mathrm{w}}^2 = -50$ dBm.



(a) $K = 1$ and $M = 256$. (b) $K = 4$ and $M = 256$. (c) $K = 8$ and $M = 256$.

Figure 2.10: Sum-rate versus transmit power of uplink transmission with perfect CSI for lossless condition: (a) single user, (b) four users, and (c) eight users.

occurrence of multiple reflections. Therefore, the interference effect due to multiple reflections becomes more critical when reducing the spacing, as shown in Case II of the figure. This issue can be addressed and performance can be improved by jointly optimizing the beamforming process. Furthermore, Fig. 2.9 shows the variation of sum-rate with different element spacing. For this simulation, we utilize the path-loss at the reference distance $C_0 = -20$ dB, the noise variance $\sigma_{\mathrm{w}}^2 = -50$ dBm, instead of equivalent basic previous parameters. It is noticeable that modifying the antenna spacing can significantly impact the sum-rate in the physically-consistent model. When the RIS elements are placed in close proximity, optimization can lead to improved performance.

2.4.6 Performance of uplink transmission

Figure 2.10 shows the variation of the sum-rate corresponding to the uplink transmission against transmit power with perfect CSI. In this context, we employ an antenna spacing of $\lambda/2$ without any phase shift losses. Similar to the downlink case, joint optimizing active and passive beamforming with the physically-consistent RIS model outperforms

the other cases. However, compared to the downlink scenario, the overall sum-rate values are lower. As there is a separate power allocation constraint for the downlink, power for each user can be optimized accordingly. Compared to the downlink scenario, each user in the uplink transmission has fixed power. This is the reason for having lower sum-rate values for the uplink compared to the downlink. Nevertheless, the interference effect due to the multiple reflections can also be visible and it can be mitigated by joint optimization.

2.4.7 Validation of the simulation results

To validate our results, we have employed the Hooke-Jeeves algorithm to optimize the phase shifters considering downlink transmission. For the active precoding optimization, we utilized the same Lagrange optimization method as described in previous sections. However, it is important to note that this algorithm exhibits slower performance, particularly with higher numbers of RIS elements. Hence, we have limited its usage to fewer RIS elements. Our primary objective is to use this algorithm solely for validating the simulation results and confirming the performance achieved by our proposed approach, which utilizes the gradient descent method.

Hooke-Jeeves method

The Hooke-Jeeves method is a heuristic approach that operates without requiring gradient information of the objective function. Instead, it relies on a combination of exploratory moves to identify potential directions and pattern moves to travel along those directions. In this method, orthogonal unit vectors supplied by the user serve as coordinate directions. Given that our problem involves complex vectors and matrices, we utilized the Gram-Schmidt algorithm to find random orthogonal vectors. This algorithm enables the generation of a unitary matrix, which is equivalent to an orthonormal

matrix in the complex domain. The columns of this matrix are utilized as coordinate directions, effectively serving as search directions within the optimization process [63].

Gram-Schmidt Algorithm

This method is utilized for orthonormalizing linearly independent sets of vectors [65]. The steps for orthonormalizing is as follows:

- Given the vectors $\mathbf{v}_1$ and $\mathbf{v}_2$ are linearly independent vectors in $\mathbb{C}^n$. Constructing the vector $\mathbf{q}_1 = \frac{\mathbf{v}_1}{\|\mathbf{v}_1\|}$ such that $\|\mathbf{q}_1\|_2 = 1$.
- Then constructing a vector $\mathbf{u}_2 = \mathbf{v}_2 - (\mathbf{v}_2, \mathbf{q}_1)\mathbf{q}_1$ such that $(\mathbf{u}_2, \mathbf{q}_1) = 0$, which means $\mathbf{u}_2$ is orthogonal to $\mathbf{q}_1$.
- Then normalizing $\mathbf{u}_2$, $\mathbf{q}_2 = \frac{\mathbf{u}_2}{\|\mathbf{u}_2\|}$. Then the vectors $\mathbf{q}_2$ and $\mathbf{q}_1$ are now orthonormal.

It was employed for the QR decomposition of the provided complex random matrix. In this process, the resulting matrix Q represents a unitary matrix comprising orthogonal vectors, while the matrix R is an upper triangular matrix. The columns of the resultant matrix Q were subsequently utilized as coordinate directions for the Hooke-Jeeves algorithm.

Obtained results from Hooke-Jeeves method

We utilized the conventional (hypothetical) model, which assumes no mutual coupling, as the benchmark for comparing the results. For implementation, we utilized the same modeling parameters as outlined in table 2.1 and used the parametric channel modeling. Note that our implementation considered a single BS with 8 antennas, a single RIS consisting of 25 elements, and $\lambda/2$ element spacing. The results were obtained by varying the number of users and adjusting the transmit power accordingly. Figure 2.11 shows the variation of sum rate with transmit power. In this scenario, we can observe similar

observations as with the proposed solution approach. Figures 2.12 and 2.13 represent the variation in sum rate with power for two and four users, respectively. Once again, we observe an improvement compared to the conventional model. It is important to note that the gap between the physically consistent model and the conventional model can be further enhanced (increased) by increasing the number of RIS elements and reducing the antenna element spacing, as described in the previous subsections. As this algorithm is very slow and gives a poor performance with a higher number of RIS elements and less element spacing, we did not consider them. Therefore, we can conclude that the simulation results obtained from our proposed optimization scheme are validated by the utilization of the Hooke-Jeeves method.

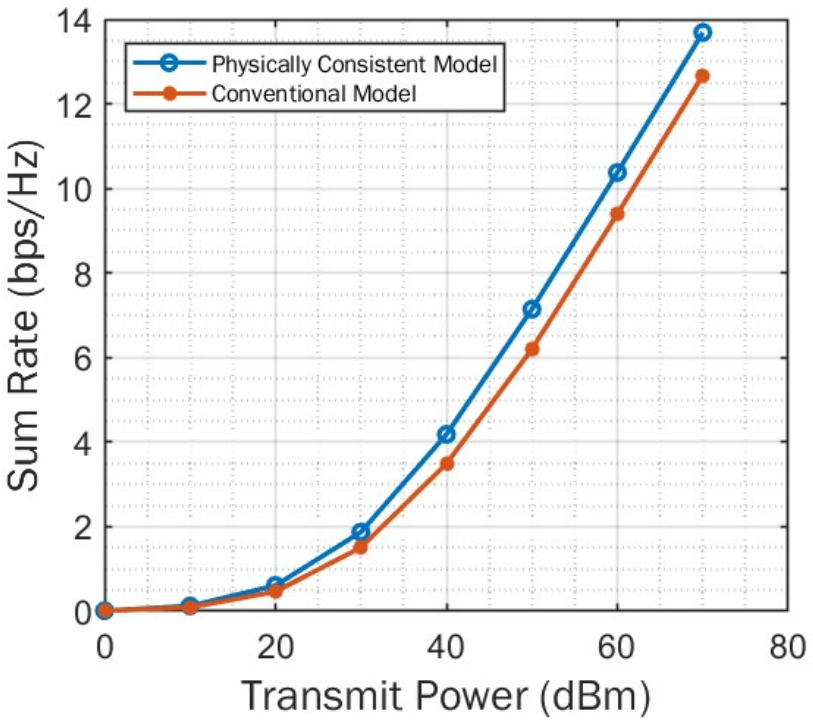


Figure 2.11: Sum-rate performance versus transmit power for single user $K = 1$, $N = 8$, $M = 25$.

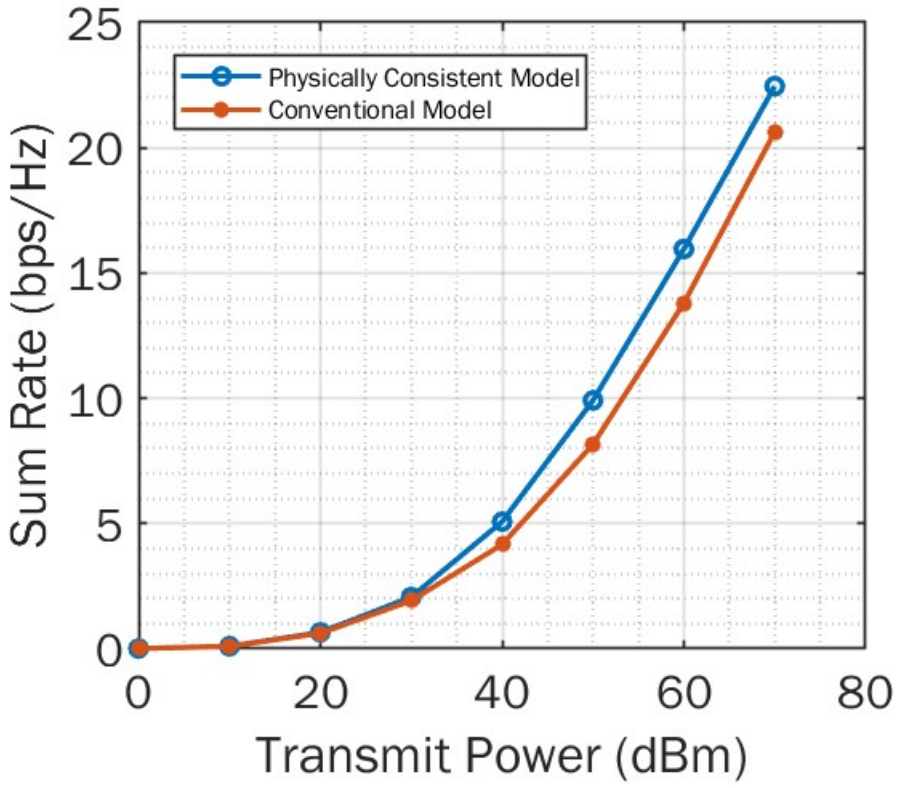


Figure 2.12: Sum-rate performance versus transmit power for two users $K = 2$, $N = 8$, $M = 25$.

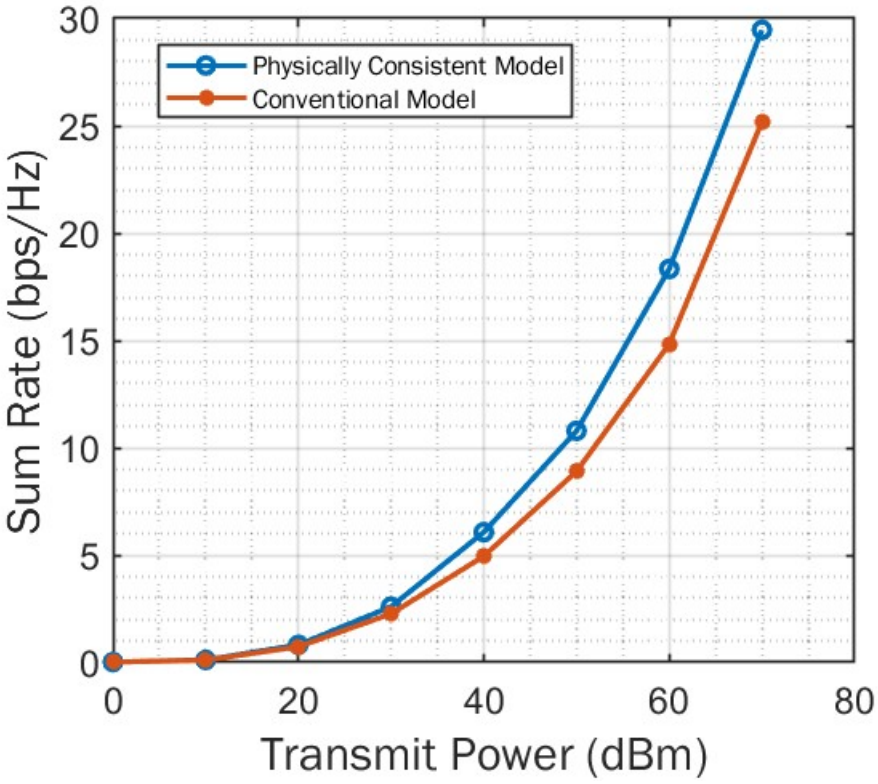


Figure 2.13: Sum-rate performance versus transmit power for four users $K = 4$, $N = 8$, $M = 25$.

Chapter 3

Physically-Consistent Modeling and Optimization of Non-local RIS-Assisted Multi-User MIMO Communication Systems

3.1 System Model and Assumptions

3.1.1 Physically-Consistent RIS Modeling of Reflective RIS

The EM properties of a RIS are explained by its transmit/receive space- and port-side scattering patterns [36]. According to Fig. 3.1, the radiation pattern of the space side is represented by the incoming and outgoing wave phasor vectors $\boldsymbol{a}_{\boldsymbol{\beta}}(\theta, \varphi)$ and $\boldsymbol{b}_{\boldsymbol{\beta}}(\theta, \varphi)$, respectively, with respect to the azimuth and elevation angles θ and φ on the horizontal and vertical axes, respectively. Moreover, the forward and backward traveling wave phasors of each element corresponding to the port-side scattering are denoted by $a_{\alpha,m}$ and $b_{\alpha,m}$, respectively, where m (with $m = 1, 2, \ldots, M$) denotes the RIS element

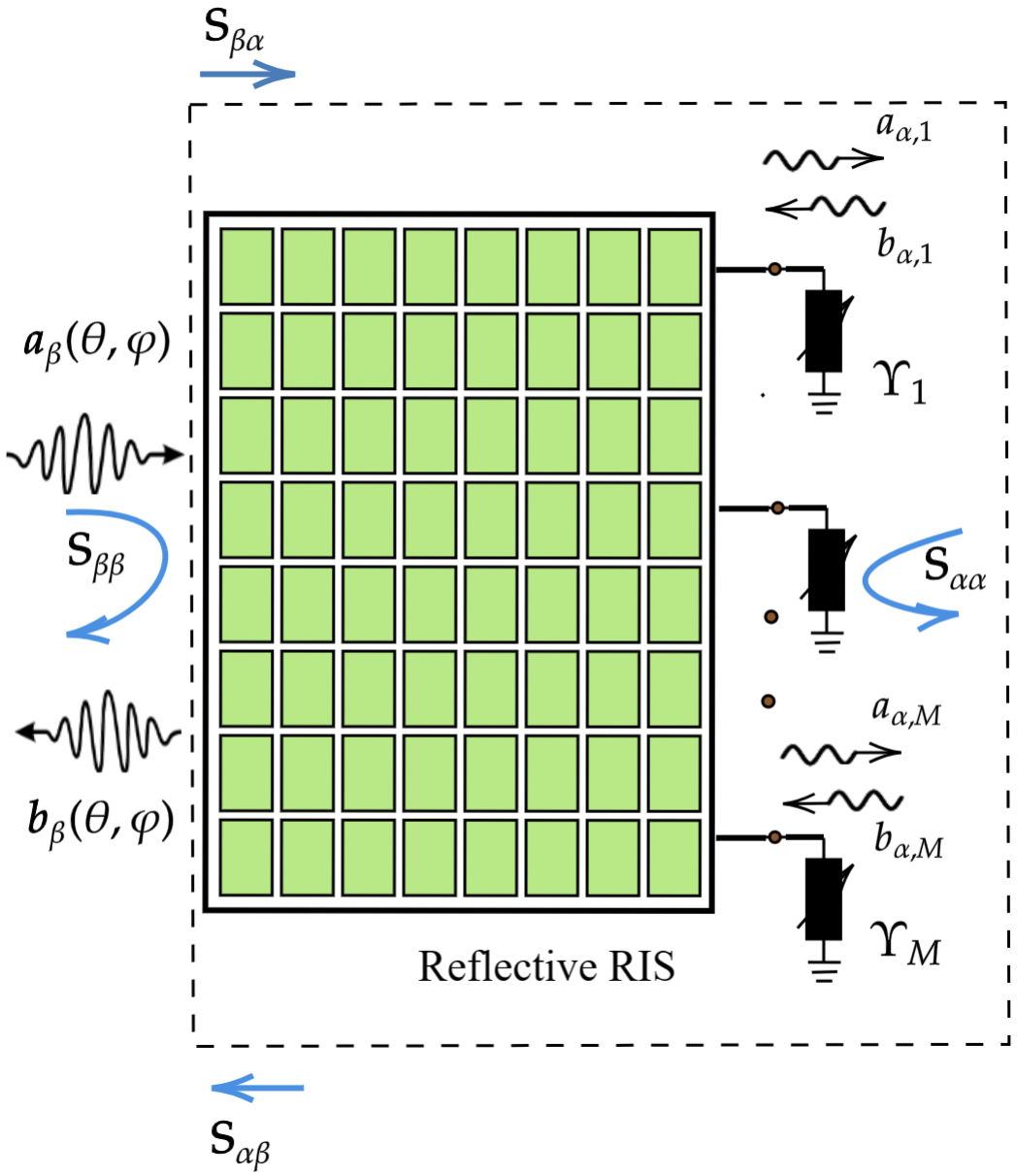


Figure 3.1: Transmit/receive space- and port-side patterns of a reflective RIS.

index. The mathematical representations of space- and port side-scattering are related

as follows:

$$\underbrace{\begin{bmatrix} \boldsymbol{b_\beta}(\theta,\varphi) \\ \boldsymbol{b_\alpha} \end{bmatrix}}_{\triangleq \boldsymbol{b}} = \underbrace{\begin{bmatrix} \mathbf{S}_{\beta\beta} & \mathbf{S}_{\alpha\beta} \\ \mathbf{S}_{\beta\alpha} & \mathbf{S}_{\alpha\alpha} \end{bmatrix}}_{\triangleq \mathbf{S}} \underbrace{\begin{bmatrix} \boldsymbol{a_\beta}(\theta,\varphi) \\ \boldsymbol{a_\alpha} \end{bmatrix}}_{\triangleq \boldsymbol{a}}. \tag{3.1}$$

In this expression, the matrix $\mathbf{S}$ denotes the total scattering matrix of the RIS. $\mathbf{S}_{\alpha\alpha}$ and $\mathbf{S}_{\beta\beta}$ are called "multi-port" and "wave" scattering matrices, respectively. In addition, the matrices $\mathbf{S}_{\alpha\beta}$ and $\mathbf{S}_{\beta\alpha}$ imply transmit and receive radiation patterns indicating MC. The terms $\boldsymbol{a_\alpha} \triangleq [a_{\alpha,1}, a_{\alpha,2}, ..., a_{\alpha,M}]^\mathsf{T}$ and $\boldsymbol{b_\alpha} \triangleq [b_{\alpha,1}, b_{\alpha,2}, ..., b_{\alpha,M}]^\mathsf{T}$ include respectively the incoming and outgoing wave vectors at all RIS port sides.

In RIS deployments, each port side is terminated by a load. The interaction between incoming and outgoing signals at each port side is characterized by the "load" scattering matrix $\boldsymbol{\Upsilon}$, which is represented as follows:

$$\boldsymbol{b_\alpha} = \boldsymbol{\Upsilon} \boldsymbol{a_\alpha}. \tag{3.2}$$

Expanding expression (3.1) yields the following equations:

$$\boldsymbol{b_\beta}(\theta,\varphi) = \mathbf{S}_{\beta\beta}\boldsymbol{a_\beta}(\theta,\varphi) + \mathbf{S}_{\alpha\beta}\boldsymbol{a_\alpha}, \tag{3.3}$$

$$\boldsymbol{b_\alpha} = \mathbf{S}_{\beta\alpha}\boldsymbol{a_\beta}(\theta,\varphi) + \mathbf{S}_{\alpha\alpha}\boldsymbol{a_\alpha}. \tag{3.4}$$

By utilizing (3.2), (3.3), and (3.4), the total outgoing wave phasor $\boldsymbol{b_\beta}(\theta,\varphi)$ at the space-side scattering pattern is given by

$$\boldsymbol{b_\beta}(\theta,\varphi) = \underbrace{\mathbf{S}_{\beta\alpha}\left(\boldsymbol{\Upsilon}^{-1} - \mathbf{S}_{\alpha\alpha}\right)^{-1}\mathbf{S}_{\alpha\beta}\boldsymbol{a_\beta}(\theta,\varphi)}_{\text{Adaptive scattering}} + \underbrace{\mathbf{S}_{\beta\beta}\boldsymbol{a_\beta}(\theta,\varphi)}_{\text{Residual scattering}}. \tag{3.5}$$

The term $\left(\boldsymbol{\Upsilon}^{-1} - \mathbf{S}_{\alpha\alpha}\right)^{-1}$ can be expanded as a summation of higher order terms, as shown in the following expression known as "Neumann series approximation" [31, 66]:

$$(\boldsymbol{\Upsilon}^{-1} - \mathbf{S}_{\alpha\alpha})^{-1} = \sum_{l=0}^{\infty}(\boldsymbol{\Upsilon}\mathbf{S}_{\alpha\alpha})^l\boldsymbol{\Upsilon} = \boldsymbol{\Upsilon} + \boldsymbol{\Upsilon}\mathbf{S}_{\alpha\alpha}\boldsymbol{\Upsilon} + \cdots. \tag{3.6}$$

Conventional models utilize a simplified version of (3.5) that ignores residual scattering and assume that $\mathbf{S}_{\alpha\beta} = \mathbf{S}_{\beta\alpha} = \mathbf{I}_M$ and absence of multiple reflections, i.e., $\mathbf{S}_{\alpha\alpha} = \mathbf{0}_M$. In

this Chapter, we incorporate MC and multiple reflections [30, 31, 66], and optimize the associated scattering parameters $\mathbf{S}_{\alpha\beta}$, $\mathbf{S}_{\beta\alpha}$, and $\mathbf{S}_{\alpha\alpha}$ for a communications' objective. For simplicity, we also ignore the residual scattering part in (3.5).

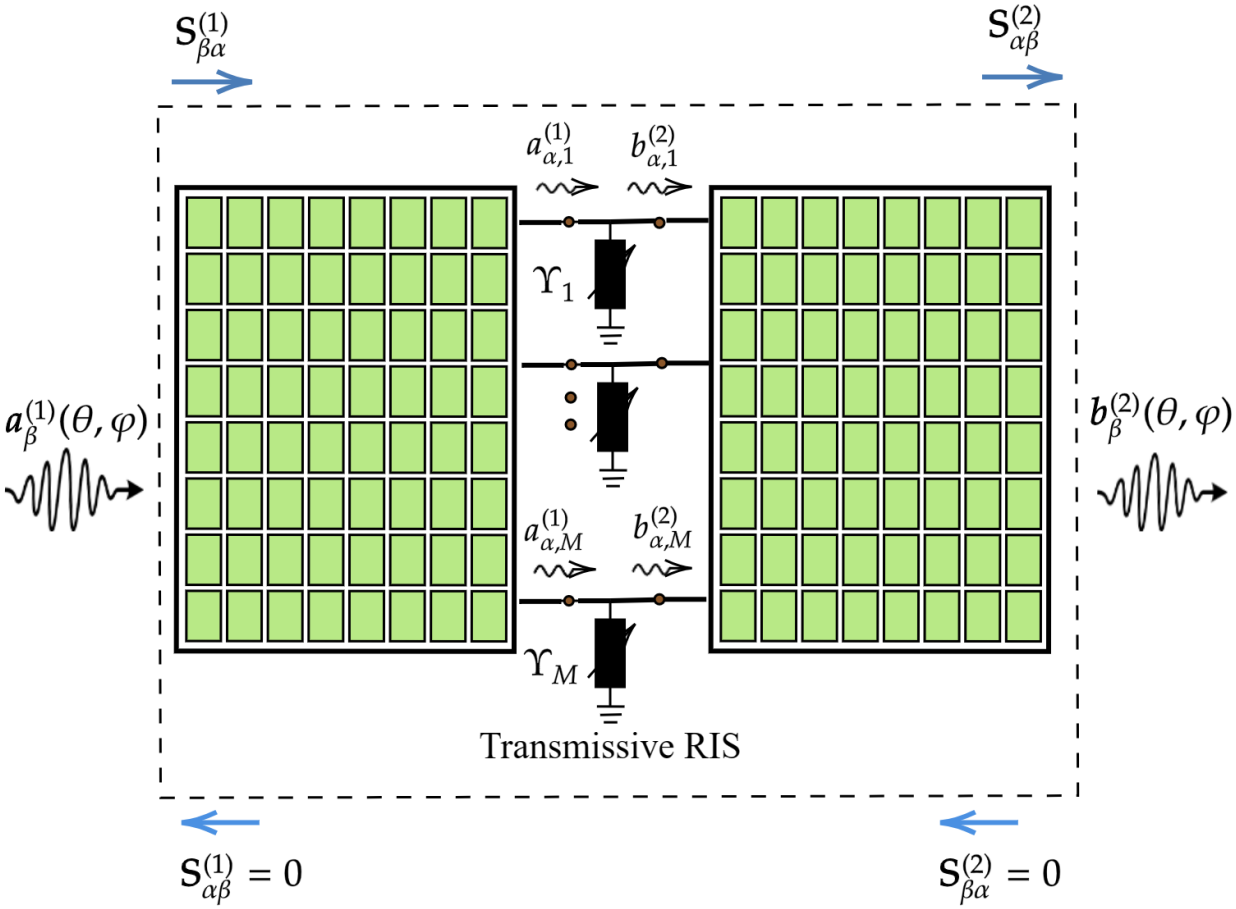


Figure 3.2: Transmit/receive space- and port-side patterns of a transmissive RIS.

3.1.2 Physically-Consistent RIS Modeling of Transmissive RIS

Figure 3.2 displays the space-port side pattern of a transmissive RIS. It contains two-element arrays connected by phase shifters, with separate space and port sides per array.

These port sides are interconnected to enable the functionality of the transmissive RIS. In other words, the transmissive RIS is modeled by arranging two reflective RISs in a back-to-back configuration. Note that, this represents just one possible approach to modeling transmissive RIS. For simplicity of the analysis, we assumed that the RIS is fully transmitting with zero reflection. The space side incoming radiation pattern is denoted by $\boldsymbol{a}_{\boldsymbol{\beta}}^{(\mathbf{1})}(\theta,\varphi)$ while outgoing radiation pattern is $\boldsymbol{b}_{\boldsymbol{\beta}}^{(\mathbf{2})}(\theta,\varphi)$. The forward traveling wave phasor of each element, denoted by $a_{\alpha,m}^{(1)}$, corresponds to the port-side scattering of the first side. Here, the backward traveling wave phasor is zero ($b_{\alpha,m}^{(1)}=0$), indicating no reflection. Conversely, for the second side, the forward traveling wave phasor towards the port side is zero ($a_{\alpha,m}^{(2)}=0$), while the backward traveling wave phasor from the port side $b_{\alpha,m}^{(2)}$ remains. In our ideal transmission scenario, we assume no multiple reflections, resulting in zero multiport scattering matrices for both sides, i.e, $\mathbf{S}_{\alpha\alpha}^{(\mathbf{1})}=\mathbf{S}_{\alpha\alpha}^{(\mathbf{2})}=\mathbf{0}_M$. Since the reflection towards the incident side is zero, wave scattering matrices (reflections of space side) are also zero, i.e., $\mathbf{S}_{\beta\beta}^{(\mathbf{1})}=\mathbf{S}_{\beta\beta}^{(\mathbf{2})}=\mathbf{0}_M$. In addition, as no signals return to incident directions, matrices $\mathbf{S}_{\alpha\beta}^{(\mathbf{1})}=\mathbf{S}_{\beta\alpha}^{(\mathbf{2})}=\mathbf{0}_M$. The equations below represent the space-port sides relationships of the transmissive RIS.

$$\boldsymbol{b}_{\boldsymbol{\beta}}^{(\mathbf{2})}(\theta,\varphi)=\mathbf{S}_{\alpha\beta}^{(\mathbf{2})}\boldsymbol{b}_{\boldsymbol{\alpha}}^{(\mathbf{2})}, \tag{3.7}$$

$$\boldsymbol{a}_{\boldsymbol{\alpha}}^{(\mathbf{1})}=\mathbf{S}_{\beta\alpha}^{(\mathbf{1})}\boldsymbol{a}_{\boldsymbol{\beta}}^{(\mathbf{1})}(\theta,\varphi), \tag{3.8}$$

$$\boldsymbol{b}_{\boldsymbol{\alpha}}^{(\mathbf{2})}=\boldsymbol{\Upsilon}\boldsymbol{a}_{\boldsymbol{\alpha}}^{(\mathbf{1})}. \tag{3.9}$$

Here the vectors $\boldsymbol{a}_{\boldsymbol{\alpha}}^{(\mathbf{1})}$ and $\boldsymbol{b}_{\boldsymbol{\alpha}}^{(\mathbf{2})}$ are denoted by $\boldsymbol{a}_{\boldsymbol{\alpha}}^{(\mathbf{1})}\triangleq[a_{\alpha,1}^{(1)},a_{\alpha,2}^{(1)},...,a_{\alpha,M}^{(1)}]^{\mathsf{T}}$ and $\boldsymbol{b}_{\boldsymbol{\alpha}}^{(\mathbf{2})}\triangleq[b_{\alpha,1}^{(2)},b_{\alpha,2}^{(2)},...,b_{\alpha,M}^{(2)}]^{\mathsf{T}}$ respectively. By combining the above equations (3.7), (3.8), and (3.9), the outgoing wave phasor $\boldsymbol{b}_{\boldsymbol{\beta}}^{(\mathbf{2})}(\theta,\varphi)$ is

$$\boldsymbol{b}_{\boldsymbol{\beta}}^{(\mathbf{2})}(\theta,\varphi)=\mathbf{S}_{\alpha\beta}^{(\mathbf{2})}\boldsymbol{\Upsilon}\mathbf{S}_{\beta\alpha}^{(\mathbf{1})}\boldsymbol{a}_{\boldsymbol{\beta}}^{(\mathbf{1})}(\theta,\varphi). \tag{3.10}$$

Hence, in the transmissive RIS scenario, we optimize the scattering parameters $\mathbf{S}_{\alpha\beta}^{(\mathbf{2})}$ and $\mathbf{S}_{\beta\alpha}^{(\mathbf{1})}$, which were traditionally assumed to be identity matrices.

3.1.3 System Model

Consider a multi-user downlink MIMO system comprising a single BS equipped with N antennas, $K < N$ single-antenna users, and a single RIS with $M > K$ elements. The direct channel between the BS and any of the users is assumed to be blocked. The matrix $\mathbf{H}_{\text{r-u}} \triangleq [\mathbf{h}_{\text{r-u},1}, \mathbf{h}_{\text{r-u},2}, \ldots, \mathbf{h}_{\text{r-u},K}]$ includes all the RIS-user channel gains, with $\mathbf{h}_{\text{r-u},k} \in \mathbb{C}^M$ representing each RIS-k-th user link ($k = 1, 2, \ldots, K$), while the matrix $\mathbf{H}_{\text{b-r}} \in \mathbb{C}^{M \times N}$ corresponds to the BS-RIS channel gains and satisfies the condition: $\text{Rank}(\mathbf{H}_{\text{b-r}}) \geq K$.

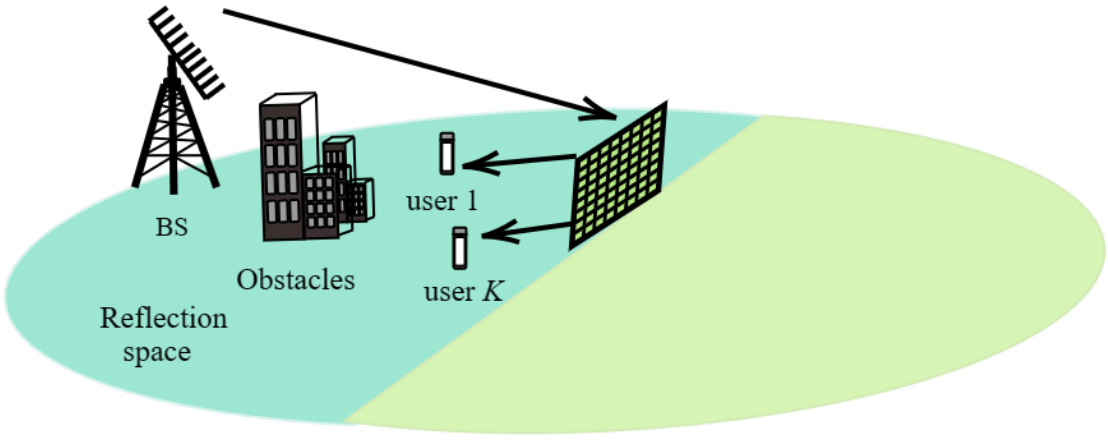


Figure 3.3: The considered RIS-aided multi-user MIMO system model for reflective RIS.

System model for reflective RIS setup

Figure 3.3 shows the system model of the reflective RIS setup. It follows from the previous subsection that the effective RIS phase shift matrix is given by $\mathbf{\Phi} \triangleq (\mathbf{\Upsilon}^{-1} - \mathbf{S}_{\alpha\alpha})^{-1}$, where $\mathbf{\Upsilon} = \text{Diag}(\boldsymbol{\upsilon}) \in \mathbb{C}^{M \times M}$ with $\boldsymbol{\upsilon} \in \mathbb{C}^M$ including the tunable phase shifts applied to the ports: $\upsilon_m = e^{\text{j}\theta_m}$ ($m = 1, 2, \ldots, M$) with $|\upsilon_m| = 1$ and $\theta_m \in [0, 2\pi]$. Putting

all above together, the baseband received signals at all K users can be mathematically expressed as follows:

$$\mathbf{y} \triangleq \rho \mathbf{H}_{\text{r-u}}^{\mathsf{H}} \mathbf{S}_{\beta\alpha} (\boldsymbol{\Upsilon}^{-1} - \mathbf{S}_{\alpha\alpha})^{-1} \mathbf{S}_{\alpha\beta} \mathbf{H}_{\text{b-r}} \mathbf{F} \mathbf{s} + \rho \mathbf{w}, \tag{3.11}$$

where $\mathbf{y} = [\mathsf{y}_1, \mathsf{y}_2, \ldots, \mathsf{y}_K]^{\mathsf{T}}$ and $\mathbf{s} \triangleq [\mathsf{s}_1, \mathsf{s}_2, \ldots, \mathsf{s}_K]^{\mathsf{T}}$ represents the transmit symbol vector such that $\mathsf{s}_k \sim \mathcal{CN}(0,1)\ \forall k$. The receiver scaling factor $\rho \in \mathbb{R}$ and is assumed to be common for all users. The vector $\mathbf{w}$ denotes the Additive White Gaussian Noise (AWGN) containing the noise components $\mathsf{w}_k \sim \mathcal{CN}(0, \sigma_{\text{w}}^2)$. Matrix $\mathbf{F} \triangleq [\mathbf{f}_1, \mathbf{f}_2, \ldots, \mathbf{f}_K]$ with $\mathbf{f}_k \in \mathbb{C}^N$ represents the transmit beamforming matrix for which it holds $\mathbb{E}_{\mathbf{s}} \left\{ \|\mathbf{F}\mathbf{s}\|^2 \right\} = P$ with P being the total transmit power. We define the end-to-end RIS-parametrized multi-user channel as follows:

$$\mathbf{H}^{\mathsf{H}} \triangleq \mathbf{H}_{\text{r-u}}^{\mathsf{H}} \mathbf{S}_{\beta\alpha} (\boldsymbol{\Upsilon}^{-1} - \mathbf{S}_{\alpha\alpha})^{-1} \mathbf{S}_{\alpha\beta} \mathbf{H}_{\text{b-r}}. \tag{3.12}$$

It is noted that, in [31], it was assumed for the radiation patterns that $\mathbf{S}_{\beta\alpha} = \mathbf{S}_{\alpha\beta} = \mathbf{I}_M$.

We now define the Minimum Mean Squared Error (MMSE) criterion [31, 46, 57] which will serve as our system's design optimization objective. The connection between the achievable sum-rate performance and the MMSE of each of the K users can be expressed as follows [31, 46]:

$$C \ = \ \sum_{k=1}^{K} \log_2 \left(\frac{1}{\text{MMSE}_k} \right) \ = \ \log_2 \left(\prod_{k=1}^{K} \frac{1}{\text{MMSE}_k} \right), \tag{3.13}$$

with the total MSE for all K users given by: [31, 46]:

$$\sum_{k=1}^{K} \mathbb{E}_{\mathsf{y}_k, \mathsf{s}_k} \left\{ |\mathsf{y}_k - \mathsf{s}_k|^2 \right\} \ = \mathbb{E}_{\mathbf{y}, \mathbf{s}} \left\{ \|\mathbf{y} - \mathbf{s}\|_2^2 \right\}, \tag{3.14}$$

with $\mathbb{E}_{\mathsf{y}_k, \mathsf{s}_k} \left\{ |\mathsf{y}_k - \mathsf{s}_k|^2 \right\}$ indicating the MSE of the received symbol for each k-th user.

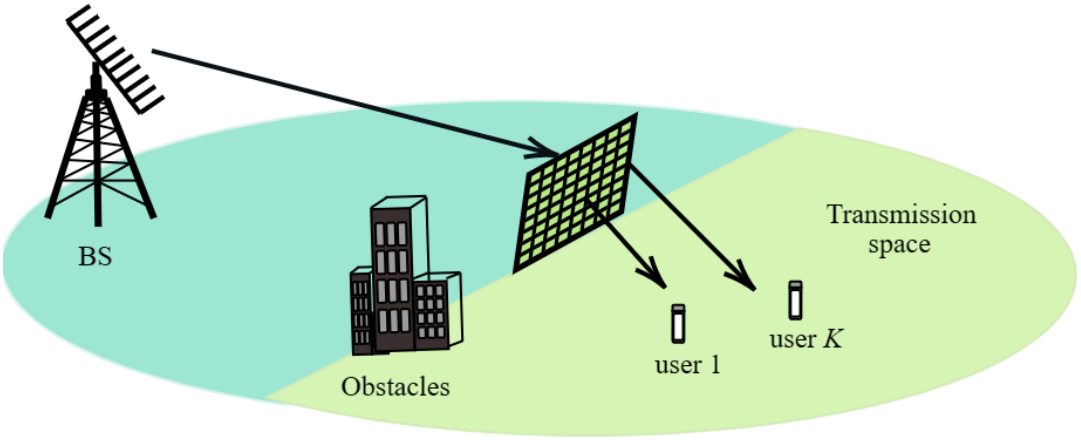


Figure 3.4: The considered RIS-aided multi-user MIMO system model for transmissive RIS.

System model for transmissive RIS setup

Figure 3.4 shows the system model of the transmissive RIS setup. According to the previous subsection, the phase shift matrix corresponding to transmissive RIS is given by $\mathbf{\Phi} \triangleq \mathbf{S}^{(\mathbf{2})}_{\alpha\beta}\mathbf{\Upsilon}\mathbf{S}^{(\mathbf{1})}_{\beta\alpha}$. Similar to reflective setup, $\mathbf{\Upsilon} = \text{Diag}(\boldsymbol{\upsilon}) \in \mathbb{C}^{M\times M}$ with $\boldsymbol{\upsilon} \in \mathbb{C}^{M}$. Even though the matrices $\mathbf{S}^{(\mathbf{1})}_{\alpha\alpha} = \mathbf{S}^{(\mathbf{2})}_{\alpha\alpha} = \mathbf{0}_M$, the phase shift matrix is non-diagonal as it incorporated $\mathbf{S}^{(\mathbf{2})}_{\alpha\beta}$, and $\mathbf{S}^{(\mathbf{1})}_{\beta\alpha}$. Thus, the total received signal can be expressed as:

$$\mathbf{y} \triangleq \rho\mathbf{H}^{\mathsf{H}}_{\text{r-u}}\mathbf{S}^{(\mathbf{2})}_{\alpha\beta}\mathbf{\Upsilon}\mathbf{S}^{(\mathbf{1})}_{\beta\alpha}\mathbf{H}_{\text{b-r}}\mathbf{F}\mathbf{s} + \ \rho\mathbf{w}. \tag{3.15}$$

The vectors $\mathbf{y}$, $\mathbf{s}$, $\mathbf{w}$, and the scaling factor ρ carry the same meaning as in the reflective scenario. Furthermore, the matrix $\mathbf{F}$ with $\mathbf{f}_k \in \mathbb{C}^N$ is the transmit beamforming matrix ensuring the $\mathbb{E}\left\{\|\mathbf{F}\mathbf{s}\|^2\right\} = P$. Thereby, the end-to-end RIS-parametrized multi-user channel for the transmissive RIS as follows:

$$\mathbf{H}^{\mathsf{H}}_{\text{T}} \triangleq \mathbf{H}^{\mathsf{H}}_{\text{r-u}}\mathbf{S}^{(\mathbf{2})}_{\alpha\beta}\mathbf{\Upsilon}\mathbf{S}^{(\mathbf{1})}_{\beta\alpha}\mathbf{H}_{\text{b-r}}. \tag{3.16}$$

Note that the MMSE criterion was employed to formulate the problem similarly to the reflective setup.

3.2 Problem Formulation

Finding the scattering parameters for each channel that optimize the sum rate in (3.13) on the fly (i.e., in the online phase) is impractical since this will require fabricating the RIS accordingly. Therefore, in this paper, we focus on optimizing the scattering matrices $\mathbf{S}_{\alpha\alpha}$, $\mathbf{S}_{\alpha\beta}$, and $\mathbf{S}_{\beta\alpha}$ offline for a particular class of channels corresponding to reflective RIS and the scattering matrices $\mathbf{S}_{\alpha\beta}^{(\mathbf{2})}$ and $\mathbf{S}_{\beta\alpha}^{(\mathbf{1})}$ are optimized offline for a particular class of channels for transmissive RIS. Then, optimize $\mathbf{F}$ and $\boldsymbol{\Upsilon}$ online (i.e., for each channel realization) for both models.

3.2.1 Problem Formulation for Reflective RIS

In particular, we formulate the following optimization problem for reflective RIS:

$$\underset{\mathbf{S}_{\alpha\alpha},\mathbf{S}_{\alpha\beta},\mathbf{S}_{\beta\alpha}}{\arg\min} \quad \mathbb{E}_{\mathbf{H}_{\text{r-u}},\mathbf{H}_{\text{b-r}}}\left[\underset{\rho,\mathbf{F},\boldsymbol{\Upsilon}}{\arg\min}\, \mathbb{E}_{\mathbf{y},\mathbf{s}}\left\{\|\mathbf{y}-\mathbf{s}\|_2^2\right\}\right], \tag{3.17a}$$

$$\text{subject to} \quad \mathbb{E}_{\mathbf{s}}\left\{\|\mathbf{F}\mathbf{s}\|_2^2\right\} = P, \tag{3.17b}$$

$$\upsilon_{im} = 0, \quad \forall i \neq m, \tag{3.17c}$$

$$|\upsilon_{ii}| = 1, \quad \forall i = 1, 2, \ldots, M, \tag{3.17d}$$

$$\mathbf{S}_{\alpha\alpha}{\mathbf{S}_{\alpha\alpha}}^{\mathsf{H}} + \mathbf{S}_{\alpha\beta}{\mathbf{S}_{\alpha\beta}}^{\mathsf{H}} = \mathbf{I}_M, \tag{3.17e}$$

$$\mathbf{S}_{\alpha\beta} = {\mathbf{S}_{\beta\alpha}}^{\mathsf{T}} \text{ and } \mathbf{S}_{\alpha\alpha} = {\mathbf{S}_{\alpha\alpha}}^{\mathsf{T}}. \tag{3.17f}$$

Constraint (3.17b) is the power allocation constraint, whereas (3.17c) and (3.17d) are unimodular constraints related to the RIS phase shift matrix. Furthermore, the last two constraints stand for the losslessness (conservation of power) and reciprocity of the

reflective structure respectively. Using (3.11) and (3.12), the optimization problem is rewritten as:

$$\underset{\mathbf{S}_{\alpha\alpha},\mathbf{S}_{\alpha\beta},\mathbf{S}_{\beta\alpha}}{\arg\min} \quad \mathbb{E}_{\mathbf{H}_{\text{r-u}},\mathbf{H}_{\text{b-r}}}\Big\{\underset{\rho,\mathbf{F},\boldsymbol{\Upsilon}}{\arg\min} \left\|\rho\mathbf{H}^{\mathsf{H}}\mathbf{F}-\mathbf{I}_K\right\|_{\text{F}}^2 + K\rho^2\sigma_{\text{w}}^2\Big\}, \tag{3.18a}$$

$$\text{subject to} \quad \|\mathbf{F}\|_{\text{F}}^2 = P, \tag{3.18b}$$

$$\upsilon_{im} = 0, \quad \forall i \neq m, \tag{3.18c}$$

$$|\upsilon_{ii}| = 1, \quad \forall i = 1, 2, \ldots, M, \tag{3.18d}$$

$$\mathbf{S}_{\alpha\alpha}{\mathbf{S}_{\alpha\alpha}}^{\mathsf{H}} + \mathbf{S}_{\alpha\beta}{\mathbf{S}_{\alpha\beta}}^{\mathsf{H}} = \mathbf{I}_M, \tag{3.18e}$$

$$\mathbf{S}_{\alpha\beta} = {\mathbf{S}_{\beta\alpha}}^{\mathsf{T}} \text{ and } \mathbf{S}_{\alpha\alpha} = {\mathbf{S}_{\alpha\alpha}}^{\mathsf{T}}, \tag{3.18f}$$

where σ_{w}^2 is the noise variance. By incorporating the equality $\mathbf{S}_{\alpha\beta} = \mathbf{S}_{\beta\alpha}^{\mathsf{T}}$ into the problem's objective function (3.18a) and using the matrix notation $\hat{\mathbf{H}}^{\mathsf{H}} \triangleq \mathbf{H}_{\text{r-u}}^{\mathsf{H}}\mathbf{S}_{\alpha\beta}^{\mathsf{T}}(\boldsymbol{\Upsilon}^{-1} - \mathbf{S}_{\alpha\alpha})^{-1}\mathbf{S}_{\alpha\beta}\mathbf{H}_{\text{b-r}}$, yields the following reformulation of the considered design problem:

$$\underset{\mathbf{S}_{\alpha\alpha},\mathbf{S}_{\alpha\beta}}{\arg\min} \quad \mathbb{E}_{\mathbf{H}_{\text{r-u}},\mathbf{H}_{\text{b-r}}}\Big\{\underset{\rho,\mathbf{F},\boldsymbol{\Upsilon}}{\arg\min} \left\|\rho\hat{\mathbf{H}}^{\mathsf{H}}\mathbf{F}-\mathbf{I}_K\right\|_{\text{F}}^2 + K\rho^2\sigma_{\text{w}}^2\Big\}, \tag{3.19a}$$

$$\text{subject to} \quad \|\mathbf{F}\|_{\text{F}}^2 = P, \tag{3.19b}$$

$$\upsilon_{im} = 0, \quad \forall i \neq m, \tag{3.19c}$$

$$|\upsilon_{ii}| = 1, \quad \forall i = 1, 2, \cdots, M, \tag{3.19d}$$

$$\mathbf{S}_{\alpha\alpha}{\mathbf{S}_{\alpha\alpha}}^{\mathsf{H}} + \mathbf{S}_{\alpha\beta}{\mathbf{S}_{\alpha\beta}}^{\mathsf{H}} = \mathbf{I}_M, \tag{3.19e}$$

$$\mathbf{S}_{\alpha\alpha} = {\mathbf{S}_{\alpha\alpha}}^{\mathsf{T}}. \tag{3.19f}$$

Note that, the matrices $\mathbf{S}_{\alpha\alpha}$ and $\mathbf{S}_{\alpha\beta}$ demonstrate partial isometry and exhibit a 2D Toeplitz structure [36]. Consequently, these matrices can be diagonalized using 2-Discrete Fourier Transform (2-DFT) matrices. This facilitates a reduction in the number of uncoupled ports for the same setup and mitigates beam coupling by avoiding oversampling [36]. Hence, to reduce the complexity of the optimization process and

to enhance mathematical tractability, we express the scattering matrices as $\mathbf{S}_{\alpha\alpha} = \mathbf{U}\boldsymbol{\Sigma}_{\alpha\alpha}\mathbf{V}^{\mathsf{H}}$ and $\mathbf{S}_{\alpha\beta} = \mathbf{U}\boldsymbol{\Sigma}_{\alpha\beta}\mathbf{V}^{\mathsf{H}}$, where $\boldsymbol{\Sigma}_{\alpha\alpha}$ and $\boldsymbol{\Sigma}_{\alpha\beta}$ are complex diagonal matrices whose respective M non-zero elements will be optimized, and the $M \times M$ matrices $\mathbf{U}$ and $\mathbf{V}$ will be considered as fixed throughout the process. In particular, both $\mathbf{U}$ and $\mathbf{V}$ are generated as $\mathbf{D} \otimes \mathbf{D}$ with $\mathbf{D} \in \mathbb{C}^{\sqrt{M}\times\sqrt{M}}$ being the DFT matrix. The utilization of 2-DFT matrices is also inspired by the concept of cell sectorization, which holds practical significance. Using this simplification and the notation $\tilde{\mathbf{H}}^{\mathsf{H}} \triangleq \mathbf{H}_{\text{r-u}}^{\mathsf{H}}(\mathbf{U}\boldsymbol{\Sigma}_{\alpha\beta}\mathbf{V}^{\mathsf{H}})^{\mathsf{T}}(\boldsymbol{\Upsilon}^{-1} - \mathbf{U}\boldsymbol{\Sigma}_{\alpha\alpha}\mathbf{V}^{\mathsf{H}})^{-1}(\mathbf{U}\boldsymbol{\Sigma}_{\alpha\beta}\mathbf{V}^{\mathsf{H}})\mathbf{H}_{\text{b-r}}$, the following formulation is deduced:

$$\underset{\boldsymbol{\Sigma}_{\alpha\alpha},\boldsymbol{\Sigma}_{\alpha\beta}}{\arg\min} \quad \mathbb{E}_{\mathbf{H}_{\text{r-u}},\mathbf{H}_{\text{b-r}}}\Big\{\underset{\rho,\mathbf{F},\boldsymbol{\Upsilon}}{\arg\min} \left\|\rho\tilde{\mathbf{H}}^{\mathsf{H}}\mathbf{F} - \mathbf{I}_K\right\|_{\text{F}}^2 + K\rho^2\sigma_{\text{w}}^2\Big\}, \tag{3.20a}$$

$$\text{subject to} \quad \|\mathbf{F}\|_{\text{F}}^2 = P, \tag{3.20b}$$

$$\upsilon_{im} = 0, \quad \forall i \neq m, \tag{3.20c}$$

$$|\upsilon_{ii}| = 1, \quad \forall i = 1, 2, \cdots, M, \tag{3.20d}$$

$$\boldsymbol{\Sigma}_{\alpha\alpha}{\boldsymbol{\Sigma}_{\alpha\alpha}}^{\mathsf{H}} + \boldsymbol{\Sigma}_{\alpha\beta}{\boldsymbol{\Sigma}_{\alpha\beta}}^{\mathsf{H}} = \mathbf{I}_M, \tag{3.20e}$$

$$\mathbf{U}\boldsymbol{\Sigma}_{\alpha\alpha}\mathbf{V}^{\mathsf{H}} = (\mathbf{U}\boldsymbol{\Sigma}_{\alpha\alpha}\mathbf{V}^{\mathsf{H}})^{\mathsf{T}}. \tag{3.20f}$$

3.2.2 Problem Formulation for Transmissive RIS

The optimization problem for the transmissive RIS is formulated as follows:

$$\underset{\mathbf{S}^{(\mathbf{2})}_{\alpha\beta},\mathbf{S}^{(\mathbf{1})}_{\beta\alpha}}{\arg\min} \quad \mathbb{E}_{\mathbf{H}_{\text{r-u}},\mathbf{H}_{\text{b-r}}}\left[\underset{\rho,\mathbf{F},\boldsymbol{\Upsilon}}{\arg\min}\, \mathbb{E}_{\mathbf{y},\mathbf{s}}\left\{\|\mathbf{y}-\mathbf{s}\|_2^2\right\}\right], \tag{3.21a}$$

$$\text{subject to} \quad \mathbb{E}_{\mathbf{s}}\left\{\|\mathbf{F}\mathbf{s}\|_2^2\right\} = P, \tag{3.21b}$$

$$\upsilon_{im} = 0, \quad \forall i \neq m, \tag{3.21c}$$

$$|\upsilon_{ii}| = 1, \quad \forall i = 1, 2, \ldots, M, \tag{3.21d}$$

$$\mathbf{S}^{(\mathbf{1})}_{\beta\alpha}\mathbf{S}^{(\mathbf{1})\mathsf{H}}_{\beta\alpha} = \mathbf{I}_M, \tag{3.21e}$$

$$\mathbf{S}^{(\mathbf{2})}_{\alpha\beta}\mathbf{S}^{(\mathbf{2})\mathsf{H}}_{\alpha\beta} = \mathbf{I}_M. \tag{3.21f}$$

By leveraging the equations (3.15) and (3.16), the formulated problem for transmissive RIS can be rewritten as follows:

$$\underset{\mathbf{S}^{(\mathbf{2})}_{\alpha\beta},\mathbf{S}^{(\mathbf{1})}_{\beta\alpha}}{\arg\min}\, \mathbb{E}_{\mathbf{H}_{\text{r-u}},\mathbf{H}_{\text{b-r}}}\Big\{\underset{\rho,\mathbf{F},\boldsymbol{\Upsilon}}{\arg\min}\, \left\|\rho\mathbf{H}_{\text{T}}^{\mathsf{H}}\mathbf{F} - \mathbf{I}_K\right\|_{\text{F}}^2 + K\rho^2\sigma_{\text{w}}^2\Big\}, \tag{3.22a}$$

$$\text{subject to} \quad \|\mathbf{F}\|_{\text{F}}^2 = P, \tag{3.22b}$$

$$\upsilon_{im} = 0, \quad \forall i \neq m, \tag{3.22c}$$

$$|\upsilon_{ii}| = 1, \quad \forall i = 1, 2, \ldots, M, \tag{3.22d}$$

$$\mathbf{S}^{(\mathbf{1})}_{\beta\alpha}\mathbf{S}^{(\mathbf{1})\mathsf{H}}_{\beta\alpha} = \mathbf{I}_M, \tag{3.22e}$$

$$\mathbf{S}^{(\mathbf{2})}_{\alpha\beta}\mathbf{S}^{(\mathbf{2})\mathsf{H}}_{\alpha\beta} = \mathbf{I}_M. \tag{3.22f}$$

Here $\mathbf{H}_{\text{T}}^{\mathsf{H}}$ is given by equation (3.16). The constraints (3.22e) and (3.22f) imply the conservation of power property. Unlike the reflective problem, we are not using any transformations here. We directly proceed with the above-formulated problem.

3.3 Proposed EM-Consistent Design

The problems become complex as they involve outer and inner minimization problems and is thus difficult to solve. In this section, we present a decomposition of the overall system design into two sub-problems and deploy an alternating optimization approach for both setups.

3.3.1 Proposed Solution for Reflective RIS

The two subproblems are given as follows:

1. For given $\mathbf{\Sigma}_{\alpha\alpha}$ and $\mathbf{\Sigma}_{\alpha\beta}$, solve:

$$\begin{aligned} \underset{\rho,\mathbf{F},\mathbf{\Upsilon}}{\arg\min} \quad & \left\| \rho\tilde{\mathbf{H}}^{\mathsf{H}}\mathbf{F} - \mathbf{I}_K \right\|_{\mathrm{F}}^2 + K\rho^2\sigma_{\mathrm{w}}^2, && \text{(3.23a)} \\ \text{subject to} \quad & \|\mathbf{F}\|_{\mathrm{F}}^2 = P, && \text{(3.23b)} \\ & \upsilon_{im} = 0, \quad \forall i \neq m, && \text{(3.23c)} \\ & |\upsilon_{ii}| = 1, \quad \forall i = 1, 2, \ldots, M. && \text{(3.23d)} \end{aligned}$$

2. For given ρ, $\mathbf{F}$, and $\mathbf{\Upsilon}$, solve:

$$\begin{aligned} \underset{\mathbf{\Sigma}_{\alpha\alpha},\mathbf{\Sigma}_{\alpha\beta}}{\arg\min} \quad & \mathbb{E}_{\mathbf{H}_{\text{r-u}},\mathbf{H}_{\text{b-r}}}\left\{ \left\| \rho\tilde{\mathbf{H}}^{\mathsf{H}}\mathbf{F} - \mathbf{I}_K \right\|_{\mathrm{F}}^2 + K\rho^2\sigma_{\mathrm{w}}^2 \right\}, && \text{(3.24a)} \\ \text{subject to} \quad & \mathbf{\Sigma}_{\alpha\alpha}\mathbf{\Sigma}_{\alpha\alpha}^{\mathsf{H}} + \mathbf{\Sigma}_{\alpha\beta}\mathbf{\Sigma}_{\alpha\beta}^{\mathsf{H}} = \mathbf{I}_M, && \text{(3.24b)} \\ & \mathbf{U}\mathbf{\Sigma}_{\alpha\alpha}\mathbf{V}^{\mathsf{H}} = (\mathbf{U}\mathbf{\Sigma}_{\alpha\alpha}\mathbf{V}^{\mathsf{H}})^{\mathsf{T}}. && \text{(3.24c)} \end{aligned}$$

The first sub-problem can be efficiently solved via the method described in [18,31] based on the gradient descent algorithm. The solution to the second sub-problem, which gives the optimal scattering matrices, is next obtained via a gradient step and a closed-form projection step for both $\mathbf{\Sigma}_{\alpha\alpha}$ and $\mathbf{\Sigma}_{\alpha\beta}$ matrices. Recall from the previous section that we have used the assumptions $\mathbf{S}_{\alpha\alpha} = \mathbf{U}\mathbf{\Sigma}_{\alpha\alpha}\mathbf{V}^{\mathsf{H}}$ and $\mathbf{S}_{\alpha\beta} = \mathbf{U}\mathbf{\Sigma}_{\alpha\beta}\mathbf{V}^{\mathsf{H}}$, where $\mathbf{\Sigma}_{\alpha\alpha}$ and

$\boldsymbol{\Sigma}_{\alpha\beta}$ are complex diagonal matrices and $\mathbf{U}$ and $\mathbf{V}$ are fixed DFT-based matrices. Hence, the optimization of the second sub-problem is conducted with respect to $\boldsymbol{\Sigma}_{\alpha\alpha}$ and $\boldsymbol{\Sigma}_{\alpha\beta}$ instead of their corresponding scattering parameters.

Projected Gradient Descent

For each channel realization, we define the following MSE function:

$$f(\boldsymbol{\Sigma}_{\alpha\alpha}, \boldsymbol{\Sigma}_{\alpha\beta}) \triangleq \left\| \rho \tilde{\mathbf{H}}^{\mathsf{H}} \mathbf{F} - \mathbf{I}_K \right\|_{\mathrm{F}}^2 + K\rho^2 \sigma_{\mathrm{w}}^2. \tag{3.25}$$

Since the objective function (3.24a) involves an expectation with respect to $\mathbf{H}_{\text{r-u}}$ and $\mathbf{H}_{\text{b-r}}$, Monte Carlo sampling can be used to compute the gradients for a set of Q channel samples. To this end, the overall gradient of this objective function with respect to $\boldsymbol{\Sigma}_{\alpha\alpha}$ is computed as follows:

$$\mathbf{G}_{\boldsymbol{\Sigma}_{\alpha\alpha}} \triangleq \frac{1}{Q} \sum_{q=1}^{Q} \left[\frac{\partial f(\boldsymbol{\Sigma}_{\alpha\alpha}, \boldsymbol{\Sigma}_{\alpha\beta})}{\partial \boldsymbol{\Sigma}_{\alpha\alpha}} \right]_q, \tag{3.26}$$

where the gradient for each q-th channel sample is given by:

$$\begin{aligned} \left[\frac{\partial f(\boldsymbol{\Sigma}_{\alpha\alpha}, \boldsymbol{\Sigma}_{\alpha\beta})}{\partial \boldsymbol{\Sigma}_{\alpha\alpha}} \right]_q &= 2\mathbf{U}^{\mathsf{T}} (\boldsymbol{\Upsilon}_q^{-1} - \mathbf{U}\boldsymbol{\Sigma}_{\alpha\alpha}\mathbf{V}^{\mathsf{H}})^{-\mathsf{T}} \left(\rho_q \mathbf{H}_{\text{r-u}}^{\mathsf{H}} (\mathbf{U}\boldsymbol{\Sigma}_{\alpha\beta}\mathbf{V}^{\mathsf{H}})^{\mathsf{T}} \right)^{\mathsf{T}} \\ &\left(\rho_q \mathbf{H}_{\text{r-u}}^{\mathsf{H}} (\mathbf{U}\boldsymbol{\Sigma}_{\alpha\beta}\mathbf{V}^{\mathsf{H}})^{\mathsf{T}} (\boldsymbol{\Upsilon}_q^{-1} - \mathbf{U}\boldsymbol{\Sigma}_{\alpha\alpha}\mathbf{V}^{\mathsf{H}})^{-1} (\mathbf{U}\boldsymbol{\Sigma}_{\alpha\beta}\mathbf{V}^{\mathsf{H}}) \mathbf{H}_{\text{b-r}} \mathbf{F}_q - \mathbf{I}_K \right)^{*} \\ &\left((\mathbf{U}\boldsymbol{\Sigma}_{\alpha\beta}\mathbf{V}^{\mathsf{H}}) \mathbf{H}_{\text{b-r}} \mathbf{F}_q \right)^{\mathsf{T}} (\boldsymbol{\Upsilon}_q^{-1} - \mathbf{U}\boldsymbol{\Sigma}_{\alpha\alpha}\mathbf{V}^{\mathsf{H}})^{-\mathsf{T}} \mathbf{V}^{*}, \end{aligned} \tag{3.27}$$

with $\boldsymbol{\Upsilon}_q$, $\mathbf{F}_q$, and ρ_q indicating the optimum RIS phase configuration, BS precoding matrix, and scaling parameter values for each q-th channel sample. Similarly, the gradient of (3.24a) with respect to $\boldsymbol{\Sigma}_{\alpha\beta}$ can be approximated as:

$$\mathbf{G}_{\boldsymbol{\Sigma}_{\alpha\beta}} \triangleq \frac{1}{Q} \sum_{q=1}^{Q} \left[\frac{\partial f(\boldsymbol{\Sigma}_{\alpha\alpha}, \boldsymbol{\Sigma}_{\alpha\beta})}{\partial \boldsymbol{\Sigma}_{\alpha\beta}} \right]_q, \tag{3.28}$$

where the gradient for each q-th channel sample is given by:

$$\begin{aligned}
&\left[\frac{\partial f(\boldsymbol{\Sigma}_{\alpha\alpha}, \boldsymbol{\Sigma}_{\alpha\beta})}{\partial \boldsymbol{\Sigma}_{\alpha\beta}}\right]_q = 2\left(\rho_q \mathbf{H}_{\text{r-u}}^{\mathsf{H}} (\mathbf{V}^{\mathsf{H}})^{\mathsf{T}}\right)^{\mathsf{T}} \\
&\left(\rho_q \mathbf{H}_{\text{r-u}}^{\mathsf{H}} (\mathbf{U}\boldsymbol{\Sigma}_{\alpha\beta}\mathbf{V}^{\mathsf{H}})^{\mathsf{T}} (\boldsymbol{\Upsilon}_q^{-1} - \mathbf{U}\boldsymbol{\Sigma}_{\alpha\alpha}\mathbf{V}^{\mathsf{H}})^{-1} (\mathbf{U}\boldsymbol{\Sigma}_{\alpha\beta}\mathbf{V}^{\mathsf{H}})\mathbf{H}_{\text{b-r}}\mathbf{F}_q - \mathbf{I}_K\right)^* \\
&(\mathbf{V}^{\mathsf{H}}\mathbf{H}_{\text{b-r}}\mathbf{F}_q)^{\mathsf{T}} \boldsymbol{\Sigma}_{\alpha\beta}{}^{\mathsf{T}} \left(\mathbf{U}^{\mathsf{T}} (\boldsymbol{\Upsilon}_q^{-1} - \mathbf{U}\boldsymbol{\Sigma}_{\alpha\alpha}\mathbf{V}^{\mathsf{H}})^{-1}\mathbf{U}\right)^{\mathsf{T}} + \\
&2\left(\mathbf{U}^{\mathsf{T}} (\boldsymbol{\Upsilon}_q^{-1} - \mathbf{U}\boldsymbol{\Sigma}_{\alpha\alpha}\mathbf{V}^{\mathsf{H}})^{-1}\mathbf{U}\right)^{\mathsf{T}} \boldsymbol{\Sigma}_{\alpha\beta}{}^{\mathsf{T}} \left(\rho_q \mathbf{H}_{\text{r-u}}^{\mathsf{H}} (\mathbf{V}^{\mathsf{H}})^{\mathsf{T}}\right)^{\mathsf{T}} \\
&\left(\rho_q \mathbf{H}_{\text{r-u}}^{\mathsf{H}} (\mathbf{U}\boldsymbol{\Sigma}_{\alpha\beta}\mathbf{V}^{\mathsf{H}})^{\mathsf{T}} (\boldsymbol{\Upsilon}_q^{-1} - \mathbf{U}\boldsymbol{\Sigma}_{\alpha\alpha}\mathbf{V}^{\mathsf{H}})^{-1} (\mathbf{U}\boldsymbol{\Sigma}_{\alpha\beta}\mathbf{V}^{\mathsf{H}})\mathbf{H}_{\text{b-r}}\mathbf{F}_q - \mathbf{I}_K\right)^* \\
&(\mathbf{V}^{\mathsf{H}}\mathbf{H}_{\text{b-r}}\mathbf{F}_q)^{\mathsf{T}}.
\end{aligned} \tag{3.29}$$

Once the gradients are computed, the parameters $\boldsymbol{\Sigma}_{\alpha\alpha}$ and $\boldsymbol{\Sigma}_{\alpha\beta}$ are updated as follows:

$$\tilde{\boldsymbol{\Sigma}}_{\alpha\alpha} = \boldsymbol{\Sigma}_{\alpha\alpha} - \mu \mathbf{Diag}\left\{\mathbf{G}^*_{\boldsymbol{\Sigma}_{\alpha\alpha}}\right\}, \tag{3.30}$$

$$\tilde{\boldsymbol{\Sigma}}_{\alpha\beta} = \boldsymbol{\Sigma}_{\alpha\beta} - \mu \mathbf{Diag}\left\{\mathbf{G}^*_{\boldsymbol{\Sigma}_{\alpha\beta}}\right\}, \tag{3.31}$$

where we have adopted the conjugate of the gradients as the search directions and μ denotes the fixed step size. To maintain the symmetry of $\tilde{\boldsymbol{\Sigma}}_{\alpha\alpha}$ as per (3.24c), we symmetrically update its diagonal elements at positions $\tilde{\sigma}_{\alpha\alpha,ii}$ and $\tilde{\sigma}_{\alpha\alpha,(M-i+2)(M-i+2)}$, $\forall i = 2, 3, \ldots, M/2$ by taking their average, as follows:

$$\tilde{\sigma}_{\alpha\alpha,ii} = \frac{\tilde{\sigma}_{\alpha\alpha,ii} + \tilde{\sigma}_{\alpha\alpha,(M-i+2)(M-i+2)}}{2}, \tag{3.32}$$

$$\tilde{\sigma}_{\alpha\alpha,(M-i+2)(M-i+2)} = \frac{\tilde{\sigma}_{\alpha\alpha,ii} + \tilde{\sigma}_{\alpha\alpha,(M-i+2)(M-i+2)}}{2}, \tag{3.33}$$

Clearly, when $\boldsymbol{\Sigma}_{\alpha\alpha}$ is symmetric, $\boldsymbol{\Sigma}_{\alpha\beta}$ becomes symmetric as well since $\mathbf{I}_M$ is also symmetric. Therefore, we also make $\tilde{\boldsymbol{\Sigma}}_{\alpha\beta}$ symmetric using similar expressions to (3.32) and (3.33).

We now formulate the following projection problem that considers both constraints (3.24b) and (3.24c) for the design of symmetric $\boldsymbol{\Sigma}_{\alpha\alpha}$ and $\boldsymbol{\Sigma}_{\alpha\beta}$:

$$\underset{\boldsymbol{\Sigma}_{\alpha\alpha}, \boldsymbol{\Sigma}_{\alpha\beta}}{\arg\min} \quad \left\|\boldsymbol{\Sigma}_{\alpha\alpha} - \tilde{\boldsymbol{\Sigma}}_{\alpha\alpha}\right\|_{\text{F}}^2 + \left\|\boldsymbol{\Sigma}_{\alpha\beta} - \tilde{\boldsymbol{\Sigma}}_{\alpha\beta}\right\|_{\text{F}}^2, \tag{3.34a}$$

$$\text{subject to} \quad \boldsymbol{\Sigma}_{\alpha\alpha}\boldsymbol{\Sigma}_{\alpha\alpha}^{\mathsf{H}} + \boldsymbol{\Sigma}_{\alpha\beta}\boldsymbol{\Sigma}_{\alpha\beta}^{\mathsf{H}} = \mathbf{I}_M. \tag{3.34b}$$

We then simplify the term $\left\|\boldsymbol{\Sigma}_{\alpha\alpha} - \tilde{\boldsymbol{\Sigma}}_{\alpha\alpha}\right\|_{\mathrm{F}}^2$ as follows:

$$\begin{aligned}\left\|\boldsymbol{\Sigma}_{\alpha\alpha} - \tilde{\boldsymbol{\Sigma}}_{\alpha\alpha}\right\|_{\mathrm{F}}^2 &= \mathrm{Tr}\big((\boldsymbol{\Sigma}_{\alpha\alpha} - \tilde{\boldsymbol{\Sigma}}_{\alpha\alpha})(\boldsymbol{\Sigma}_{\alpha\alpha} - \tilde{\boldsymbol{\Sigma}}_{\alpha\alpha})^{\mathsf{H}}\big) \\ &= \sum_{i=1}^{M} |\sigma_{\alpha\alpha,ii}|^2 - 2\mathrm{Re}\{\sigma_{\alpha\alpha,ii}\tilde{\sigma}_{\alpha\alpha,ii}{}^*\} + |\tilde{\sigma}_{\alpha\alpha,ii}|^2,\end{aligned} \tag{3.35}$$

where $\sigma_{\alpha\alpha,ii}$ and $\tilde{\sigma}_{\alpha\alpha,ii}$ are the diagonal elements of $\boldsymbol{\Sigma}_{\alpha\alpha}$ and $\tilde{\boldsymbol{\Sigma}}_{\alpha\alpha}$, respectively. Similarly, the other term $\left\|\boldsymbol{\Sigma}_{\alpha\beta} - \tilde{\boldsymbol{\Sigma}}_{\alpha\beta}\right\|_{\mathrm{F}}^2$ in the objective can be simplified as:

$$\begin{aligned}\left\|\boldsymbol{\Sigma}_{\alpha\beta} - \tilde{\boldsymbol{\Sigma}}_{\alpha\beta}\right\|_{\mathrm{F}}^2 &= \mathrm{Tr}\big((\boldsymbol{\Sigma}_{\alpha\beta} - \tilde{\boldsymbol{\Sigma}}_{\alpha\beta})(\boldsymbol{\Sigma}_{\alpha\beta} - \tilde{\boldsymbol{\Sigma}}_{\alpha\beta})^{\mathsf{H}}\big) \\ &= \sum_{i=1}^{M} |\sigma_{\alpha\beta,ii}|^2 - 2\mathrm{Re}\{\sigma_{\alpha\beta,ii}\tilde{\sigma}_{\alpha\beta,ii}{}^*\} + |\tilde{\sigma}_{\alpha\beta,ii}|^2,\end{aligned} \tag{3.36}$$

where the scalars $\sigma_{\alpha\beta,ii}$ and $\tilde{\sigma}_{\alpha\beta,ii}$ represent the diagonal elements of the matrices $\boldsymbol{\Sigma}_{\alpha\beta}$ and $\tilde{\boldsymbol{\Sigma}}_{\alpha\beta}$, respectively. Using the latter expressions, the projection problem in (3.34a) and (3.34b) can be reformulated as follows:

$$\begin{aligned}\underset{\sigma_{\alpha\alpha,ii},\sigma_{\alpha\beta,ii}}{\arg\min} \; &\sum_{i=1}^{M}(|\sigma_{\alpha\alpha,ii}|^2 - 2\mathrm{Re}\{\sigma_{\alpha\alpha,ii}\tilde{\sigma}_{\alpha\alpha,ii}{}^*\} + |\tilde{\sigma}_{\alpha\alpha,ii}|^2) \\ &+ \sum_{i=1}^{M}(|\sigma_{\alpha\beta,ii}|^2 - 2\mathrm{Re}\{\sigma_{\alpha\beta,ii}\tilde{\sigma}_{\alpha\beta,ii}{}^*\} + |\tilde{\sigma}_{\alpha\beta,ii}|^2),\end{aligned} \tag{3.37a}$$

$$\text{subject to } |\sigma_{\alpha\alpha,ii}|^2 + |\sigma_{\alpha\beta,ii}|^2 = 1, \; \forall i = 1, 2, \ldots, M. \tag{3.37b}$$

In the sequel, we present closed-form solutions for the latter problem using the method of Lagrange multipliers.

Closed-Form Solution

The Lagrangian function of the latter optimization problem is defined as follows:

$$\begin{aligned}L(\sigma_{\alpha\alpha,ii}, \sigma_{\alpha\beta,ii}, \lambda_i) &\triangleq \sum_{i=1}^{M}(|\sigma_{\alpha\alpha,ii}|^2 - 2\mathrm{Re}\{\sigma_{\alpha\alpha,ii}\tilde{\sigma}_{\alpha\alpha,ii}{}^*\} + |\tilde{\sigma}_{\alpha\alpha,ii}|^2) \\ &+ \sum_{i=1}^{M}(|\sigma_{\alpha\beta,ii}|^2 - 2\mathrm{Re}\{\sigma_{\alpha\beta,ii}\tilde{\sigma}_{\alpha\beta,ii}{}^*\} + |\tilde{\sigma}_{\alpha\beta,ii}|^2) + \sum_{i=1}^{M}\lambda_i(|\sigma_{\alpha\alpha,ii}|^2 + |\sigma_{\alpha\beta,ii}|^2 - 1).\end{aligned} \tag{3.38}$$

Its partial derivatives with respect to the optimization variables $\sigma_{\alpha\alpha,ii}$, $\sigma_{\alpha\beta,ii}$, and λ_i are then computed as:

$$\frac{\partial L(\sigma_{\alpha\alpha,ii}, \sigma_{\alpha\beta,ii}, \lambda_i)}{\partial \sigma_{\alpha\alpha,ii}} = 2\sigma_{\alpha\alpha,ii} - 2\tilde{\sigma}_{\alpha\alpha,ii} + 2\lambda_i \sigma_{\alpha\alpha,ii}, \tag{3.39}$$

$$\frac{\partial L(\sigma_{\alpha\alpha,ii}, \sigma_{\alpha\beta,ii}, \lambda_i)}{\partial \sigma_{\alpha\beta,ii}} = 2\sigma_{\alpha\beta,ii} - 2\tilde{\sigma}_{\alpha\beta,ii} + 2\lambda_i \sigma_{\alpha\beta,ii}, \tag{3.40}$$

$$\frac{\partial L(\sigma_{\alpha\alpha,ii}, \sigma_{\alpha\beta,ii}, \lambda_i)}{\partial \lambda_i} = |\sigma_{\alpha\alpha,ii}|^2 + |\sigma_{\alpha\beta,ii}|^2 - 1. \tag{3.41}$$

By equating each of the latter expressions to zero and then solving the resulting 3×3 system of equations, the diagonal elements of $\mathbf{\Sigma}_{\alpha\alpha}$ and $\mathbf{\Sigma}_{\alpha\beta}$ are obtained in closed-form as:

$$\sigma_{\alpha\alpha,ii} = \frac{\tilde{\sigma}_{\alpha\alpha,ii}}{\sqrt{|\tilde{\sigma}_{\alpha\alpha,ii}|^2 + |\tilde{\sigma}_{\alpha\beta,ii}|^2}}, \tag{3.42}$$

$$\sigma_{\alpha\beta,ii} = \frac{\tilde{\sigma}_{\alpha\beta,ii}}{\sqrt{|\tilde{\sigma}_{\alpha\alpha,ii}|^2 + |\tilde{\sigma}_{\alpha\beta,ii}|^2}}. \tag{3.43}$$

The solution of the overall system design problem is summarized in **Algorithm 3**.

Algorithm 3 Proposed System Design for Reflective RIS

Input: Q samples of the channel matrices $\mathbf{H}_{\text{b-r}}$ and $\mathbf{H}_{\text{r-u}}$.

Set $\mathbf{U} = \mathbf{V} = \mathbf{D} \otimes \mathbf{D}$.
Initialize $\widehat{\mathbf{\Sigma}}_{\alpha\alpha_0}$ and $\widehat{\mathbf{\Sigma}}_{\alpha\beta_0}$.
for $k = 1, 2, \ldots, I_{\max}$ **do**
  for $q = 1, 2, \ldots, Q$ **do**
    Initialize $\mathbf{\Upsilon}_k$.
    Initialize ρ_k and $\mathbf{F}_k$ as described in [18, 31].
    Compute $\mathbf{\Upsilon}_k$ and $\mathbf{F}_k$ solving sub-problem 1) via the method in [18, 31].
    Compute $\left[\frac{\partial f(\mathbf{\Sigma}_{\alpha\alpha,k}, \mathbf{\Sigma}_{\alpha\beta,k})}{\partial \mathbf{\Sigma}_{\alpha\alpha,k}}\right]_q$.
    Compute $\left[\frac{\partial f(\mathbf{\Sigma}_{\alpha\alpha,k}, \mathbf{\Sigma}_{\alpha\beta,k})}{\partial \mathbf{\Sigma}_{\alpha\beta,k}}\right]_q$.
  end for
  Compute $\mathbf{G}_{\mathbf{\Sigma}_{\alpha\alpha,k}}$ using (3.26).
  Compute $\mathbf{G}_{\mathbf{\Sigma}_{\alpha\beta,k}}$ using (3.28).
  Compute $\mathbf{\Sigma}_{\alpha\alpha,k} = \mathbf{\Sigma}_{\alpha\alpha,k-1} - \mu \mathbf{Diag}\left\{\mathbf{G}^*_{\mathbf{\Sigma}_{\alpha\alpha,k}}\right\}$.
  Compute $\mathbf{\Sigma}_{\alpha\beta,k} = \mathbf{\Sigma}_{\alpha\beta,k-1} - \mu \mathbf{Diag}\left\{\mathbf{G}^*_{\mathbf{\Sigma}_{\alpha\beta,k}}\right\}$.
  Use (3.32) and (3.33) to symmetrify $\mathbf{\Sigma}_{\alpha\alpha,k}$ and $\mathbf{\Sigma}_{\alpha\beta,k}$.
  Compute $\widehat{\sigma}_{\alpha\alpha,ii,k} = \frac{\sigma_{\alpha\alpha,ii,k}}{\sqrt{|\sigma_{\alpha\alpha,ii,k}|^2 + |\sigma_{\alpha\beta,ii,k}|^2}}$.
  Compute $\widehat{\sigma}_{\alpha\beta,ii,k} = \frac{\sigma_{\alpha\beta,ii,k}}{\sqrt{|\sigma_{\alpha\alpha,ii,k}|^2 + |\sigma_{\alpha\beta,ii,k}|^2}}$.
  Set $\mathbf{S}_{\alpha\alpha,k} = \mathbf{U}\text{Diag}\left([\widehat{\sigma}_{\alpha\alpha,11,k}, \ldots, \widehat{\sigma}_{\alpha\alpha,MM,k}]\right)\mathbf{V}^{\mathsf{H}}$.
  Set $\mathbf{S}_{\alpha\beta,k} = \mathbf{U}\text{Diag}\left([\widehat{\sigma}_{\alpha\beta,11,k}, \ldots, \widehat{\sigma}_{\alpha\beta,MM,k}]\right)\mathbf{V}^{\mathsf{H}}$.
end for

Output: $\rho_{I_{\max}}$, $\mathbf{F}_{I_{\max}}$, $\mathbf{\Upsilon}_{I_{\max}}$, $\mathbf{S}_{\alpha\alpha,I_{\max}}$, and $\mathbf{S}_{\alpha\beta,I_{\max}}$.

3.3.2 Proposed Solution for Transmissive RIS

The two subproblems of transmissive RIS are outlined as follows:

1. For given $\mathbf{S}_{\beta\alpha}^{(\mathbf{1})}$ and $\mathbf{S}_{\alpha\beta}^{(\mathbf{2})}$, solve:

$$\underset{\rho,\mathbf{F},\boldsymbol{\Upsilon}}{\arg\min} \quad \left\|\rho\mathbf{H}_{\mathrm{T}}^{\mathsf{H}}\mathbf{F}-\mathbf{I}_K\right\|_{\mathrm{F}}^2+K\rho^2\sigma_{\mathrm{w}}^2, \tag{3.44a}$$

$$\text{subject to} \quad \|\mathbf{F}\|_{\mathrm{F}}^2=P, \tag{3.44b}$$

$$\upsilon_{im}=0, \quad \forall i\neq m, \tag{3.44c}$$

$$|\upsilon_{ii}|=1, \quad \forall i=1,2,\cdots,M. \tag{3.44d}$$

2. For given ρ, $\mathbf{F}$, and $\boldsymbol{\Upsilon}$, solve:

$$\underset{\mathbf{S}_{\alpha\beta}^{(\mathbf{2})},\mathbf{S}_{\beta\alpha}^{(\mathbf{1})}}{\arg\min} \quad \mathbb{E}_{\mathbf{H}_{\text{r-u}},\mathbf{H}_{\text{b-r}}} \quad \left\{\left\|\rho\mathbf{H}_{\mathrm{T}}^{\mathsf{H}}\mathbf{F}-\mathbf{I}_K\right\|_{\mathrm{F}}^2+K\rho^2\sigma_{\mathrm{w}}^2\right\}, \tag{3.45a}$$

$$\text{subject to} \quad \mathbf{S}_{\beta\alpha}^{(\mathbf{1})}\mathbf{S}_{\beta\alpha}^{(\mathbf{1})^{\mathsf{H}}}=\mathbf{I}_M, \tag{3.45b}$$

$$\mathbf{S}_{\alpha\beta}^{(\mathbf{2})}\mathbf{S}_{\alpha\beta}^{(\mathbf{2})^{\mathsf{H}}}=\mathbf{I}_M. \tag{3.45c}$$

Here $\mathbf{H}_{\mathrm{T}}^{\mathsf{H}}$ is given by equation (3.16). The first sub-problem is addressed utilizing the method detailed in [46] via the VAMP algorithm. Subsequently, the solution to the second sub-problem, which determines the optimal radiation pattern matrices, is obtained by employing the projection gradient descent method, as described below.

Projection Gradient Descent Method

The MSE function for each channel realization corresponding to the transmissive model is presented below:

$$f(\mathbf{S}_{\alpha\beta}^{(\mathbf{2})},\mathbf{S}_{\beta\alpha}^{(\mathbf{1})})=\left\|\rho\mathbf{H}_{\mathrm{T}}^{\mathsf{H}}\mathbf{F}-\mathbf{I}_K\right\|_{\mathrm{F}}^2+K\rho^2\sigma_{\mathrm{w}}^2. \tag{3.46}$$

Here also the objective function (3.45a) contains an expectation involving the channels $\mathbf{H}_{\text{r-u}}$ and $\mathbf{H}_{\text{b-r}}$. Consequently, Monte Carlo sampling is employed to compute the gradient for each parameter, considering a set of Q channels as samples, similar to the

reflective case. Thereby, the overall gradient of the function with respect to $\mathbf{S}^{(\mathbf{1})}_{\beta\alpha}$ is

$$\mathbf{G}_{\mathbf{S}^{(1)}_{\beta\alpha}} = \frac{1}{Q}\sum_{q=1}^{Q}\left[\frac{\partial f(\mathbf{S}^{(\mathbf{2})}_{\alpha\beta}, \mathbf{S}^{(\mathbf{1})}_{\beta\alpha})}{\partial \mathbf{S}^{(\mathbf{1})}_{\beta\alpha}}\right]_q, \tag{3.47}$$

where,

$$\left[\frac{\partial f(\mathbf{S}^{(\mathbf{2})}_{\alpha\beta}, \mathbf{S}^{(\mathbf{1})}_{\beta\alpha})}{\partial \mathbf{S}^{(\mathbf{1})}_{\beta\alpha}}\right]_q = 2\left(\rho_q \mathbf{H}^{\mathsf{H}}_{\text{r-u}} \mathbf{S}^{(\mathbf{2})}_{\alpha\beta} \boldsymbol{\Upsilon}_q\right)^{\mathsf{T}} \left(\rho_q \mathbf{H}^{\mathsf{H}}_{\text{r-u}} \mathbf{S}^{(\mathbf{2})}_{\alpha\beta} \boldsymbol{\Upsilon}_q \mathbf{S}^{(\mathbf{1})}_{\beta\alpha} \mathbf{H}_{\text{b-r}} \mathbf{F}_q - \mathbf{I}_K\right)^{*} (\mathbf{H}_{\text{b-r}} \mathbf{F}_q)^{\mathsf{T}}, \tag{3.48}$$

similar to the reflection case, $\boldsymbol{\Upsilon}_q$, $\mathbf{F}_q$, ρ_q denote optimum phase shift, precoding and scaling parameter values for each channel (sample). Then, the gradient of the function with respect to $\mathbf{S}^{(\mathbf{2})}_{\alpha\beta}$ is given by

$$\mathbf{G}_{\mathbf{S}^{(2)}_{\alpha\beta}} = \frac{1}{Q}\sum_{q=1}^{Q}\left[\frac{\partial f(\mathbf{S}^{(\mathbf{2})}_{\alpha\beta}, \mathbf{S}^{(\mathbf{1})}_{\beta\alpha})}{\partial \mathbf{S}^{(\mathbf{2})}_{\alpha\beta}}\right]_q, \tag{3.49}$$

where

$$\left[\frac{\partial f(\mathbf{S}^{(\mathbf{2})}_{\alpha\beta}, \mathbf{S}^{(\mathbf{1})}_{\beta\alpha})}{\partial \mathbf{S}^{(\mathbf{2})}_{\alpha\beta}}\right]_q = 2\left(\rho_q \mathbf{H}^{\mathsf{H}}_{\text{r-u}}\right)^{\mathsf{T}} \left(\rho_q \mathbf{H}^{\mathsf{H}}_{\text{r-u}} \mathbf{S}^{(\mathbf{2})}_{\alpha\beta} \boldsymbol{\Upsilon}_q \mathbf{S}^{(\mathbf{1})}_{\beta\alpha} \mathbf{H}_{\text{b-r}} \mathbf{F}_q - \mathbf{I}_K\right)^{*} \left(\boldsymbol{\Upsilon}_q \mathbf{S}^{(\mathbf{1})}_{\beta\alpha} \mathbf{H}_{\text{b-r}} \mathbf{F}_q\right)^{\mathsf{T}}, \tag{3.50}$$

Once the gradients are calculated, the next step would be to update the matrices $\mathbf{S}^{(\mathbf{1})}_{\beta\alpha}$ and $\mathbf{S}^{(\mathbf{2})}_{\alpha\beta}$.

$$\tilde{\mathbf{S}}^{(\mathbf{1})}_{\beta\alpha} = \mathbf{S}^{(\mathbf{1})}_{\beta\alpha} - \mu\Big\{\mathbf{G}^{*}_{\mathbf{S}^{(1)}_{\beta\alpha}}\Big\}, \tag{3.51}$$

$$\tilde{\mathbf{S}}^{(\mathbf{2})}_{\alpha\beta} = \mathbf{S}^{(\mathbf{2})}_{\alpha\beta} - \mu\Big\{\mathbf{G}^{*}_{\mathbf{S}^{(2)}_{\alpha\beta}}\Big\}. \tag{3.52}$$

Given our operation within the complex domain, we employ Wirtinger derivatives. Consequently, we utilize the conjugate of the gradient as the search direction. Subsequently,

we perform the economy size Singular Value Decomposition (SVD) on each updated matrix, as outlined below:

$$\left[\tilde{\mathbf{U}}^{(\mathbf{1})}_{\beta\alpha}, \tilde{\mathbf{\Sigma}}^{(\mathbf{1})}_{\beta\alpha}, \tilde{\mathbf{V}}^{(\mathbf{1})}_{\beta\alpha}\right] = SVD\left(\tilde{\mathbf{S}}^{(\mathbf{1})}_{\beta\alpha}\right), \tag{3.53}$$

$$\left[\tilde{\mathbf{U}}^{(\mathbf{2})}_{\alpha\beta}, \tilde{\mathbf{\Sigma}}^{(\mathbf{2})}_{\alpha\beta}, \tilde{\mathbf{V}}^{(\mathbf{2})}_{\alpha\beta}\right] = SVD\left(\tilde{\mathbf{S}}^{(\mathbf{2})}_{\alpha\beta}\right). \tag{3.54}$$

Next, we proceed with the projection to satisfy the constraints specified in (3.45b) and (3.45c).

$$\mathbf{S}^{(\mathbf{1})}_{\beta\alpha} = \tilde{\mathbf{U}}^{(\mathbf{1})}_{\beta\alpha}\left(\tilde{\mathbf{V}}^{(\mathbf{1})}_{\beta\alpha}\right)^{\mathsf{H}}, \tag{3.55}$$

$$\mathbf{S}^{(\mathbf{2})}_{\alpha\beta} = \tilde{\mathbf{U}}^{(\mathbf{2})}_{\alpha\beta}\left(\tilde{\mathbf{V}}^{(\mathbf{2})}_{\alpha\beta}\right)^{\mathsf{H}}. \tag{3.56}$$

The matrices undergo iterative updates until convergence is achieved. The overall algorithm for the transmissive model is given in **Algorithm 4**.

Algorithm 4 Proposed System Design for Transmissive RIS

Input: Q samples of the channel matrices $\mathbf{H}_{\text{b-r}}$ and $\mathbf{H}_{\text{r-u}}$

Initialize $\widehat{\mathbf{S}}^{(1)}_{\beta\alpha_0}$ and $\widehat{\mathbf{S}}^{(2)}_{\alpha\beta_0}$.
for $k = 1, 2, \ldots, I_{\max}$ **do**
  for $q = 1, 2, \ldots, Q$ **do**
    Initialize $\boldsymbol{\Upsilon}_k$.
    Initialize ρ_k and $\mathbf{F}_k$ as described in [46].
    Compute $\boldsymbol{\Upsilon}_k$ and $\mathbf{F}_k$ solving sub-problem 1) via the method in [46].
    Compute $\left[\frac{\partial f(\mathbf{S}^{(2)}_{\alpha\beta,k}, \mathbf{S}^{(1)}_{\beta\alpha,k})}{\partial \mathbf{S}^{(1)}_{\beta\alpha,k}}\right]_q$.
    Compute $\left[\frac{\partial f(\mathbf{S}^{(2)}_{\alpha\beta,k}, \mathbf{S}^{(1)}_{\beta\alpha,k})}{\partial \mathbf{S}^{(2)}_{\alpha\beta,k}}\right]_q$.
  end for
  Compute $\mathbf{G}_{\mathbf{S}^{(1)}_{\beta\alpha,k}}$ using (3.47).
  Compute $\mathbf{G}_{\mathbf{S}^{(2)}_{\alpha\beta,k}}$ using (3.49).
  Compute $\mathbf{S}^{(1)}_{\beta\alpha,k} = \mathbf{S}^{(1)}_{\beta\alpha,k-1} - \mu\left\{\mathbf{G}^*_{\mathbf{S}^{(1)}_{\beta\alpha,k}}\right\}$.
  Compute $\mathbf{S}^{(2)}_{\alpha\beta,k} = \mathbf{S}^{(2)}_{\alpha\beta,k-1} - \mu\left\{\mathbf{G}^*_{\mathbf{S}^{(2)}_{\alpha\beta,k}}\right\}$.
  Compute $\left[\mathbf{U}^{(1)}_{\beta\alpha,k}, \boldsymbol{\Sigma}^{(1)}_{\beta\alpha,k}, \mathbf{V}^{(1)}_{\beta\alpha,k}\right] = SVD\left(\mathbf{S}^{(1)}_{\beta\alpha,k}\right)$.
  Compute $\left[\mathbf{U}^{(2)}_{\alpha\beta,k}, \boldsymbol{\Sigma}^{(2)}_{\alpha\beta,k}, \mathbf{V}^{(2)}_{\alpha\beta,k}\right] = SVD\left(\mathbf{S}^{(2)}_{\alpha\beta,k}\right)$.
  Compute $\widehat{\mathbf{S}}^{(1)}_{\beta\alpha,k} = \mathbf{U}^{(1)}_{\beta\alpha,k}\left(\mathbf{V}^{(1)}_{\beta\alpha,k}\right)^{\mathsf{H}}$.
  Compute $\widehat{\mathbf{S}}^{(2)}_{\alpha\beta,k} = \mathbf{U}^{(2)}_{\alpha\beta,k}\left(\mathbf{V}^{(2)}_{\alpha\beta,k}\right)^{\mathsf{H}}$.
end for

Output: $\rho_{I_{\max}}$, $\mathbf{F}_{I_{\max}}$, $\boldsymbol{\Upsilon}_{I_{\max}}$, $\mathbf{S}^{(1)}_{\beta\alpha,I_{\max}}$, and $\mathbf{S}^{(2)}_{\alpha\beta,I_{\max}}$.

3.3.3 Analysis of Computational Complexity

Reflective RIS setup

Consider the **Algorithm 1**. Initializing the 2-DFT matrices $\mathbf{U}$ and $\mathbf{V}$ requires $\mathcal{O}(M^2)$ operations each, summing up to a total of $2\mathcal{O}(M^2)$. The overall computational complexity of the subsequent inner optimization process, aimed at determining the optimal values for $\boldsymbol{\Upsilon}$, ρ, and $\mathbf{F}$ using the method described in [18], entails a total complexity of $Q\big(3\mathcal{O}(M^3) + 4\mathcal{O}(N^3) + \mathcal{O}(KNM + KN^2)\big)$ [18]. Computing the two partial derivatives inside the inner loop incurs a computational complexity of $2Q\mathcal{O}(M^3)$. Subsequently, updating the two matrices $\boldsymbol{\Sigma}_{\alpha\alpha}$, $\boldsymbol{\Sigma}_{\alpha\beta}$, and ensuring the symmetry of the matrices involves a total complexity of $4\mathcal{O}(M^2)$. Finally, the two matrix-matrix products related to SVD steps require $2\mathcal{O}(M^3)$ operations. Hence, the overall complexity is $I_{\max}\Big(Q\left(3\mathcal{O}(M^3) + 4\mathcal{O}(N^3) + \mathcal{O}(KNM + KN^2)\right) + 6\mathcal{O}(M^2) + 2(Q+1)\mathcal{O}(M^3)\Big)$, where $I_{\max}$ is the number of iterations of outer loop. Note that the dominant complexity factor is $\mathcal{O}(M^3)$, primarily due to calculating the matrix inverse. However, despite this seemingly critical complexity, optimization of the scattering parameters on-the-fly helps mitigate its impact.

Transmissive RIS setup

Consider the **Algorithm 4**. The overall complexity of computing the optimal $\boldsymbol{\Upsilon}$, ρ, and $\mathbf{F}$, as outlined in [46] requires the total complexity of $Q\mathcal{O}(KNM + KN^2)$. Similar to the reflective case, computing partial derivatives takes $2Q\mathcal{O}(M^3)$. Then updating the matrices incurs a complexity of $2\mathcal{O}(M^2)$. Finally calculating the SVD and matrix-matrix multiplication adds a complexity of $4\mathcal{O}(M^3)$. Hence, the overall complexity of the algorithm is $I_{\max}\big(Q\mathcal{O}(KNM + KN^2) + 2\mathcal{O}(M^2) + 2(Q+2)\mathcal{O}(M^3)\big)$.

3.3.4 Practical Limitations

In this design, we have assumed a lossless structure and narrow-band systems. To address losses and broaden its applicability, further investigation from an EM perspective is needed. The design itself is still an open area, but it remains a more reasonable option than existing non-local fully-connected complex designs.

3.4 Numerical Results

In this section, we present simulation results for the proposed scheme. The channels $\mathbf{H}_{\text{b-r}}$ and $\mathbf{H}_{\text{r-u}}$ are characterized using both parametric channel modeling [18, 31] and geometric channel modeling [67, 68] for both RIS models. We present and analyze the simulation results based on these two distinct modeling approaches separately.

In the setup, a uniform linear array with N antennas is considered for the base station (BS), while a square uniform planar array containing M reflective elements serves for the RIS. The channel between the BS and RIS is characterized by [18, 31]

$$\mathbf{H}_{\text{b-r}} = \sqrt{L(d_{\text{RIS}})}\sum_{q=1}^{Q_{\text{b-r}}} \mathsf{c}_q \mathbf{a}_{\text{RIS}}(\varphi_q, \psi_q)\mathbf{a}_{\text{BS}}(\phi_q)^{\mathsf{T}}. \tag{3.57}$$

The channel vectors corresponding to the links between the RIS and the k-th users are as follows [18, 31]:

$$\mathbf{h}_{\text{r-u},k} = \sqrt{L(d'_k)}\sum_{q=1}^{Q_{\text{r-u}}} \mathsf{c}_{k,q}\mathbf{a}_{\text{IRS}}(\varphi_{k,q}, \psi_{k,q}); \quad k = 1, \cdots, K. \tag{3.58}$$

Here, the angles φ and ψ denote the elevation angle and azimuth angle, respectively. The vectors $\mathbf{a}_{\text{RIS}}(\varphi, \psi)$ and $\mathbf{a}_{\text{BS}}(\phi)$ represent the array response of the RIS and the BS [18, 31]. The distance-dependent path-loss factor, represented by the term $L(d)$, is defined as follows [18, 31]:

$$L(d) = C_0(d/d_0)^{\eta}. \tag{3.59}$$

The sum-rate C is the performance metric [18, 31]. Note that, existing non-local RIS structures rely on hardware architectures with higher overhead. Consequently, comparing simulation results between these models and ours may not provide a fair comparison, as both designs are not in the same common ground. Therefore, we consider some other benchmarks for the comparison outlined in the subsequent subsections.

Table 3.1: Simulation parameters

Simulation parameter	**Notation**
Reference distance	$d_0 = 1$ m
RIS-BS distance	$d_{\text{RIS}} = 500$ m
RIS-User distance	$d' \in [10, 50]$ m
Channel path numbers of RIS-BS	$Q_{\text{b-r}} = 8$
Channel path numbers of RIS-user	$Q_{\text{r-u}} = 2$
Channel path gain	$c_q \sim \mathcal{CN}(0, 1)$
RIS-BS, RIS-user, path-loss exponent	$\eta = 2.5$
BS-user path-loss exponent	$\eta = 3.7$
Path-loss at the reference distance	$C_0 = -30$ dB
Noise variance	$\sigma_{\text{w}}^2 = -100$ dBm

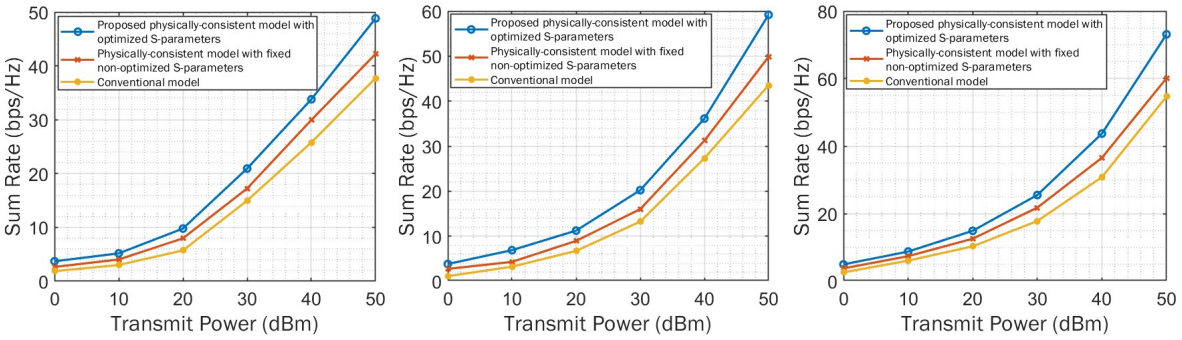
3.4.1 Simulation results for reflective RIS Model

For benchmarking for the reflective RIS model, we consider the following two baselines:

- The physically consistent design of [18, 31] that incorporates MC, but considers a fixed non-optimized scattering matrix $\mathbf{S}_{\alpha\alpha}$ and $\mathbf{S}_{\alpha\beta} = \mathbf{I}_M$.
- The conventional (hypothetical) model described in [46] that does not incorporate MC in the analysis.

Parametric Channel Modeling Setup

In this modeling approach, azimuth and elevation angles are chosen randomly. The azimuth angle ψ_q follows a uniform distribution across the interval $[0, 2\pi]$, while the elevation angle φ_q follows a uniform distribution across the interval $[0, \pi]$. [18, 31]. Table I illustrates the simulation parameters considered [18, 31]. Note that the distances between the BS and RIS, as well as between the RIS and users, were solely utilized for path loss calculations.



(a) 10 channels, $K = 6$, $M = 64$. (b) 10 channels, $K = 8$, $M = 64$. (c) 6 channels, $K = 8$, $M = 100$.

Figure 3.5: Sum-rate performance versus the transmit power P in dBm with $N = 32$ for various system parameters using parametric channel modeling considering reflective RIS.

It can be observed in Fig. 3.5 that there is an improvement in the sum rate when optimizing MC via the proposed offline approach, as compared to the other two cases. Figure 3.6 illustrates the variation of the sum rate versus the M number of RIS elements, and a considerable improvement is observed with the proposed optimization approach. As M increases, the RIS elements need to be placed in close proximity to satisfy a desired surface area. Placing elements closer results in stronger MC. Consequently, an increase in M leads to higher sum-rate performance when optimizing the MC. Therefore, it can be concluded from all simulation results that it suffices for the MC to be optimized

offline to gain from its optimization.

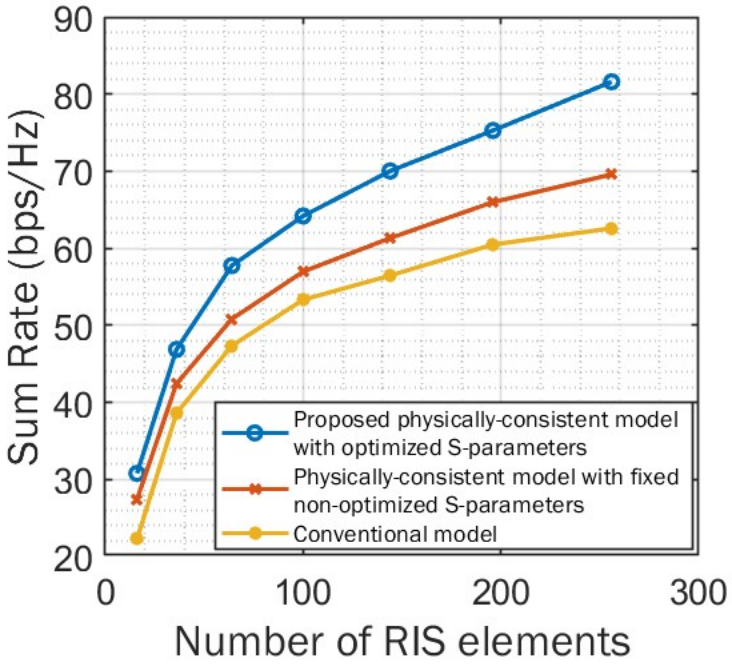


Figure 3.6: Sum-rate performance versus the number of RIS elements M considering 10 channels with $K = 5$, $P = 50$ dBm, and $N = 32$ using parametric channel modeling for reflective RIS.

Geometric Channel Modeling Setup

In this modeling framework, azimuth and elevation angles are computed based on the geometric positions of the BS, RIS, and users, ensuring the angles align with the geometry of the simulation setup. Instead of using a predetermined number of channel paths for the BS-RIS, we introduce multiple BSs. This transforms the setup into a distributed multiple-input multiple-output (MIMO) scenario [69, 70], offering increased flexibility and adaptability. In this scenario, BSs are positioned along the circumference of a circle with a radius of 500m from the RIS and users are situated within an area ranging from

10m to 50m distance from the RIS, as shown in Fig. 3.7. Figures 3.8 and 3.9 show the variation of sum-rate with power for 4 BSs and 8 BSs respectively. As the number of BSs in the setup increases, the capacity to serve more users concurrently improves. Furthermore, we notice an improvement in performance with the proposed model, even when utilizing the geometric model.

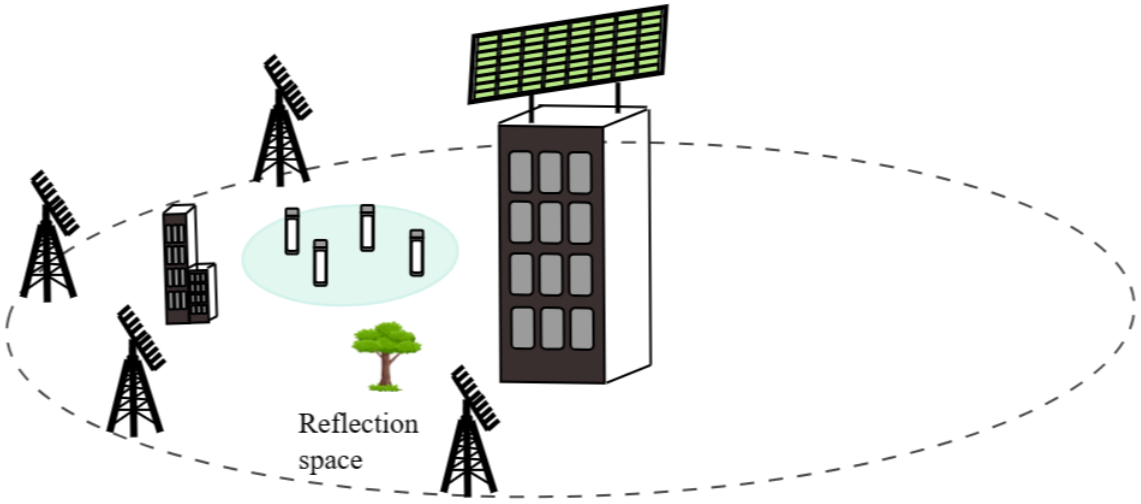
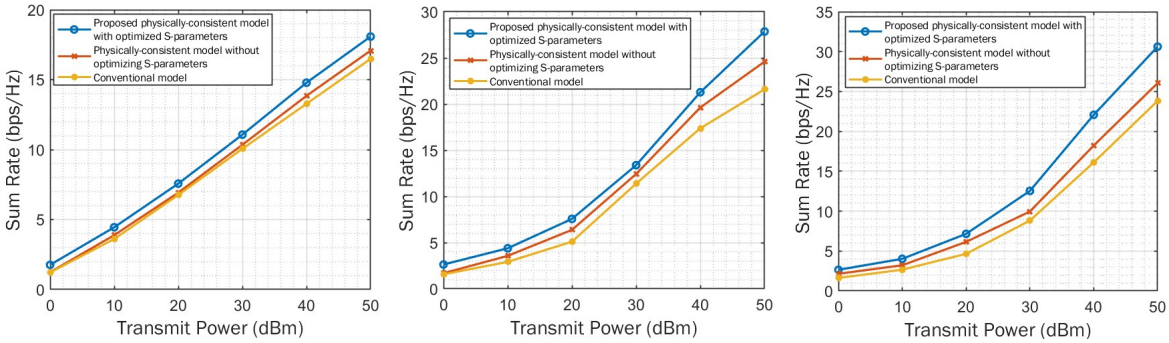
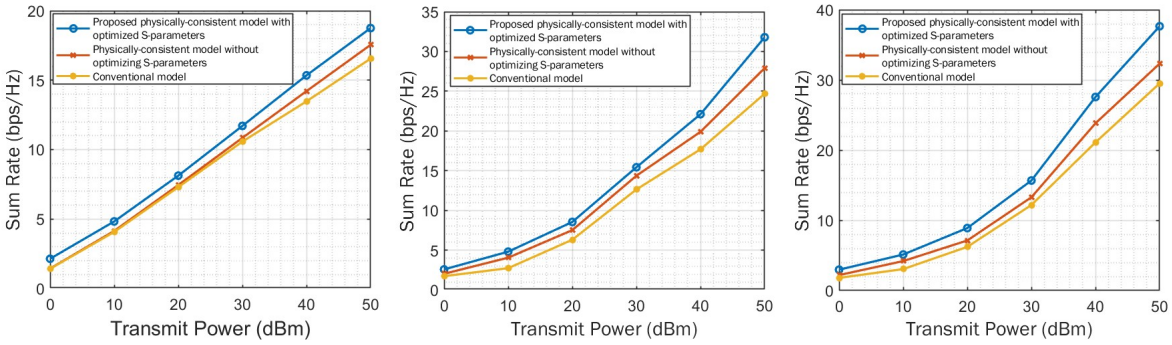


Figure 3.7: Geometric channel modeling setup for reflective RIS model



(a) 10 channels, $K = 1$, $M = 64$. (b) 10 channels, $K = 2$, $M = 64$. (c) 10 channels, $K = 3$, $M = 64$.

Figure 3.8: Sum-rate performance versus the transmit power P in dBm with $N = 32$ considering 4 BSs using geometric channel modeling for reflective RIS



(a) 10 channels, $K = 1$, $M = 64$. (b) 10 channels, $K = 3$, $M = 64$. (c) 10 channels, $K = 4$, $M = 64$.

Figure 3.9: Sum-rate performance versus the transmit power P in dBm with $N = 32$ considering 8 BSs using geometric channel modeling for reflective RIS.

3.4.2 Simulation results for transmissive RIS model

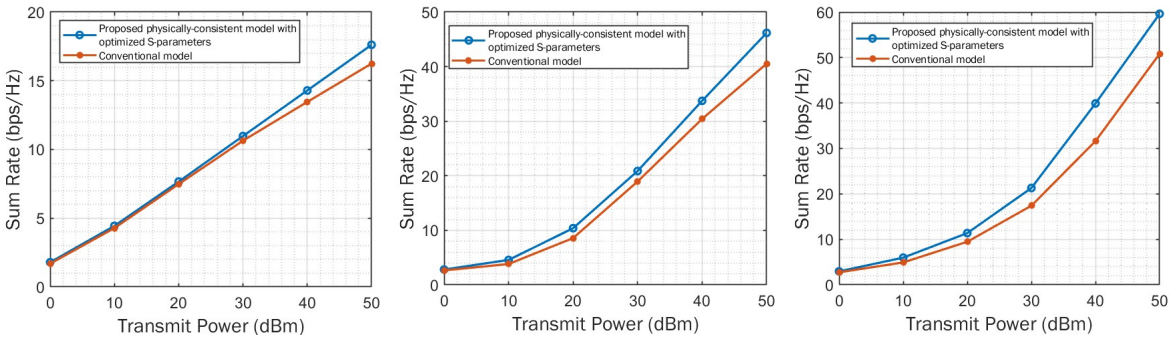
As we assumed no MC in the transmissive model, we use the following benchmarking to compare the results.

- The conventional (hypothetical) model described in [46] that does not incorporate MC in the analysis.

Parametric Channel Modeling

Here, we employ the same simulation parameters as outlined in the table 2.1. Fig. 3.10 shows the variation of sum-rate with transmit power for the transmissive model. We can observe a significant amount of gain in the proposed model compared to the conventional model. Furthermore, a noticeable improvement in the sum-rate is evident when adjusting the number of RIS elements, as illustrated in the Fig. 3.11. Figure 3.12 illustrates the sum-rate variation with the number of users. Notably, the gain rises with the user count up to 8, then declines. This behavior is attributed to the utilization of 8 RIS-BS multi-paths. Consequently, the system cannot support more than 8 users.

Having additional RIS-BS multi-paths would enable support for a larger number of users.



(a) 10 channels, $K = 1$, $M = 64$. (b) 10 channels, $K = 4$, $M = 64$. (c) 10 channels, $K = 8$, $M = 64$.

Figure 3.10: Sum-rate performance versus the transmit power P in dBm with $N = 32$ considering various parameters using parametric channel modeling for transmissive RIS.

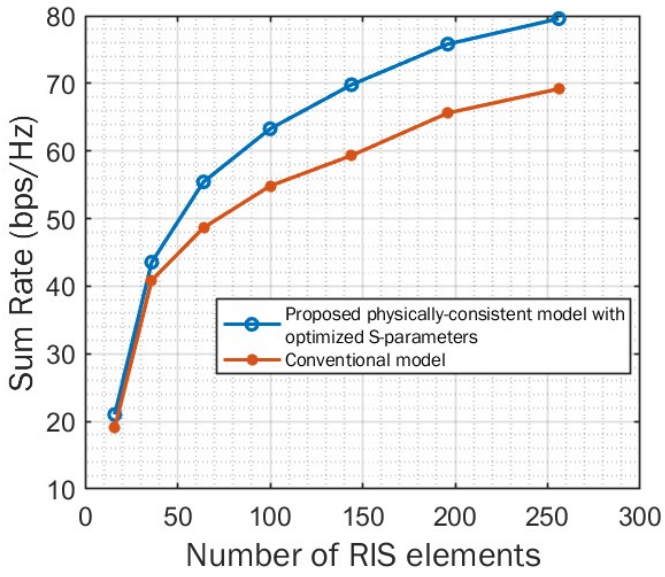


Figure 3.11: Sum-rate performance versus the number of RIS elements M considering 10 channels with $K = 6$, $P = 50$ dBm, and $N = 32$ using parametric channel modeling for transmissive RIS.

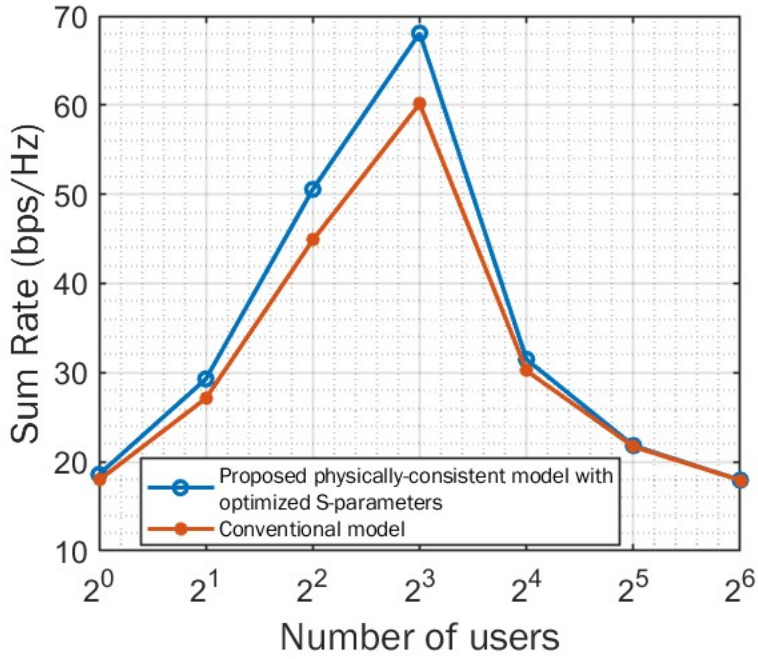


Figure 3.12: Sum-rate performance versus the number users considering 10 channels with $M = 100$, $P = 50$ dBm, and $N = 32$ using parametric channel modeling for transmissive RIS.

Geometric Channel Modeling

Utilizing the identical geometric setup as in the reflective mode, but relocating users to the transmissive space as shown in Fig. 3.13. The results showcased in Fig. 3.14 demonstrate a considerable improvement in the sum-rate with our proposed scheme compared to the hypothetical model.

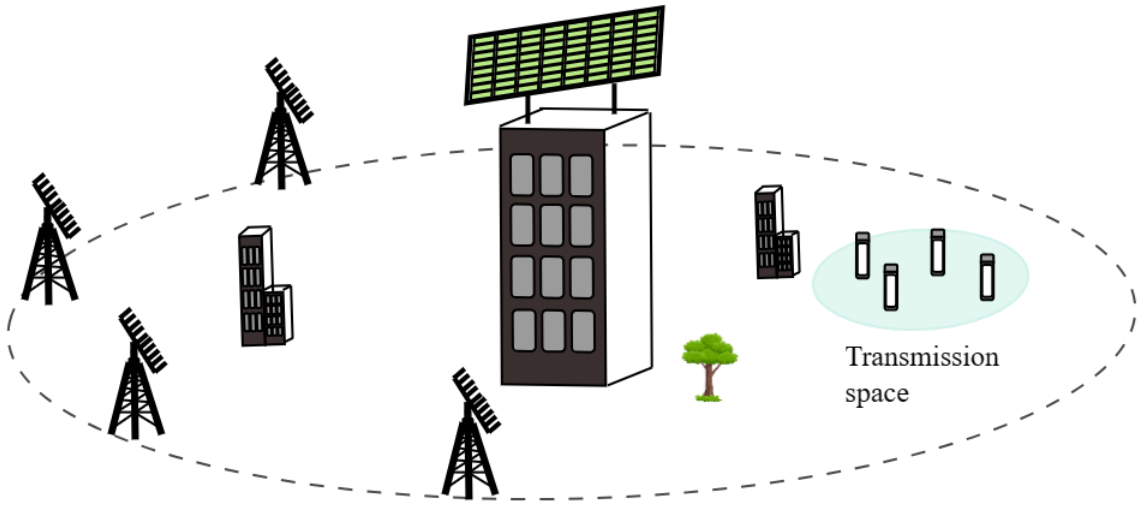
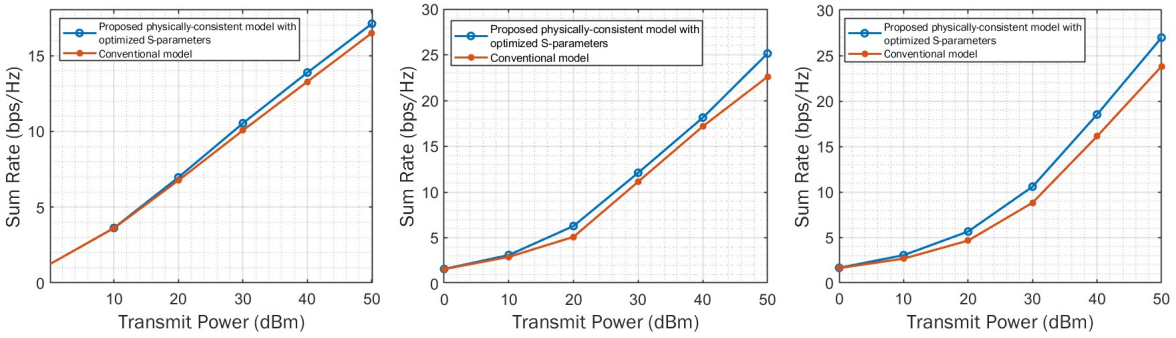


Figure 3.13: Geometric channel modeling setup for transmissive RIS model



(a) 10 channels, $K = 1$, $M = 64$. (b) 10 channels, $K = 2$, $M = 64$. (c) 10 channels, $K = 3$, $M = 64$.

Figure 3.14: Sum-rate performance versus the transmit power P in dBm with $N = 32$ considering 4 BSs using geometric channel modeling for transmissive RIS.

Chapter 4

Conclusion and Future Directions

4.1 Concluding Remarks

Here we conclude the main contributions of the thesis. In the first study, we have proposed an approach to optimize active and passive beamforming jointly in the presence of multiple reflections due to the mutual coupling among the RIS elements for a RIS-assisted wireless communications systems. We have considered a physically-consistent model with both downlink and uplink transmission scenarios. In particular, we have optimized the active precoding matrix and phase shift matrices corresponding to downlink transmission as well as the linear receiver matrix and the phase shift matrix corresponding to uplink transmission. The numerical results have confirmed that the mutual coupling effect is more pronounced in the multi-user setting than the single-user setting due to the multiple reflection effects, which result in multi-user interference. The results imply that the multiple reflection effects are less effective if the phase shifters (or the RIS itself) are lossy and the CSI is imperfect. Moreover, we show that the optimized model with multiple reflections gives better performance even when reducing the RIS element spacing or increasing the RIS size. The proposed scheme for optimized active

and passive beamforming based on a physically-consistent model under multiple reflections gives better performance compared to the conventional models, which ignore the effect of mutual coupling.

In the second study, we have proposed a novel non-local RIS framework by optimizing the RIS phase shifters and BS active precoders using a physically consistent end-to-end channel model incorporating MC and radiation patterns considering reflective and transmissive RIS models separately. We provided distinct solution approaches to reflective and transmissive RIS models. The Mutual Coupling (MC) and radiation patterns have been optimized through an offline optimization method applied to a class of channels, as opposed to state-of-the-art approaches where no MC and radiation pattern optimization are performed. The resultant nested optimization problem has been decomposed into two sub-problems. The RIS phase shifters and BS active precoding were designed using online optimization. Our approach emphasizes the potential of "engineering" MC and radiation patterns to enhance sum-rate performance without requiring on-the-fly adjustments. The simulation results are presented considering parametric and geometric channel modeling. The numerical results validated the effectiveness of the proposed scheme, demonstrating enhanced system performance through active and passive beamforming, incorporating optimized MC and radiation patterns.

4.2 Future Directions

In this section, we discuss new directions for extending our work.

4.2.1 Improving the Problem Formulation Methods and Solution Approaches

In our first study, we employed the gradient descent method as the solution approach due to the complexity of the phase shift matrix, involving two inverse terms. We could investigate alternative solution approaches that offer reduced computational complexity.

In our second study, the problem involves both inner and outer optimization problems, with the outer problem incorporating an expectation operator. Instead, we can explore a method to formulate a separate optimization problem for the outer one that does not involve an expectation operation. This could potentially simplify the optimization process and reduce the computational complexity.

4.2.2 Modulating Intelligent Surfaces (MIS)

In future work, the proposed methodologies from the first and second studies can be applied to a special type of RIS known as Modulating Intelligent Surfaces (MIS) [71]. Here, the RIS serves a dual purpose. Firstly, beamforming signals for a set of users whose signals are already modulated at the base station [71]. Secondly, embedding data intended for distinct sets of users through passively modulating the carrier signals transmitted from the BS to RIS [71]. In the literature, this RIS has utilized with conventional models without accounting for mutual coupling. Therefore, the proposed scheme can be applied to MIS.

4.2.3 Simultaneously Transmitting and Reflecting (STAR) RIS

In our second study, we leveraged the reflective and transmissive RIS separately to formulate problems. However, we can apply our method to a Simultaneously Transmitting and Reflecting (STAR) RIS, where the RIS can both transmit and receive signals at

the same time [72–74]. Using STAR-RIS will lead to a new problem formulation and require a distinct solution approach, which could be quite complex.

4.2.4 Near-Field Communication

In our two studies, we have developed equations and formulas using the far-field approximation. In near-field communication, where the source is located close to the antenna array, the wavefront must be characterized by Maxwell's equations, as the electromagnetic field exhibits complex behavior with significant variations in amplitude and phase [75, 76]. Using near-field probably can be complex, but it has the potential to reveal new insights, and extending our research to include the near-field approximation would represent a significant advancement.

4.2.5 Multi-Layer RIS

Multi-layer RIS is composed of multiple layers of RIS, where each layer contains an array of reflective elements. Due to the multi-layer structure, this provides more degrees of freedom to achieve more effective beamforming [77,78]. Employing physically consistent modeling in the context of multi-layer RIS may be complicated, but it holds the potential for achieving better performance.

Appendix A

Derivations of downlink/uplink optimization problem formulations

In this section, we provide detailed derivations of the problem formulations for downlink and uplink transmissions corresponding to Chapter 2. We have the following minimization problem for the downlink transmission:

$$\underset{\alpha,\mathbf{F},\boldsymbol{\Upsilon}}{\arg\min} \quad \mathbb{E}_{\mathbf{y},\mathbf{s}}\left\{\|\mathbf{y}-\mathbf{s}\|_2^2\right\}, \tag{A.1a}$$

$$\text{subject to} \quad \mathbb{E}_{\mathbf{s}}\left\{\|\mathbf{F}\mathbf{s}\|_2^2\right\} = P, \tag{A.1b}$$

$$\upsilon_{im} = 0, \quad i \neq m, \tag{A.1c}$$

$$|\upsilon_{ii}| = 1, \quad i = 1, 2, \cdots, M. \tag{A.1d}$$

The objective function can be written as

$$\mathbb{E}_{\mathbf{y},\mathbf{s}}\left\{\|\mathbf{y}-\mathbf{s}\|_2^2\right\} = \mathbb{E}_{\mathbf{y},\mathbf{s}}\left\{(\mathbf{y}-\mathbf{s})^{\mathsf{H}}(\mathbf{y}-\mathbf{s})\right\}. \tag{A.2}$$

Since $(\mathbf{y}-\mathbf{s})^{\mathsf{H}}(\mathbf{y}-\mathbf{s}) = \mathrm{Tr}\big((\mathbf{y}-\mathbf{s})(\mathbf{y}-\mathbf{s})^{\mathsf{H}}\big)$, (A.2) can be modified as follows:

$$\mathbb{E}_{\mathbf{y},\mathbf{s}}\left\{\|\mathbf{y}-\mathbf{s}\|_2^2\right\} = \mathbb{E}_{\mathbf{y},\mathbf{s}}\left\{\mathrm{Tr}\big((\mathbf{y}-\mathbf{s})(\mathbf{y}-\mathbf{s})^{\mathsf{H}}\big)\right\}. \tag{A.3}$$

We know that

$$\mathbf{y} = \alpha \left(\mathbf{H}_{\text{s-u}}^{\mathsf{H}} (\boldsymbol{\Upsilon}^{-1} - \mathbf{S})^{-1} \mathbf{H}_{\text{b-s}} \mathbf{F}\mathbf{s} + \mathbf{H}_{\text{b-u}}^{\mathsf{H}} \mathbf{F}\mathbf{s} + \ \mathbf{w} \right), \tag{A.4}$$

which can be written as

$$\mathbf{y} = \alpha \left(\mathbf{H}^{\mathsf{H}} \mathbf{F}\mathbf{s} + \ \mathbf{w} \right), \tag{A.5}$$

where $\mathbf{H}^{\mathsf{H}} = \mathbf{H}_{\text{s-u}}^{\mathsf{H}} (\boldsymbol{\Upsilon}^{-1} - \mathbf{S})^{-1} \mathbf{H}_{\text{b-s}} + \mathbf{H}_{\text{b-u}}^{\mathsf{H}}$. The vectors $\mathbf{y} - \mathbf{s}$ and $(\mathbf{y} - \mathbf{s})^{\mathsf{H}}$ are given by (A.6) and (A.7) as follows:

$$\mathbf{y} - \mathbf{s} = \alpha \mathbf{H}^{\mathsf{H}} \mathbf{F}\mathbf{s} + \ \alpha \mathbf{w} - \mathbf{s}, \tag{A.6}$$

$$(\mathbf{y} - \mathbf{s})^{\mathsf{H}} = \alpha \mathbf{s}^{\mathsf{H}} \mathbf{F}^{\mathsf{H}} \mathbf{H} + \ \alpha \mathbf{w}^{\mathsf{H}} - \mathbf{s}^{\mathsf{H}}. \tag{A.7}$$

By using (A.6) and (A.7) the expectation function $\mathbb{E}_{\mathbf{y},\mathbf{s}} \left\{ \text{Tr}\big((\mathbf{y} - \mathbf{s})(\mathbf{y} - \mathbf{s})^{\mathsf{H}} \big) \right\}$ can be modified as follows with respect to (w.r.t.) the random vectors $\mathbf{s}$ and $\mathbf{w}$:

$$\begin{aligned}
&\mathbb{E}_{\mathbf{w},\mathbf{s}} \Big\{ \text{Tr}\big((\alpha \mathbf{H}^{\mathsf{H}} \mathbf{F}\mathbf{s} + \ \alpha \mathbf{w} - \mathbf{s})(\alpha \mathbf{s}^{\mathsf{H}} \mathbf{F}^{\mathsf{H}} \mathbf{H} + \ \alpha \mathbf{w}^{\mathsf{H}} - \mathbf{s}^{\mathsf{H}}) \big) \Big\} \\
&= \ \mathbb{E}_{\mathbf{w},\mathbf{s}} \Big\{ \text{Tr}\big(\alpha^2 \mathbf{H}^{\mathsf{H}} \mathbf{F}\mathbf{s}\mathbf{s}^{\mathsf{H}} \mathbf{F}^{\mathsf{H}} \mathbf{H} + \alpha^2 \mathbf{H}^{\mathsf{H}} \mathbf{F}\mathbf{s}\mathbf{w}^{\mathsf{H}} \\
&- \alpha \mathbf{H}^{\mathsf{H}} \mathbf{F}\mathbf{s}\mathbf{s}^{\mathsf{H}} - \alpha \mathbf{s}\mathbf{s}^{\mathsf{H}} \mathbf{F}^{\mathsf{H}} \mathbf{H} - \alpha \mathbf{s}\mathbf{w}^{\mathsf{H}} + \mathbf{s}\mathbf{s}^{\mathsf{H}} + \alpha^2 \mathbf{w}\mathbf{s}^{\mathsf{H}} \mathbf{F}^{\mathsf{H}} \mathbf{H} \\
&+ \alpha^2 \mathbf{w}\mathbf{w}^{\mathsf{H}} - \alpha \mathbf{w}\mathbf{s}^{\mathsf{H}} \big) \Big\} \\
&= \ \text{Tr}\big(\alpha^2 \mathbf{H}^{\mathsf{H}} \mathbf{F}\mathbf{F}^{\mathsf{H}} \mathbf{H} - \alpha \mathbf{H}^{\mathsf{H}} \mathbf{F} - \alpha \mathbf{F}^{\mathsf{H}} \mathbf{H} + \mathbf{I}_K \big) \\
&+ \alpha^2 \, \mathbb{E}_{\mathbf{w}} \Big\{ \text{Tr}\big(\mathbf{w}\mathbf{w}^{\mathsf{H}} \big) \Big\} \\
&= \ \text{Tr}\big((\alpha \mathbf{H}^{\mathsf{H}} \mathbf{F} - \mathbf{I}_K)(\alpha \mathbf{F}^{\mathsf{H}} \mathbf{H} - \mathbf{I}_K)^{\mathsf{H}} \big) + K \alpha^2 \sigma_{\text{w}}^2 \\
&= \ \big\| (\alpha \mathbf{H}^{\mathsf{H}} \mathbf{F} - \mathbf{I}_K) \big\|_{\text{F}}^2 + K \alpha^2 \sigma_{\text{w}}^2.
\end{aligned} \tag{A.8}$$

Hence, the resultant objective function is

$$\mathbb{E}_{\mathbf{y},\mathbf{s}} \left\{ \|\mathbf{y} - \mathbf{s}\|_2^2 \right\} = \big\| \alpha \mathbf{H}^{\mathsf{H}} \mathbf{F} - \mathbf{I}_K \big\|_{\text{F}}^2 + K \alpha^2 \sigma_{\text{w}}^2. \tag{A.9}$$

Note that, since we assume the input symbols and the noise to be Gaussian: $\mathsf{s}_k \sim \mathcal{CN}(s; 0, 1)$, $\mathsf{w} \sim \mathcal{CN}(w; 0, \sigma_{\text{w}}^2)$, the expectation terms $\mathbb{E}_{\mathbf{w},\mathbf{s}} \left\{ \mathbf{s} \right\} = 0$, $\mathbb{E}_{\mathbf{w},\mathbf{s}} \left\{ \mathbf{w} \right\} = 0$, and

$\mathbb{E}_{\mathbf{w},\mathbf{s}}\left\{\mathbf{s}\mathbf{s}^{\mathsf{H}}\right\} = \mathbf{I}_K$.

Now consider the uplink transmission scenario. The optimization problem formulation for the uplink transmission is

$$\underset{\mathbf{G},\boldsymbol{\Upsilon}}{\arg\min} \quad \mathbb{E}_{\mathbf{s},\hat{\mathbf{s}}}\left\{\|\mathbf{s}-\hat{\mathbf{s}}\|_2^2\right\}, \tag{A.10a}$$

$$\text{subject to} \quad \upsilon_{im} = 0, \quad i \neq m, \tag{A.10b}$$

$$|\upsilon_{ii}| = 1, \quad i = 1, 2, \cdots, M. \tag{A.10c}$$

The objective function can be given by

$$\mathbb{E}_{\mathbf{s},\hat{\mathbf{s}}}\left\{\|\mathbf{s}-\hat{\mathbf{s}}\|_2^2\right\} = \mathbb{E}_{\mathbf{s},\hat{\mathbf{s}}}\left\{(\mathbf{s}-\hat{\mathbf{s}})^{\mathsf{H}}(\mathbf{s}-\hat{\mathbf{s}})\right\}. \tag{A.11}$$

Since $(\mathbf{s}-\hat{\mathbf{s}})^{\mathsf{H}}(\mathbf{s}-\hat{\mathbf{s}}) = \mathrm{Tr}\big((\mathbf{s}-\hat{\mathbf{s}})(\mathbf{s}-\hat{\mathbf{s}})^{\mathsf{H}}\big)$, equation (A.11) can be modified as follows:

$$\mathbb{E}_{\mathbf{s},\hat{\mathbf{s}}}\left\{\|\mathbf{s}-\hat{\mathbf{s}}\|_2^2\right\} = \mathbb{E}_{\mathbf{s},\hat{\mathbf{s}}}\left\{\mathrm{Tr}\big((\mathbf{s}-\hat{\mathbf{s}})(\mathbf{s}-\hat{\mathbf{s}})^{\mathsf{H}}\big)\right\}. \tag{A.12}$$

Here,

$$\hat{\mathbf{s}} = \mathbf{G}\mathbf{y}. \tag{A.13}$$

For uplink transmission

$$\mathbf{y} = \mathbf{H}_{\text{s-b}}(\boldsymbol{\Upsilon}^{-1} - \mathbf{S})^{-1}\mathbf{H}_{\text{u-s}}\mathbf{s} + \mathbf{H}_{\text{u-b}}\mathbf{s} + \ \mathbf{w}. \tag{A.14}$$

Using equation (A.13) and (A.14), we have

$$\hat{\mathbf{s}} = \mathbf{G}(\mathbf{H}\mathbf{s} + \mathbf{w}), \tag{A.15}$$

where $\mathbf{H} = \mathbf{H}_{\text{s-b}}(\boldsymbol{\Upsilon}^{-1} - \mathbf{S})^{-1}\mathbf{H}_{\text{u-s}}\mathbf{s} + \mathbf{H}_{\text{u-b}}\mathbf{s}$. Substituting in (A.12), we have

$$\begin{aligned}
&\mathbb{E}_{\mathbf{w},\mathbf{s}}\left\{\mathrm{Tr}\big((\mathbf{s} - \mathbf{G}\mathbf{H}\mathbf{s} - \mathbf{G}\mathbf{w})(\mathbf{s}^{\mathsf{H}} - \mathbf{s}^{\mathsf{H}}\mathbf{H}^{\mathsf{H}}\mathbf{G}^{\mathsf{H}} - \mathbf{w}^{\mathsf{H}}\mathbf{G}^{\mathsf{H}})\big)\right\} \\
&= \mathbb{E}_{\mathbf{w},\mathbf{s}}\Big\{\mathrm{Tr}\big(\mathbf{s}\mathbf{s}^{\mathsf{H}} - \mathbf{s}\mathbf{s}^{\mathsf{H}}\mathbf{H}^{\mathsf{H}}\mathbf{G}^{\mathsf{H}} - \mathbf{s}\mathbf{w}^{\mathsf{H}}\mathbf{G}^{\mathsf{H}} - \mathbf{G}\mathbf{H}\mathbf{s}\mathbf{s}^{\mathsf{H}} - \mathbf{G}\mathbf{w}\mathbf{s}^{\mathsf{H}} \\
&+ \mathbf{G}\mathbf{H}\mathbf{s}\mathbf{s}^{\mathsf{H}}\mathbf{H}^{\mathsf{H}}\mathbf{G}^{\mathsf{H}} + \mathbf{G}\mathbf{H}\mathbf{s}\mathbf{w}^{\mathsf{H}}\mathbf{G}^{\mathsf{H}} + \mathbf{G}\mathbf{w}\mathbf{s}^{\mathsf{H}}\mathbf{H}^{\mathsf{H}}\mathbf{G}^{\mathsf{H}} + \mathbf{G}\mathbf{w}\mathbf{w}^{\mathsf{H}}\mathbf{G}^{\mathsf{H}}\big)\Big\} \\
&= \mathrm{Tr}\big(\mathbb{E}_{\mathbf{w},\mathbf{s}}\left\{\mathbf{s}\mathbf{s}^{\mathsf{H}}\right\} - \mathbb{E}_{\mathbf{w},\mathbf{s}}\left\{\mathbf{s}\mathbf{s}^{\mathsf{H}}\right\}\mathbf{H}^{\mathsf{H}}\mathbf{G}^{\mathsf{H}} - \mathbf{G}\mathbf{H}\,\mathbb{E}_{\mathbf{w},\mathbf{s}}\left\{\mathbf{s}\mathbf{s}^{\mathsf{H}}\right\} \\
&+ \mathbf{G}\mathbf{H}\,\mathbb{E}_{\mathbf{w},\mathbf{s}}\left\{\mathbf{s}\mathbf{s}^{\mathsf{H}}\right\}\mathbf{H}^{\mathsf{H}}\mathbf{G}^{\mathsf{H}} + \mathbf{G}\,\mathbb{E}_{\mathbf{w},\mathbf{s}}\left\{\mathbf{w}\mathbf{w}^{\mathsf{H}}\right\}\mathbf{G}^{\mathsf{H}}\big) \\
&= \mathrm{Tr}\big(p_k\mathbf{I}_K - p_k\mathbf{I}_K\mathbf{H}^{\mathsf{H}}\mathbf{G}^{\mathsf{H}} - p_k\mathbf{G}\mathbf{H}\mathbf{I}_K + \mathbf{G}\mathbf{H}\mathbf{I}_K\mathbf{H}^{\mathsf{H}}\mathbf{G}^{\mathsf{H}}\big) \\
&+ \sigma_{\mathrm{w}}^2\mathrm{Tr}\big(\mathbf{G}\mathbf{G}^{\mathsf{H}}\big) \\
&= p_k\left\|\mathbf{I}_K - \mathbf{G}\mathbf{H}\right\|_{\mathrm{F}}^2 + \sigma_{\mathrm{w}}^2\left\|\mathbf{G}\right\|_{\mathrm{F}}^2.
\end{aligned} \tag{A.16}$$

Here we consider $\mathbb{E}_{\mathbf{w},\mathbf{s}}\left\{\mathbf{s}\mathbf{s}^{\mathsf{H}}\right\} = p_k\mathbf{I}_K$, where p_k is the allocated power for k-th user. Thus, the objective function related to uplink transmission is given by

$$\mathbb{E}_{\mathbf{s},\hat{\mathbf{s}}}\left\{\left\|\mathbf{s} - \hat{\mathbf{s}}\right\|_2^2\right\} = p_k\left\|\mathbf{I}_K - \mathbf{G}\mathbf{H}\right\|_{\mathrm{F}}^2 + \sigma_{\mathrm{w}}^2\left\|\mathbf{G}\right\|_{\mathrm{F}}^2. \tag{A.17}$$

Bibliography

[1] E. G. Larsson, O. Edfors, F. Tufvesson, and T. L. Marzetta, "Massive MIMO for next generation wireless systems," *IEEE Communications Magazine*, vol. 52, no. 2, pp. 186–195, 2014.

[2] T. Abood, I. Hburi, and H. F. Khazaal, "Massive MIMO: An Overview, Recent Challenges, and Future Research Directions," in *2021 International Conference on Advance of Sustainable Engineering and its Application (ICASEA)*, pp. 43–48, 2021.

[3] L. Lu, G. Y. Li, A. L. Swindlehurst, A. Ashikhmin, and R. Zhang, "An Overview of Massive MIMO: Benefits and Challenges," *IEEE Journal of Selected Topics in Signal Processing*, vol. 8, no. 5, pp. 742–758, 2014.

[4] S. Buzzi, C.-L. I, T. E. Klein, H. V. Poor, C. Yang, and A. Zappone, "A Survey of Energy-Efficient Techniques for 5G Networks and Challenges Ahead," *IEEE Journal on Selected Areas in Communications*, vol. 34, no. 4, pp. 697–709, 2016.

[5] X. Wang, L. Kong, F. Kong, F. Qiu, M. Xia, S. Arnon, and G. Chen, "Millimeter Wave Communication: A Comprehensive Survey," *IEEE Communications Surveys & Tutorials*, vol. 20, no. 3, pp. 1616–1653, 2018.

[6] J. Bae and Y. S. Choi, "System capacity enhancement of MmWave based 5G mobile communication system," in *2015 International Conference on Information and Communication Technology Convergence (ICTC)*, pp. 291–294, 2015.

[7] S. Patchala and M. Sailaja, "Improvement Spectral Efficiency in mmWave MIMO system for 5G communications," in *2019 3rd International Conference on Computing Methodologies and Communication (ICCMC)*, pp. 540–544, 2019.

[8] N. K. Mallat, M. Ishtiaq, A. U. Rehman, and A. Iqbal, "Millimeter-Wave in the Face of 5G Communication Potential Applications," *IETE Journal of Research*, vol. 68, no. 4, pp. 2522–2530, 2022.

[9] Q. Wu and R. Zhang, "Intelligent reflecting surface enhanced wireless network via joint active and passive beamforming," *IEEE Transactions on Wireless Communications*, vol. 18, no. 11, pp. 5394–5409, 2019.

[10] Q. Wu and R. Zhang, "Towards smart and reconfigurable environment: Intelligent reflecting surface aided wireless network," *IEEE Communications Magazine*, vol. 58, no. 1, pp. 106–112, 2020.

[11] C. Huang, A. Zappone, G. C. Alexandropoulos, M. Debbah, and C. Yuen, "Reconfigurable intelligent surfaces for energy efficiency in wireless communication," *IEEE Transactions on Wireless Communications*, vol. 18, no. 8, pp. 4157–4170, 2019.

[12] E. Basar, M. Di Renzo, J. De Rosny, M. Debbah, M.-S. Alouini, and R. Zhang, "Wireless communications through reconfigurable intelligent surfaces," *IEEE Access*, vol. 7, pp. 116753–116773, 2019.

[13] Q. Tao, J. Wang, and C. Zhong, "Performance analysis of intelligent reflecting surface aided communication systems," *IEEE Communications Letters*, vol. 24, no. 11, pp. 2464–2468, 2020.

[14] B. G. Kashyap, P. C. Theofanopoulos, A. S. Shekhawat, A. Y. Modi, A. P. Sengar, S. K. Kumar, A. Chang, T. Osman, A. Alkhateeb, and G. C. Trichopoulos, "A reconfigurable intelligent surface for 5g wireless communication applications," in *2021 IEEE International Symposium on Antennas and Propagation and USNC-URSI Radio Science Meeting (APS/URSI)*, pp. 111–112, 2021.

[15] M. S. Ayub, P. Adasme, I. Soto, and D. Z. Rodriguez, "Reconfigurable intelligent surfaces enabling future wireless communication," in *2021 Third South American Colloquium on Visible Light Communications (SACVLC)*, pp. 1–5, 2021.

[16] E. Björnson, H. Wymeersch, B. Matthiesen, P. Popovski, L. Sanguinetti, and E. de Carvalho, "Reconfigurable Intelligent Surfaces: A signal processing perspective with wireless applications," *IEEE Signal Processing Magazine*, vol. 39, no. 2, pp. 135–158, 2022.

[17] C. Pan, H. Ren, K. Wang, W. Xu, M. Elkashlan, A. Nallanathan, and L. Hanzo, "Multicell MIMO Communications Relying on Intelligent Reflecting Surfaces," *IEEE Transactions on Wireless Communications*, vol. 19, no. 8, pp. 5218–5233, 2020.

[18] D. Wijekoon, A. Mezghani, and E. Hossain, "Phase Shifter Optimization in RIS-Aided MIMO Systems Under Multiple Reflections," *IEEE Transactions on Wireless Communications*, pp. 1–1, 2024.

[19] P. Wang, J. Fang, X. Yuan, Z. Chen, and H. Li, "Intelligent Reflecting Surface-Assisted Millimeter Wave Communications: Joint Active and Passive Precoding

Design," *IEEE Transactions on Vehicular Technology*, vol. 69, no. 12, pp. 14960–14973, 2020.

[20] H. Cao, Z. Li, and W. Chen, "Resource Allocation for IRS-Assisted Wireless Powered Communication Networks," *IEEE Wireless Communications Letters*, vol. 10, no. 11, pp. 2450–2454, 2021.

[21] Z. Li, W. Chen, Q. Wu, K. Wang, and J. Li, "Joint Beamforming Design and Power Splitting Optimization in IRS-Assisted SWIPT NOMA Networks," *IEEE Transactions on Wireless Communications*, vol. 21, no. 3, pp. 2019–2033, 2022.

[22] Z. Li, H. Hu, J. Zhang, and J. Zhang, "Coverage Analysis of Multiple Transmissive RIS-Aided Outdoor-to-Indoor mmWave Networks," *IEEE Transactions on Broadcasting*, vol. 68, no. 4, pp. 935–942, 2022.

[23] Y. Sun, K. An, C. Li, Z. Lin, H. Niu, D. W. K. Ng, J. Wang, and N. Al-Dhahir, "Joint Transmissive and Reflective RIS-Aided Secure MIMO Systems Design Under Spatially-Correlated Angular Uncertainty and Coupled PSEs," *IEEE Transactions on Information Forensics and Security*, vol. 18, pp. 3606–3621, 2023.

[24] S. Zeng, H. Zhang, B. Di, Y. Tan, Z. Han, H. V. Poor, and L. Song, "Reconfigurable Intelligent Surfaces in 6G: Reflective, Transmissive, or Both?," *IEEE Communications Letters*, vol. 25, no. 6, pp. 2063–2067, 2021.

[25] Z. Li, W. Chen, Q. Wu, X. Zhu, H. Qin, K. Wang, and J. Li, "Toward Transmissive RIS Transceiver Enabled Uplink Communication Systems: Design and Optimization," *IEEE Internet of Things Journal*, vol. 11, no. 4, pp. 6788–6801, 2024.

[26] J. Tang, M. Cui, S. Xu, L. Dai, F. Yang, and M. Li, "Transmissive RIS for B5G Communications: Design, Prototyping, and Experimental Demonstrations," *IEEE Transactions on Communications*, vol. 71, no. 11, pp. 6605–6615, 2023.

[27] J. Zhang, R. Xiong, J. Liu, T. Mi, and R. C. Qiu, "Design and Prototyping of Transmissive RIS-Aided Wireless Communication," 2024.

[28] S. Sun and M. Tao, "Characteristics of Channel Eigenvalues and Mutual Coupling Effects for Holographic Reconfigurable Intelligent Surfaces," *Sensors (Basel, Switzerland)*, vol. 22, no. 14, pp. 5297–, 2022.

[29] X. Qian and M. D. Renzo, "Mutual Coupling and Unit Cell Aware Optimization for Reconfigurable Intelligent Surfaces," *IEEE Wireless Communications Letters*, vol. 10, no. 6, pp. 1183–1187, 2021.

[30] R. Faqiri, C. Saigre-Tardif, G. C. Alexandropoulos, N. Shlezinger, M. F. Imani, and P. del Hougne, "PhysFad: Physics-based end-to-end channel modeling of RIS-parametrized environments with adjustable fading," *IEEE Trans. Wireless Commun.*, vol. 22, no. 1, pp. 580–595, 2023.

[31] D. Wijekoon, A. Mezghani, and E. Hossain, "Beamforming Optimization in RIS-Aided MIMO Systems Under Multiple-Reflection Effects," in *ICASSP 2023 - 2023 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pp. 1–5, 2023.

[32] X. Chen, S. Zhang, and Q. Li, "A Review of Mutual Coupling in MIMO Systems," *IEEE Access*, vol. 6, pp. 24706–24719, 2018.

[33] J. Wallace and M. Jensen, "Mutual coupling in MIMO wireless systems: a rigorous network theory analysis," *IEEE Transactions on Wireless Communications*, vol. 3, no. 4, pp. 1317–1325, 2004.

[34] T. Svantesson and A. Ranheim, "Mutual coupling effects on the capacity of multielement antenna systems," in *2001 IEEE International Conference on Acoustics,*

Speech, and Signal Processing. Proceedings (Cat. No.01CH37221), vol. 4, pp. 2485–2488 vol.4, 2001.

[35] A. Rabault, L. L. Magoarou, J. Sol, G. C. Alexandropoulos, N. Shlezinger, H. V. Poor, and P. del Hougne, "On the tacit linearity assumption in common cascaded models of ris-parametrized wireless channels," 2023.

[36] A. Mezghani, F. Bellili, and E. Hossain, "Reconfigurable Intelligent Surfaces for Quasi-Passive mmWave and THz Networks: Should They be Reflective or Redirective?," in *2022 56th Asilomar Conference on Signals, Systems, and Computers*, pp. 1076–1080, 2022.

[37] A. Abrardo, D. Dardari, M. Di Renzo, and X. Qian, "MIMO Interference Channels Assisted by Reconfigurable Intelligent Surfaces: Mutual Coupling Aware Sum-Rate Optimization Based on a Mutual Impedance Channel Model," *IEEE Wireless Communications Letters*, vol. 10, no. 12, pp. 2624–2628, 2021.

[38] G. Gradoni and M. Di Renzo, "End-to-end mutual coupling aware communication model for reconfigurable intelligent surfaces: An electromagnetic-compliant approach based on mutual impedances," *IEEE Wireless Communications Letters*, vol. 10, no. 5, pp. 938–942, 2021.

[39] A. Abrardo, A. Toccafondi, and M. Di Renzo, "Analysis and Optimization of Reconfigurable Intelligent Surfaces Based on S-Parameters Multiport Network Theory," *arXiv preprint arXiv:2308.16856*, 2023.

[40] D. M. Pozar, *Microwave engineering*. J. Wiley, 4th ed. ed., 2011.

[41] H. Li, S. Shen, and B. Clerckx, "Beyond Diagonal Reconfigurable Intelligent Surfaces: From Transmitting and Reflecting Modes to Single-, Group-, and

Fully-Connected Architectures," *IEEE Trans. Wireless Commun.*, vol. 22, no. 4, pp. 2311–2324, 2023.

[42] Q. Li, M. El-Hajjar, I. Hemadeh, A. Shojaeifard, A. A. M. Mourad, B. Clerckx, and L. Hanzo, "Reconfigurable Intelligent Surfaces Relying on Non-Diagonal Phase Shift Matrices," *IEEE Trans. Veh. Technol.*, vol. 71, no. 6, pp. 6367–6383, 2022.

[43] S. Abeywickrama, R. Zhang, and C. Yuen, "Intelligent reflecting surface: Practical phase shift model and beamforming optimization," in *ICC 2020 - 2020 IEEE International Conference on Communications (ICC)*, pp. 1–6, 2020.

[44] P. Wang, J. Fang, and H. Li, "Joint beamforming for intelligent reflecting surface-assisted millimeter wave communications," *arXiv.org*, 2019.

[45] X. Yu, D. Xu, and R. Schober, "MISO Wireless Communication Systems via Intelligent Reflecting Surfaces : (Invited Paper)," in *2019 IEEE/CIC International Conference on Communications in China (ICCC)*, pp. 735–740, 2019.

[46] H. Ur Rehman, F. Bellili, A. Mezghani, and E. Hossain, "Joint Active and Passive Beamforming Design for IRS-Assisted Multi-User MIMO Systems: A VAMP-Based Approach," *IEEE Transactions on Communications*, vol. 69, no. 10, pp. 6734–6749, 2021.

[47] M. Z. Siddiqi, T. Mir, M. Hao, and R. MacKenzie, "Low-complexity joint active and passive beamforming for RIS-aided MIMO systems," *Electronics letters*, vol. 57, no. 9, pp. 384–386, 2021.

[48] J. Wang, H. Wang, Y. Han, S. Jin, and X. Li, "Joint Transmit Beamforming and Phase Shift Design for Reconfigurable Intelligent Surface Assisted MIMO Systems," *IEEE Transactions on Cognitive Communications and Networking*, vol. 7, no. 2, pp. 354–368, 2021.

[49] X. Gan, C. Zhong, C. Huang, Z. Yang, and Z. Zhang, "Multiple RISs Assisted Cell-Free Networks With Two-Timescale CSI: Performance Analysis and System Design," *IEEE Transactions on Communications*, vol. 70, no. 11, pp. 7696–7710, 2022.

[50] C. Huang, Z. Yang, G. C. Alexandropoulos, K. Xiong, L. Wei, C. Yuen, Z. Zhang, and M. Debbah, "Multi-Hop RIS-Empowered Terahertz Communications: A DRL-Based Hybrid Beamforming Design," *IEEE Journal on Selected Areas in Communications*, vol. 39, no. 6, pp. 1663–1677, 2021.

[51] M.-W. Un, W.-K. Ma, and P. C. Ching, "Joint transmit beamforming optimization and uplink/downlink user selection in a full-duplex multi-user MIMO system," in *2017 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pp. 3639–3643, 2017.

[52] W. Rhee, W. Yu, and J. Cioffi, "The optimality of beamforming in uplink multiuser wireless systems," *IEEE Transactions on Wireless Communications*, vol. 3, no. 1, pp. 86–96, 2004.

[53] H. Guo and V. K. N. Lau, "Uplink Cascaded Channel Estimation for Intelligent Reflecting Surface Assisted Multiuser MISO Systems," *IEEE Transactions on Signal Processing*, vol. 70, pp. 3964–3977, 2022.

[54] D. P. Bertsekas, "Nonlinear programming," *Journal of the Operational Research Society*, vol. 48, no. 3, pp. 334–334, 1997.

[55] S. Ruder, "An overview of gradient descent optimization algorithms," 2016.

[56] Y. Chen, C. Jin, and B. Yu, "Stability and convergence trade-off of iterative optimization algorithms," 2018.

[57] R. W. Heath Jr. and A. Lozano, *Foundations of MIMO Communication*. Cambridge University Press, 2018.

[58] Q. Li, Z. Zhu, and G. Tang, "Alternating Minimizations Converge to Second-Order Optimal Solutions," in *Proceedings of the 36th International Conference on Machine Learning* (K. Chaudhuri and R. Salakhutdinov, eds.), vol. 97 of *Proceedings of Machine Learning Research*, pp. 3935–3943, PMLR, 09–15 Jun 2019.

[59] X. Yi, C. Caramanis, and S. Sanghavi, "Alternating Minimization for Mixed Linear Regression," in *Proceedings of the 31st International Conference on Machine Learning* (E. P. Xing and T. Jebara, eds.), vol. 32 of *Proceedings of Machine Learning Research*, (Bejing, China), pp. 613–621, PMLR, 22–24 Jun 2014.

[60] W. Huang, Q. Si, and M. Jin, "Alternating Optimization Based Low Complexity Hybrid Precoding in Millimeter Wave MIMO Systems," *IEEE Communications Letters*, vol. 24, no. 3, pp. 635–638, 2020.

[61] M. Joham, W. Utschick, and J. Nossek, "Linear transmit processing in MIMO communications systems," *IEEE Transactions on Signal Processing*, vol. 53, no. 8, pp. 2700–2712, 2005.

[62] X. Wang, L. Yan, and Q. Zhang, "Research on the application of gradient descent algorithm in machine learning," in *2021 International Conference on Computer Network, Electronic and Automation (ICCNEA)*, pp. 11–15, 2021.

[63] E. K. P. Chong, *An introduction to optimization* . Wiley series in discrete mathematics and optimization, Hoboken, New Jersey: Wiley, fourth edition. ed., 2013.

[64] P. Hannan, "The element-gain paradox for a phased-array antenna," *IEEE Transactions on Antennas and Propagation*, vol. 12, no. 4, pp. 423–433, 1964.

[65] J. Vargas, J. A. Quiroga, C. O. S. Sorzano, J. C. Estrada, and J. M. Carazo, "Two-step demodulation based on the gram–schmidt orthonormalization method," *Opt. Lett.*, vol. 37, pp. 443–445, Feb 2012.

[66] A. Rabault, L. Le Magoarou, J. Sol, G. C. Alexandropoulos, N. Shlezinger, H. V. Poor, and P. del Hougne, "On the Tacit Linearity Assumption in Common Cascaded Models of RIS-Parametrized Wireless Channels," *arXiv preprint arXiv:2302.04993*, 2023.

[67] G. Zhou, C. Pan, H. Ren, P. Popovski, and A. L. Swindlehurst, "Channel Estimation for RIS-Aided Multiuser Millimeter-Wave Systems," *IEEE Transactions on Signal Processing*, vol. 70, pp. 1478–1492, 2022.

[68] C. Pan, G. Zhou, K. Zhi, S. Hong, T. Wu, Y. Pan, H. Ren, M. D. Renzo, A. Lee Swindlehurst, R. Zhang, and A. Y. Zhang, "An Overview of Signal Processing Techniques for RIS/IRS-Aided Wireless Systems," *IEEE Journal of Selected Topics in Signal Processing*, vol. 16, no. 5, pp. 883–917, 2022.

[69] P. Zhang, J. Zhang, H. Xiao, X. Zhang, D. W. K. Ng, and B. Ai, "Joint Distributed Precoding and Beamforming for RIS-Aided Cell-Free Massive MIMO Systems," *IEEE Transactions on Vehicular Technology*, pp. 1–6, 2023.

[70] D. Wang, J. Wang, X. You, Y. Wang, M. Chen, and X. Hou, "Spectral Efficiency of Distributed MIMO Systems," *IEEE Journal on Selected Areas in Communications*, vol. 31, no. 10, pp. 2112–2127, 2013.

[71] H. U. Rehman, F. Bellili, A. Mezghani, and E. Hossain, "Modulating Intelligent Surfaces for Multiuser MIMO Systems: Beamforming and Modulation Design," *IEEE Transactions on Communications*, vol. 70, no. 5, pp. 3234–3249, 2022.

[72] X. Mu, Y. Liu, L. Guo, J. Lin, and R. Schober, "Simultaneously Transmitting and Reflecting (STAR) RIS Aided Wireless Communications," *IEEE Transactions on Wireless Communications*, vol. 21, no. 5, pp. 3083–3098, 2022.

[73] Q. Zhang, Y. Zhao, H. Li, S. Hou, and Z. Song, "Joint Optimization of STAR-RIS Assisted UAV Communication Systems," *IEEE Wireless Communications Letters*, vol. 11, no. 11, pp. 2390–2394, 2022.

[74] J. Zuo, Y. Liu, Z. Ding, L. Song, and H. V. Poor, "Joint Design for Simultaneously Transmitting and Reflecting (STAR) RIS Assisted NOMA Systems," *IEEE Transactions on Wireless Communications*, vol. 22, no. 1, pp. 611–626, 2023.

[75] M. Cui, Z. Wu, Y. Lu, X. Wei, and L. Dai, "Near-Field MIMO Communications for 6G: Fundamentals, Challenges, Potentials, and Future Directions," *IEEE Communications Magazine*, vol. 61, no. 1, pp. 40–46, 2023.

[76] J. Liang and D. Liu, "Passive Localization of Mixed Near-Field and Far-Field Sources Using Two-stage MUSIC Algorithm," *IEEE Transactions on Signal Processing*, vol. 58, no. 1, pp. 108–120, 2010.

[77] C. Zou, K. An, Z. Lin, Y. He, X. Zhong, G. Zheng, and N. Al-Dhahir, "Multi-Layer RIS-Assisted Anti-Jamming Communications: A Hierarchical Game Learning Approach," *IEEE Communications Letters*, vol. 27, no. 11, pp. 2998–3002, 2023.

[78] Y. Sun, K. An, Y. Zhu, G. Zheng, K.-K. Wong, S. Chatzinotas, D. W. K. Ng, and D. Guan, "Energy-Efficient Hybrid Beamforming for Multilayer RIS-Assisted Secure Integrated Terrestrial-Aerial Networks," *IEEE Transactions on Communications*, vol. 70, no. 6, pp. 4189–4210, 2022.