D2.3
Final Results in Directional Links Analysis, and Algorithm Designs

Editor:	Joonas Kokkoniemi, UOULU
Deliverable nature:	Document, report (R)
Dissemination level:	Public (PU)
Date: planned | actual	31.10.2021	15.11.2021
Version | No. of pages	1.0	100
Keywords:	Beamforming algorithms, directional links, performance analysis, RIS channel estimation, RIS communications, RIS control algorithms.

Abstract

This deliverable brings the final results on directional links' analysis and algorithm design. The results, models, and analysis herein append the results given in the previous deliverable (D2.1) of Task 2.1 of the ARIADNE project. This deliverable reports the models and analysis for practical hardware imperfections, reconfigurable intelligent surface (RIS) gain properties, RIS beamforming algorithms, and performance analysis of RIS-aided links in this deliverable.

Disclaimer

Impressum

Full project title: Artificial Intelligence Aided D-band Network for 5G Long Term Evolution

Short project title: ARIADNE

Number and title of work-package: WP2, Novel Information and Communication Theory Framework Beyond Shannon

Number and title of task: T2.1, Directional D-Band Link Analysis and Design

Document title: D2.3, Final Results in Directional Links Analysis, and Algorithm Designs

Editor: Joonas Kokkoniemi, UOULU

Work-package leader: CNRS

Copyright notice

List of authors

Company	Author	Contribution
CNRS	Marco Di Renzo, Xuewen Qian, Fadil H. Danufane	Contributions to Chapters 3 and 4.
EURESCOM	Halid Hrasnica	Final review of the document
ICOM	Georgia Ntouni	Contributions to Chapter 1. Review of the document.
UOULU	Joonas Kokkoniemi	Editor of the document. Contributions to Chapter 2.
UPM	José Manuel Riera	Review of the document.
UPRC	Sotiris Droulias, Angeliki Alexiou	Contributions to Chapter 5. Final review of the document (Angeliki Alexiou).

Deliverable D2.3:
Final Results in Directional Links Analysis, and Algorithm Design

Executive Summary

This deliverable goes through the final results on directional links' analysis and algorithm design. The initial results were given in Deliverable 2.1. Those focused on channel estimation, beamforming, and performance analysis of directional links, positioning and user tracking, hardware (HW) imperfections, and some machine learning aspects of directional communications. This deliverable adds to those with practical Tx/Rx impairments, reconfigurable intelligent surface (RIS) gain analysis and beamforming algorithm designs, and performance assessments of RIS-aided links.

To be specific, Chapter 1 goes through various HW imperfections, their modelling, and discusses how those affect the signal chain. Chapter 2 analyses the achievable RIS gain as a function of the center frequency, size of the RIS, and distance. A comparison to line-of-sight link shows that the performance of the RIS is highly dependent on the above listed parameters, and the gain of the RIS in some situations can be very low. On the other hand, the RISs can also have very high gain if the distance to and out of the RIS is small and the RIS is large. Chapter 3 considers an RIS-enhanced multiple-input single-output (MISO) wireless communication system and proposes a reflection pattern modulation (RPM) scheme for the dual-use of passive beamforming and information transfer of the RIS. It is shown that the proposed RIS-RPM scheme was able to improve the achievable rate performance despite the loss in received signal power as compared to the conventional RIS-assisted system. Chapter 4 proposes a RIS quadrature reflection modulation (RIS-QRM) scheme for simultaneous passive beamforming and information transfer of RIS. The RIS-QRM is shown to be able to improve the error performance of the additional bits delivered by the RIS without deteriorating that of the bits carried on the constellation symbol, as compared to ON/OFF-based reflection modulation schemes. Finally, in Chapter 5, optimal RIS placement is studied on the basis of the available access point (AP) gain, the available positions for the AP and RIS placement, and the minimum desired power levels at the user equipment (UE). It is demonstrated that the optimal RIS placement relies on the AP antenna gain tunability. The results illustrate that increasing the AP antenna gain does not guarantee an increase in received power and the results provide guidelines for the optimal selection of the AP antenna gain and the RIS placement.

The concepts and results presented herein are very important for the other ARIADNE work packages. For instance, the directional links analysis with and without RISs feed directly into the system level performance evaluations in T1.3 where final ARIADNE system level evaluations are presented in the end of October 2022. The same applies for practical hardware impairments that are important to take into account in the system level designs and evaluations. The beamforming algorithms presented in D2.1 and in this deliverable are very important in order to understand how to maximize the received power in high frequency communication schemes that always require directional links to obtain enough antenna gain. This is especially the case with RISs where the channel can be controlled to some degree. On the other hand, the RISs make the full Tx-Rx link estimation and optimization more complex. In this regard, machine learning algorithms and artificial intelligent are seen as very promising tools for link and partial system optimization. Therefore, the results herein feed into WP4 of ARIADNE focusing on machine learning algorithms. All in all, this deliverable and D2.1 provide a lot of link level information that can be further utilized in the various WPs of the ARIADNE project.

Introduction

The future communication system seek the extreme data rates from the millimeter wave (mmWave, 30–300 GHz) and the THz band (0.3–10 THz). These very extremely high frequencies offer very wide frequency bands, making ot possible to achieve very large data rates even with relatively modest modulation methods. The specific focus of the ARIADNE project is the D band (110–170 GHz). This band is located in the middle of the mmWave band and it offers very respectable continuous bandwidths up to 12.5 GHz [1]. As the D band is close to existing specified fifth generation (5G) maximum frequencies, it is on the radar for the beyond 5G (B5G) and sixth generation (6G) communication systems.

When we move into higher frequencies, directional antennas become necessity due to need for high antenna/beamforming gain in all situations. Directional communications cause some notable problems, such as problems with channel acquisition, channel estimation, blockage of signals, and beam tracking. The reconfigurable intelligent surfaces (RISs) are seen as a very potential technologies to help with some of these issues by giving base station a partial control over the propagation channel. However, there are a lot of research problems with directional communications with RISs. Partially those are similar with the issues described above for generic directional communications, but with additional degree of difficulty coming from the additional link and the control of RIS(s) in base station (BS)-RIS-user equipment (UE) compound link.

ARIADNE Task 2.1 has been focusing on single directional link analysis with emphasis on the RIS communications. Deliverable 2.1 [2] gave the initial results on directional communications with topics covering hardware and other imperfections, machine learning aspects of the beamforming, and RIS related channel estimation, beam tracking, and performance evaluations. This deliverable appends on those results by topics listed below. Putting together D2.1 and D2.3, we have a very comprehensive look into the directional link level communications in the mmWave and THz bands.

The topics covered in this deliverable are as follows.

- Chapter 1 provides an overview of the practical hardware imperfections. There are various mechanism how hardware distorts the signal in various components of the radio frequency (RF) chain. Those include, e.g., non-linearity of the components and phase noise, to mention couple. These imperfections have impact on the communication link performance and should be taken into account in the system and link modelling.

- Chapter 2 looks into the gain characteristics of RISs as a function of the distance, frequency, and the size of the RIS. The apparent size of the RISs decrease as a function of the distance. The size of the RIS in general case decreases as a function of frequency. These have a great impact on the achievable gain of the RIS that can very from very small to very large depending on the link distances, frequency, and the physical size of the RIS.

- Chapters 3 and 4 look into beamforming design and performance analysis of RISs. These chapters propose RIS reflection pattern modulation (RIS-RPM) and RIS quadrature reflection modulation (RIS-QRM) schemes, respectively. The proposed RIS-RPM scheme is show to be able to improve the achievable rate performance despite the loss in received signal power as compared to the conventional

RIS-assisted system. The RIS-QRM is shown to improve the error performance without deteriorating that of the bits carried on the constellation symbol.

- Chapter 5 presents the performance analysis of RIS-aided links. One focal point therein is to study the optimal placement of the RISs by taking into account the access point (AP) gain, the available space for the AP and RIS locations, and the minimum desired power levels at the UE. It is shown that the optimal RIS placement relies on the AP antenna gain tunability.

Chapter 1

Practical Tx/Rx Impairments

This chapter provides an overview of the sources and effects of hardware impairments of ultra high carrier frequency wireless communication transceivers rendering their performance costly and hindering their exploitation as commercial products. The more the transmission bandwidths and carrier frequencies increase, the more the effect of these impairments worsens. In the following, we refer to the most critical components and the impairments they induce to the system. This is a brief overview and the interested reader is encouraged to check in detail various textbooks dealing with this topic, as for instance [3] and [4].

AD/DA converters Communication systems targeting high data rates need to utilize high bandwidth which along with the converters resolution, critically affect the system feasibility and power consumption. For high bandwidth systems, Digital-to-Analog converters (DACs) are somehow easier to be implemented, while Analog-to-Digital Converters (ADCs) act more as a bottleneck due to the limited resolution of high-speed ADCs. Some of the typical impairments introduced in ADCs are the quantization noise, aperture jitter, non-linearities, and saturation. DACs can also introduce strong harmonic components in the transmitted signal due to baseband non-linearities.

Mixers and local Oscillators In both the transmitter (Tx) and receiver (Rx) mixers, the effect of in-phase and quadrature (I/Q) imbalance is important and can be observed as images in the spectrum. If the I and Q channels of a mixer are not orthogonal, i.e., they do not have a 90° phase difference, and do not introduce the same gain value, then the output signal has I/Q imbalance. The local oscillators (LOs) that are used in order to generate the carrier frequency, suffer from various impairments, but among them the most critical is the phase noise. Meanwhile, the generated carrier frequency at the Tx and the Rx LOs is not identical, as LOs are not ideal. This frequency difference is referred to as carrier frequency offset and should be compensated. Finally, DC offset due to leakage is also a common impairment induced by these components, as LO's signal leaking from the LO towards the mixer's RF or IF input is self-mixed and results in adding a DC component to the input signal.

Power amplifier Power amplifiers (PAs) are utilized to compensate for the loss of energy, i.e., the attenuation of a signal, which occurs when a signal is transmitted through a communication medium. Apart from their high cost, PAs are among the most power

consuming elements in the transmitter chain and the most critical for the linearity of the transmission. The intrinsic non-linear behaviour of PAs generates intermodulation distortion (IMD) products, which produce out-of-band distortion. This distortion combined with the use of modulated signals of multiple levels, results in spectral regrowth, i.e., spectrum extension on adjacent channels. Besides, the non-linear PA behaviour also produces in-band distortion, which affects the transmitted signal constellation quality, while PAs might also introduce a DC component to the output signal. In general, using extremely large bandwidth values tends to increase the non-linearities induced by the PA.

Efficient power amplifiers need to operate close to saturation, i.e., have a low back-off. Thus, for some signal input values, they exhibit an undesired non-linear behavior, generating a distorted spectrum several times wider than the allocated channel. This is critical for commercial products and their certification as the transmitted power might not be compliant with the mandatory spectrum mask. Finally, the PA power affects the system coverage and might impose severe limitations, especially due to the fact that maximum power scales down with carrier frequency.

1.1   Non-linearities

Non-linearities induced by the PA are evident when it is operated in its non-linear region, i.e., close to its maximum output power where significant compression of the output signal can be observed. Non-linear distortions degrade the modulation accuracy and produce unwanted out-of-band emissions, due to spectral regrowth, causing interference to adjacent channels.

The simplest way to model RF non-linearities is by using a Taylor series with a maximum degree l_{max} [3]. Assuming x and y stand for the input and output signals, respectively, we have

$$y = \sum_{l=1,\, l\, odd}^{l_{max}} k_l x^l, \textit{ where } \quad k_l \in \mathcal{R}. \tag{1.1}$$

Apart for this simplified model, more accurate ones exist and can be utilized. The distortions of a real PA are generally modelled using amplitude-to-amplitude (AM/AM) and amplitude-to-phase (AM/PM) curves, which are extracted via measurements [5] and can be represented by mathematical functions. If we denote the complex input signal as

$$x(t) = a(t) \exp[j\phi(t)], \tag{1.2}$$

the output of the PA can be expressed as

$$y(t) = A(a(t)) \exp\left(j[\phi(t) + P(a(t))]\right), \tag{1.3}$$

where $A(a(t))$ is the AM/AM function of the PA output amplitude as a function of the PA input signal amplitude $a(t)$. Moreover, $P(a(t))$ is the AM/PM function describing the PA output phase as a function of the PA input signal amplitude. There exist various ways to extract the two functions above, namely $A(a(t))$ and $P(a(t))$, such as creating a lookup table through measurements and using linear extrapolation when required, as well as utilizing more complex models depending on the available measurements [5] such as the cubic polynomial, the Saleh model [6], the Rapp and modified Rapp models [7]

and others [8]. The AM/AM causes amplitude distortion, for instance the corner points of the constellation might move toward the origin due to PA gain compression, whereas AM/PM introduces phase shift, i.e., constellation rotation.

The non-linearities induced by the PA can be classified as memoryless or memory-based. For the former case, the distorted output of the PA does not depend on the past but only on the instantaneous input. For the latter case, the non-linearities model output is also affected by past values of the input. The PA memory changes the frequency response of the generated distortion, while it also increases the number of coefficients needed to model the induced non-linearity. The memory effects exhibited by PAs can be neglected for narrowband signals but are noticeable in the case of wide bandwidth input signals, such as those comprising multiple carrier signals.

1.2 Phase Noise

Generating a stabilized local oscillator (LO) signal with sufficient power to drive the transceiver's mixers is a challenge for high frequency communication systems. In an ideal LO, the power would be concentrated exactly at the carrier frequency, but in real world phase noise exists and results in spreading the power over frequencies around the carrier frequency. Phase noise induced by the transceivers' LOs is considered a critical impairment at high carrier frequencies. In most of the cases, frequency multipliers along with amplifiers are used in series in order to produce the high carrier frequency. This results in an undesirable increase of the phase noise floor in the pass-band. Typically, the phase noise level increases by 6 dB for each doubling of the frequency. The influence of the phase noise on the system performance depends on the system bandwidth. At lower bandwidths, the correlated part of the phase noise has the most critical impact and should be compensated, while, at higher bandwidths, the phase noise floor has the dominant impact.

To this end, the phase noise induced by the LO is commonly described as a superposition of a correlated Wiener [4] and an uncorrelated Gaussian process. For high-rate communication systems, the phase noise simulated as a Wiener process, does not affect the system performance as much as the Gaussian, thus its impact can be neglected [9]. As a result, the noise floor of the LO contributes to the observed phase noise significantly. A criterion indicating whether it is appropriate to utilize solely the Gaussian process for the phase noise modelling is the fulfilment of the following inequality

$$N f_c^2 T^2 \leq \frac{\ln(2)}{2\pi^2}, \qquad (1.4)$$

where N stands for the number of symbols per frame, f_c for the carrier frequency and T for the symbol period. This holds true for wideband systems where the carrier frequency is small compared to the system bandwidth. Phase noise can be also modelled using its power spectral density (PSD). In the literature, there exist various mathematical functions for the PSD, for instance [8, 10, 11], that can be utilized along with experimental measurements in order to approach a real LO phase noise PSD.

It should be noted that phase noise is a time varying random component of the phase that causes an arching of the signal constellation points, while a constant phase offset might be also present at the distorted signal. In general, the phase noise is time varying, while the phase noise processes of distinct receivers are uncorrelated.

1.3   I/Q imbalance

I/Q imbalance corresponds to the amplitude and phase mismatch between the in-phase and quadrature components of the signal at a transceiver. Its most significant part is caused by an non-ideal mixer, while any amplitude mismatch at the DAC's components will also appear as I/Q imbalance. Furthermore, there might be differences in conversion losses between the output ports of the I and Q channel mixers, while filters and ADCs in the I and Q paths might not be perfectly matched.

In most cases, the dominant I/Q phase imbalance is inserted when the phase between the I and Q channels of the quadrature modulator (up-converter) at the Tx and/or the quadrature demodulator (down-converter) at the Rx, is not equal to 90°, while the I/Q amplitude imbalance occurs when there is a gain mismatch between the I and Q channels mentioned above, respectively. I/Q phase imbalance causes a nominally square constellation to become rhombic, while I/Q amplitude imbalance turns a square constellation into a rectangular one.

The simplest way to model I/Q imbalance when a mixer does not respect the amplitude balance and the orthogonality between the I and Q channels is presented below [12]. Assuming that y is the complex signal impacted by I/Q imbalance and x is the complex input signal, we have

$$y = \mathfrak{Re}\{x\}\, \alpha_r \, \exp\left(-j\,0.5\,\Delta\phi\right) + j\, \mathfrak{Im}\{x\}\, \alpha_i \exp\left(+j\,0.5\,\Delta\phi\right), \qquad (1.5)$$

where $\Delta\phi$ is the phase imbalance in rad and A stands for the amplitude imbalance in dB. Thus, α_r and α_i can be expressed as

$$\alpha_r + j\,\alpha_i = 10^{A/10} + j\,10^{-A/10} \qquad (1.6)$$

I/Q imbalances are modelled as frequency flat in most of the studies. However, advanced models taking into account frequency selective imbalances and more complex effects can be found in the literature [4, 13, 14].

1.4   Carrier phase and frequency synchronization errors

Frequency offsets in wireless systems are generated by both carrier frequency offsets (CFOs), introduced by mismatch between the local oscillators of the Tx and Rx, and Doppler shifts, caused by movements in the wireless channel. The CFO corresponds to the frequency difference between the transmitted signal carrier and the carrier of the receiver due to the non-ideality of the LOs. A constant phase offset might be also present along with the phase distortion caused by phase noise. The phase offset can be assumed to be constant and uniformly distributed in $[0, 2\pi]$, while in most cases, it is related to the real hardware and the received signal path lengths.

Denoting f_c as the desired carrier frequency, the actual frequency can be defined as $f = f_c + \Delta f$, where Δf represents the random CFO which is uniformly distributed across $[-\zeta_{max}, +\zeta_{max}]$. The bound ζ_{max} is the LO precision tolerance which is expressed in parts per million (ppm) of the carrier frequency. For instance, a carrier frequency equal to 140 GHz and an LO specification of 100 ppm correspond to an LO of poor

quality, where the maximum CFO is 14 MHz. When the combined frequency offset among the Tx LO and the Rx LO needs to be considered, we denote $\Delta f_{tot} = f_{Rx} - f_{Tx}$, where the f_{Tx} and f_{Rx} correspond to the actual frequencies of the LOs at the Tx and the Rx, respectively, as defined above [4].

1.5   DC offset

The contribution of the RF downconversion stage to the DC offset is due to LO signal leaking and self-mixing as well as to transistor mismatches in the signal path. Imperfections of the ADC's circuitry contribute also to the DC offset. The DC offset is often modelled as a Gaussian random variable with zero mean and a given standard deviation, although in some hardware realizations, it is practically almost constant. While the DC offset is independent of the input signal's frequency, it critically depends on the level of the input signal.

1.6   Unified imperfections model

In their theoretical studies, researchers sometimes need to utilize a model that accounts for all the hardware imperfections of a communication system. Such a model has been proposed by [4] and utilized in various ARIADNE's studies, for instance [15]. In such cases, the received signal can be expressed as

$$y = A\left(x + \zeta_{Tx}\right) + \zeta_{Rx} + n, \tag{1.7}$$

where x and n stand for the transmitted signal and the additive white Gaussian noise (AWGN), respectively, and A represents the channel, whose model depends on the considered scenario. The transmitter's and receiver's hardware imperfections are represented by ζ_{Tx} and ζ_{Rx}, respectively, which are modelled as independent random variables that have zero mean complex Gaussian distributions with variances

$$\sigma_{Tx}^2 = \kappa_{Tx}^2 P_{Tx} \ \text{and} \ \sigma_{Rx}^2 = \kappa_{Rx}^2 A^2 P_{Tx}, \tag{1.8}$$

respectively. Besides, κ_{Tx} and κ_{Rx} are the error vector magnitudes of the transmitter and the receiver, while P_{Tx} stands for the transmitted power.

1.7   Concluding remarks

It should be emphasized that theoretical and simulation analysis of assumed hardware imperfections is critically affected by the uncertainty and inaccuracy of the utilized models. These render the suppression of the impairments' effects difficult, while their identification through measurements as well as experimental verification are essential for the implementation of a real-world system.

Chapter 2

Reflective Phased Array Type RIS gain analysis in mmWave Networks

In this chapter, we analyze the gain properties of mmWave channels when utilizing ideal RISs and compare those to the LoS case. We start with a general channel model for a RIS and derive the gain compared to LoS path in order to analyse the impact of carrier frequency on the achievable RIS gain. As the antenna structures may be very large measured by the number of elements while still having moderate physical size at high frequencies, the near field distance is also analyzed with respect to size and configuration of the array. We will show that the frequency and the number of antenna elements will have a significant impact on the gain. Furthermore, depending on the application and environment, the shape of the RIS may also be one of the design criteria if the near field of antenna need to be avoided. In short, the gain of the RIS with fixed number of antenna elements will increase at lower frequencies due to larger antenna array size, but the near field of the antenna array will increase linearly with the wavelength at the same time.

2.1   RIS and Channel Models

There are many ways to implement RISs. Two of the most common ones considered in the literature are metasurfaces and reflective phased arrays. Metasurface reflectors are structures formed of small meta-atoms that tend to be mutually coupled and the reconfigurability of the metasurface is obtained, for instance, by geometric design and tunable materials [16]. There are practical test set-ups for metasurfaces and also various models to characterize their performance. However, the models are often rather complicated and not easy for wireless network planning yet.

A simpler solution from analysis and also from the deployment points of view are the reflective phased arrays. They are essentially reflective antenna arrays without the baseband or radio frequency (RF) processing chains. These structures are often considered to be fully passive, disregarding some semi-passive structures where a subset of the RIS elements are able to receive the signals. Hence, those can help with the channel estimation [17]. A hybrid relay-reflecting RIS (HR-RIS) has been recently proposed in [18], but herein we focus on the passive reflecting structures. The antenna array type approach assumes individual phase control at each antenna element. The beam steering is then similar to that of the regular antenna array.
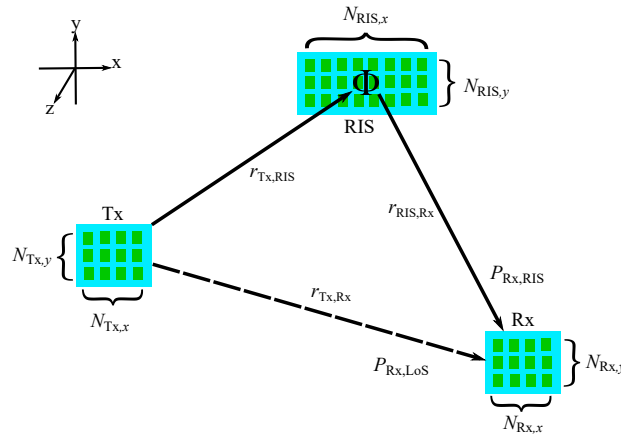
Figure 2.1: The system model used herein.

There are numerous papers on channel modelling for RIS assisted links [16, 19–28]. The core link model in all of them is a compound channel model formed of Tx-RIS and RIS-Rx links. In the most of the above channel modeling themed works, the links are assumed to be LoS. The main differences come from the handling of the beamforming. The antenna gains are important in the RIS modelling as it is possible with high gain antennas to be in such a close range to the RIS that the Tx antenna pattern is only partially illuminating the RIS [29]. This leads to decreased RIS response, but on the other hand, the same occurs in the more common case where RIS captures only a small part of the incoming radiation. This latter point is analyzed herein when we show that the RIS gain decreases with distance.

The channel model herein follows the work presented in [28] and it is commonly used in the mmWave band. However, we assume optimized beamformers such that the channel gain is given in terms of the antenna gains only. This model is detailed in the next section with the main focus on the channel gain by RIS and the near field radii of different RIS configurations.

2.2 Theoretical RIS Response

2.2.1 System Model Used in Analysis

The system model used to analyse the RIS gain is given in Fig. 2.1. It comprises a RIS, a transmitter (Tx), and a receiver (Rx). All of those utilize phased arrays with $N_{\text{RIS},y} \times N_{\text{RIS},x}$ antenna elements at RIS, $N_{\text{Tx},y} \times N_{\text{Tx},x}$ antenna elements at Tx, and likewise $N_{\text{Rx},y} \times N_{\text{Rx},x}$ antenna elements at Rx. The distances between the network elements have been given as follows: $r_{\text{Tx},\text{Rx}}$ is the LoS distance from the Tx to the Rx, $r_{\text{Tx},\text{RIS}}$ is the distance from the Tx to the RIS, and $r_{\text{RIS},\text{Rx}}$ is the distance from the RIS to the Rx. The RIS phase Φ is not modelled in this work as we use stochastic gains based on the far field propagation models as detailed below.

2.2.2 Path Loss Models

We use the standard LoS channel model in all links. That is given in the mmWave bands by the Friis' transmission equation with molecular absorption loss [30]. Thereby the receive power with the LoS channel is

$$P_{\text{Rx,LoS}}(r, f) = P_{\text{Tx}}G_{\text{Tx}}G_{\text{Rx}}\frac{c^2 e^{-\kappa_{\text{a}}(f)r_{\text{Tx,Rx}}}}{(4\pi r_{\text{Tx,Rx}}f)^2}, \tag{2.1}$$

where P_{Tx} is the transmit power, $\kappa_{\text{a}}(f)$ is the absorption coefficient, f is the frequency, c is the speed of light, and G_{Tx} and G_{Rx} are the Tx and Rx antenna gains, respectively.

We utilize a RIS model given in [28]. The model therein gives the receive power as

$$P_{\text{Rx,RIS}} = P_{\text{Tx}}G_{\text{Tx}}G_{\text{Rx}}G_{\text{RIS}}\Gamma_{\text{RIS}}\frac{e^{-\kappa_a(r_{\text{Tx,RIS}}+r_{\text{RIS,Rx}})}c^2 A_{\text{RIS}}}{(4\pi)^3 f^2 r_{\text{Tx,RIS}}^2 r_{\text{RIS,Rx}}^2}, \tag{2.2}$$

where $r_{\text{Tx,RIS}}$ and $r_{\text{RIS,Rx}}$ are the Tx–RIS and RIS–Rx distances (see also Fig. 2.1), A_{RIS} is the RIS area (or capture area for the incoming energy) and G_{RIS} is the total gain of the RIS. We added in a RIS reflection coefficient, or reflection efficiency Γ_{RIS} that accounts for the non-ideal reflection power, although, it is assumed to be one in the sequel. It should be noted that this is the maximum possible RIS channel gain. That is, the beamformers are fully optimized in the far field of the array. As a consequence, we obtain the full channel gain. In reality, the beamformers are estimated from the channel response and gain will be decreased due to any possible estimation errors at the Tx, Rx, or RIS, as well as due to possible hardware imperfections.

The RIS gain on the receive powers compared to LoS is given by division of (2.2) and (2.1) by setting the total distance in the both models to equal $r_{\text{Tx,RIS}} + r_{\text{RIS,Rx}}$. The resultant gain is given by

$$G_{\text{RIS,LoS}} = G_{\text{RIS}}A_{\text{RIS}}\Gamma_{\text{RIS}}\frac{(r_{\text{Tx,RIS}} + r_{\text{RIS,Rx}})^2}{4\pi r_{\text{Tx,RIS}}^2 r_{\text{RIS,Rx}}^2}. \tag{2.3}$$

Even with ideally reflecting RIS, we can see that the gain of the RIS compared to LoS path has certain bounds. Those include the gain and surface area of the RIS and the distance from the RIS. With a fully covered RIS surface and separation distance of d between the individual elements, the gain and the size of the RIS are interconnected. Herein we assume $d = \lambda/2$ separation between the antenna elements. By assuming fully covered RIS, the total area of the RIS becomes

$$A_{\text{RIS}} = [(N_x - 1)d_x + \epsilon_x] \times [(N_y - 1)d_y + \epsilon_y], \tag{2.4}$$

where $(N_x - 1)d_x + \epsilon_x$ and $(N_y - 1)d_y + \epsilon_y$ are the horizontal and vertical lengths of the antenna array, N_x and N_y are the number of antenna elements in the horizontal and vertical directions, d_x and d_y are the corresponding antenna element separations, and ϵ_x and ϵ_y give an additional room for realistic antenna element sizes. That is, the added length in the horizontal and vertical directions due to physical size of the antenna elements. Our assumptions on fully covered RIS surface hints that $\epsilon_x = \epsilon_y = d_x = d_y = d$. With ideal antennas without any specific antenna size, $\epsilon_x = \epsilon_y = 0$. In any case, the total effective capture area for energy at RIS has an impact on the total energy for the reflection towards the Rx. Due to limited size, the distance from the Tx or Rx to the RIS

has impact on the energy captured by this area. Therefore, a finite size RIS tends to lose in gain to LoS path, but this depends on the size of RIS and distance to it as it will be shown in the numerical results. The above also gives away that the gain of the RIS is dependent on the size of the elements via their area. This in an ideal setting leads to the diminishing RIS size as a function of the wavelength, which leads to a smaller gain with a fixed number of antenna elements when moving to higher frequencies.

From the above, we can show the gain of an ideal square RIS compared to LoS gain is

$$G_{\text{RIS,LoS}} = N_{\text{RIS}}^2 d^2 \frac{(r_{\text{Tx,RIS}} + r_{\text{RIS,Rx}})^2}{4\pi r_{\text{Tx,RIS}}^2 r_{\text{RIS,Rx}}^2}, \tag{2.5}$$

where $N_{R}IS$ is the total number of antenna elements in the RIS, and hence, this is the maximum gain of the RIS G_{RIS}. The square RIS also has the smallest near field among the rectangular RIS structures. This can be a very appealing structure as the near field of the RIS can be very large with large numbers of antenna elements as discussed in the next section.

2.2.3 Near Field of the RIS

It should be evident that the above models for RIS propagation are only valid in the far field of the RIS (or any other antenna in the system). However, as the future antenna systems and especially RISs can be very large to obtain the required gain levels, the near field of an antenna can be very large. Therefore, we analyze the near field size of an array of various phased antenna array configurations. We only focus on rectangular arrays here with different aspect ratios. Linear arrays are omitted as those are unlikely in arrays with large numbers of antenna elements.

The near field radius of an antenna array in general case is given by

$$r_{\text{NF}} = \frac{2D^2}{\lambda}, \tag{2.6}$$

where r_{NF} is the near field radius, D is the maximum dimension of the antenna, and λ is the wavelength. The maximum dimension of the antenna array is assumed to be formed by the maximum dimension of the area confined by the antenna elements as

$$D = \sqrt{[(N_x - 1)d_x + \epsilon_x]^2 + [(N_y - 1)d_y + \epsilon_y]^2}, \tag{2.7}$$

where the maximum distance across the antenna D herein with rectangular arrays is from a corner to corner. However, in reality, the actual dimensions of the antennas depend on the type of the antenna. Here we use easily tractable rectangular arrays and the above holds for those.

The RIS near field radius is given by

$$r_{\text{NF}} = \frac{2}{\lambda} \left([(N_x - 1)d_x + \epsilon_x]^2 + [(N_y - 1)d_y + \epsilon_y]^2 \right). \tag{2.8}$$

Furthermore, for ideal square RIS without the ϵ, the near field becomes

$$r_{\text{NF}} = \frac{4}{\lambda}(\sqrt{N} - 1)^2 d^2, \tag{2.9}$$

or $r_{\text{NF}} = 4Nd^2/\lambda$ if we assume ϵ to be equal to the antenna separation d. As mentioned above, and using the above equations, we show that the antenna configuration will have a large impact on the near field size close to large antenna arrays.
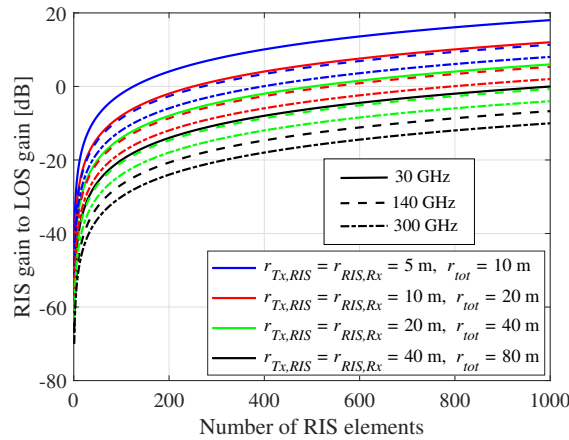
Figure 2.2: RIS gain as a function RIS antenna elements for various distances and three center frequencies given by solid lines for 30 GHz, dashed lines for 140 GHz, and dash-dotted lines for 300 GHz.

2.3   Numerical Results

In this section, we give the numerical results on impact of the RIS size at different frequencies on the achievable channel gain and the size of the near field. The frequencies herein are 30, 140, and 300 GHz, i.e., the far ends of the mmWave band, and the middle point of the D band ($110 - 170$ GHz).

The achievable RIS gains with optimized beamformers given by (2.3) are shown in Fig. 2.2. The gains are given for equal distances from Tx-RIS and RIS-Rx distances that are varied from 5 to 40 meters. This translates into total distances varying from 10 to 80 meters. The three different frequencies are given by solid lines for 30 GHz, dashed lines for 140 GHz, and dash-dotted lines for 300 GHz. We can see that the RIS gain obviously increases as a function of the RIS antenna elements. This directly follows from increased surface area for capturing the energy and more antenna element to bring more gain. We further see that for fixed number of antenna elements, the gain decreases when frequency increases. Although the gain remains constant across the frequencies with a fixed number of RIS antenna elements, the surface area of the RIS decreases with increasing frequency. This leads to the observed loss in the gain. Lastly, as we can see in (2.3), increasing distance decreases the RIS gain. This is due to decreased energy density at the surface of the RIS when distance is increased.

These results are inline with intuitive behavior of the RISs. The surface area of the RIS determines how much energy can be redirected at most, the number of antenna elements determine the maximum gain, and the distance determines the energy distribution across the surface. Therefore, while very potential technologies, the RISs have fundamental limits how much gain they can give at maximum. This gain is therefore an important system parameter when determining the cell sizes in RIS assisted networks.

Figures 2.3–2.5 give the near field radii of the RIS (or any array) as a function of the antenna elements. The results are given for two options, with the ϵ parameter equal to zero or the antenna element separation distance d, i.e., when the antenna element size is equivalent to the antenna element separation distance d. Although, this latter case only has minor impact on the near field radii due to d being rather small. Furthermore, we give the results for various array aspect ratios as this will have great impact on the
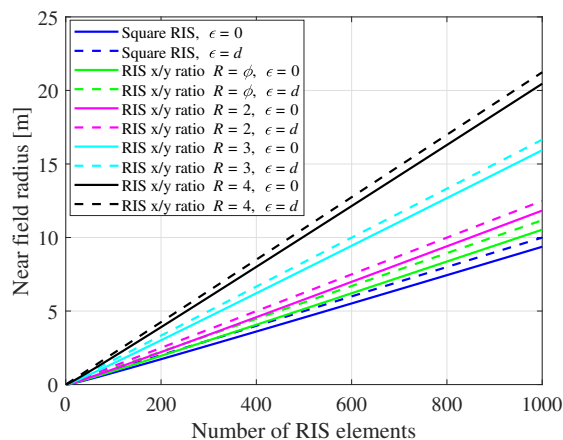
Figure 2.3: Near field radius of RIS for various RIS aspect ratios as a function of the number of RIS elements at 30 GHz center frequency.

near field of the array. That is, the ratio R equals the difference in the horizontal length of the array against the vertical length. As an example, $R = 2$ means that there are twice the number of element in horizontal direction than in the vertical direction. For these results, we used pure mathematical ratios and the number of elements in x- and y-direction is obtained from the total number of antenna elements N with ratio R as $N_x = \sqrt{RN}$ and $N_y = \sqrt{N/R}$. To have a step between square RIS and one with aspect ratio 2/1, we used golden ratio ($\phi = (1 + \sqrt{5})/2 \approx 1.618$) as one of the ratios.

We can see in Figs. 2.3–2.5 that as expected, the near field radius increases as a function of the number of the antenna elements. It should be noted that the above propagation equations are derived by using far field assumptions. Those overestimate the received power in the near field due to spherical wave propagation in this region. We can see that the aspect ratio of the array has a major impact on the near field distance. This follows from increasing maximum dimension as the antenna gets smaller toward some other dimension. The maximum near field radius would therefore be given by a linear array. However, the linear arrays are very unlikely in RIS type applications due to large size and no elevation control. The main observation here is related to the gain versus the frequency. When looking at Fig. 2.2, we can see that a RIS can theoretically exceed the LoS gain in certain setups. This is most likely with large numbers of RIS antenna elements at relatively low transmit frequency. These high gain setups, on the other hand lead to large near field close to antenna array. In certain situation, such as in indoors, the near field size has to be small enough to allow high and predictable gain. Especially, in mobile applications where near field introduces spatial fluctuation on the signal levels. As such, the near field radius may impose some application and environment specific limits on the RIS sizes even if the physical sizes of high frequency RISs are most likely quite modest.

Finally, Figs. 2.6 and 2.7 give the spatial distribution of the LoS power and RIS power, respectively, with 32-by-32 RIS, 16-by-16 Tx, and 4-by-4 Rx arrays at 140 GHz. We can see that the LoS link gives roughly at least 20 dB higher Rx power compared to RIS link when assuming that the LoS link is obstructed in RIS communications. If we look at Fig. 2.2, we see that the distances in order of about 20 meters or less with 1024 RIS antenna elements gives in fact more gain than the LoS link. However, the increased distance via the reflected path from Tx to Rx favors the direct LoS link. This
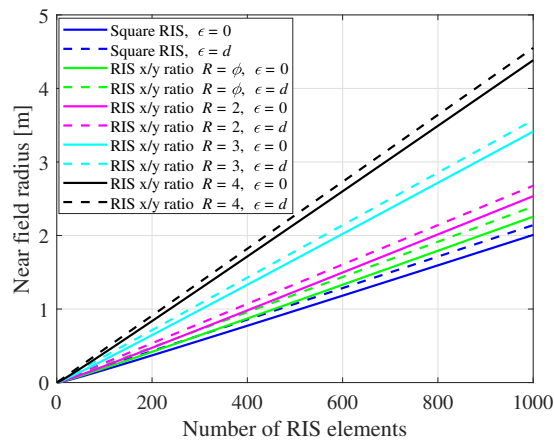
Figure 2.4: Near field radius of RIS for various RIS aspect ratios as a function of the number of RIS elements at 140 GHz center frequency.
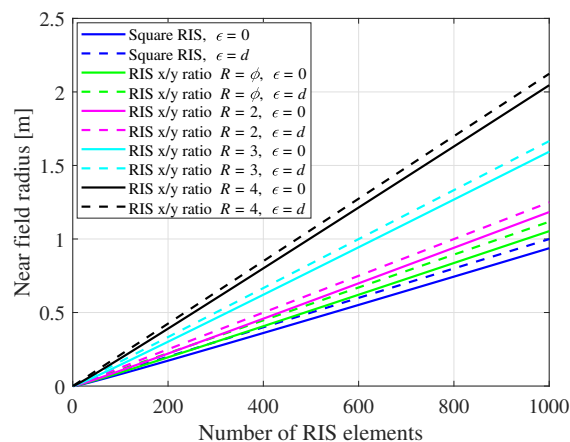
Figure 2.5: Near field radius of RIS for various RIS aspect ratios as a function of the number of RIS elements at 300 GHz center frequency.

is very much expected as it is well known from literature than the RISs should be mostly beneficial in situations where the LoS link is not available.

As a final summary, the RIS gains depend on many factors that do not always favor their gain. Especially in realistic deployment scenarios where such massive surfaces may be difficult to realize. However, the ability to steer the beams behind the corner make them very potential technologies for increasing cell sizes, decreasing the impact of blockage, serving cell edge users, creating virtual multiple-input multiple-output (MIMO) channels, and in many more applications. Thus, it can be expected that the RISs will be an important part of the high frequency small cell 5G and beyond networks.

2.4 Conclusion

We presented and went over the channel modelling for RIS assisted systems and analyzed the size of the near-field in realistic mmWave system assumptions in this chapter. The achievable gain with respect to the LoS gain was analyzed against frequency and
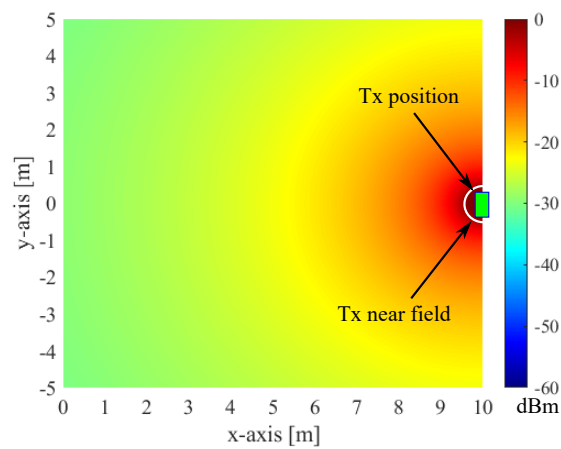
Figure 2.6: Received power via LoS link with 16-by-16 Tx and 4-by-4 Rx.
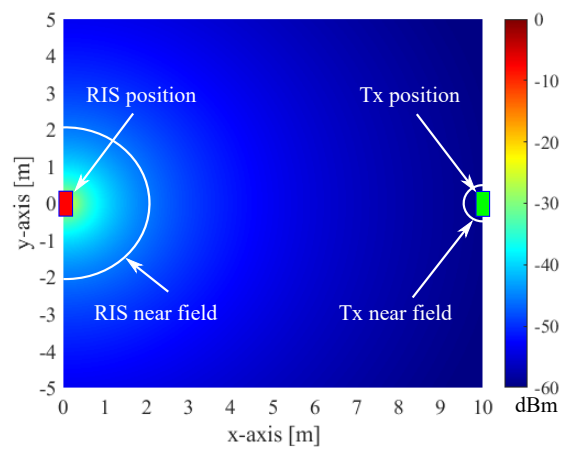
Figure 2.7: Received power at 140 GHz via 32-by-32 RIS with 16-by-16 Tx and 4-by-4 Rx.

size of the RIS. It was shown that the RIS gain decreases with frequency due to decreasing capture area for the reflected energy. However, at the same time, the near field radius of the RIS decreases with frequency due to the same reason. This means that, depending on the application, there may be limitations on the size of the RIS (and therefore gain) when far field communications need to be guaranteed. Whereas the RIS gain can exceed the LoS gain with large enough RIS size, the LoS link tends to give a better overall gain due to shorter propagation distance. The ability to control the reflection direction at RIS makes them very potential in high blockage probability scenarios, such as indoor and dense city scenarios. Thus, the RISs will be essential part of the future high frequency communication systems where good channels are very important.

Chapter 3

Reconfigurable Intelligent Surfaces with Reflection Pattern Modulation: Beamforming Design and Performance Analysis

Recent considerations for RISs assume that RISs can convey information by reflection without the need of transmit radio frequency chains, which, however is a challenging task. In this chapter, we propose an RIS-enhanced multiple-input single-output system with reflection pattern modulation, where the RIS can configure its reflection state for boosting the received signal power via passive beamforming and simultaneously conveying its own information via reflection. We formulate an optimization problem to maximize the *average* received signal power by jointly optimizing the active beamforming at the access point (AP) and passive beamforming at the RIS for the case where the RIS's state information is statistically known by the AP, and propose a high-quality suboptimal solution based on the alternating optimization technique. We analyze the asymptotic outage probability of the proposed scheme under Rayleigh fading channels, for which a closed-form expression is derived. The achievable rate of the proposed scheme is also investigated for the case where the transmitted symbol is drawn from a finite constellation. Simulation results validate the effectiveness of the proposed scheme and reveal the effect of various system parameters on the achievable rate performance. It is shown that the proposed scheme outperforms the conventional RIS-assisted system without information transfer in terms of achievable rate performance.

3.1 Introduction

To meet the demanding requirements for 5G wireless communication in, e.g., enhanced data rate, massive connectivity, low latency, ultra reliability, etc., multiple key technologies including mmWave communications, massive multiple-input multiple-output (MIMO) systems, and ultra-dense networks (UDNs) have been extensively investigated in the last decade [31]. However, most of those technologies generally require increased implementation complexity and result in considerably increased energy consumption [32, 33]. Leveraging the recent advances in reconfigurable metasurfaces [34, 35], RISs (a.k.a. intelligent reflecting surfaces) have emerged as a revolu-

tionary technology for improving the coverage and energy/spectrum efficiency of future wireless communications [36–42]. Specifically, RISs are planar metasurfaces consisting of a large number of low-cost unit cell elements, each of which reflects the incident signals according to its reflection state. By configuring the reflection amplitude and/or phase shift at each unit cell element according to the dynamic wireless channels, the signals reflected by the RIS can constructively or destructively combine with the signals from other paths at the receiver for signal power enhancement or co-channel interference suppression [43]. Compared to traditional relays, RISs are envisioned to work in a full-duplex mode without incurring self-interference and thermal noise, and yet possess substantially reduced hardware cost and energy consumption owing to their nearly passive components [44–47].

To achieve the theoretical passive beamforming gain offered by RISs in RIS-assisted systems, the global channel state information (CSI) was assumed to be available at the RIS side [48], which is practically difficult to implement. The acquisition/estimation of channels involving RIS is in practice a challenging task due to the absence of signal processing capability of passive RISs and their large number of unit cell elements, which has spurred rapidly growing interests [49–56]. Existing RIS channel acquisition methods can be classified into two categories. The first category [49, 50] includes the approaches considering that RISs are embedded with some receive radio frequency (RF) chains, which enable them to explicitly estimate the channels from the network's transmitters/receivers to RIS. In the second category, instead of explicitly estimating the individual channels from the network's transmitters/receivers to RIS, the cascaded transmitter-RIS-receiver channels are estimated at the receiver by adopting an element-by-element ON/OFF-based reflection pattern at RIS without the need of costly receive RF chains [52]. However, this approach incurs prohibitive channel estimation overhead due to the large number of RIS elements. To reduce the channel estimation overhead, a grouping strategy for the RIS elements was proposed in [54], where adjacent RIS elements are grouped together to share a common reflection coefficient. Based on this grouping strategy, a channel estimation method using the discrete Fourier transform (DFT)-based reflection pattern was first presented in [55] for frequency-selective fading channels, which significantly improves the channel estimation accuracy by leveraging the large aperture gain of RIS. In [57], a DFT-based reflection pattern was designed for channel estimation in RIS-assisted narrowband systems. More recently, some preliminary works (e.g., [58, 59]) have appeared for channel estimation in RIS-assisted multi-user systems.

Prior works on RISs mainly focus on their utilization for maximizing the received signal strength of links between the access point (AP) and users, e.g., [60]. In order to improve the spectral efficiency of RIS-assisted systems, the authors of [61] introduced the RIS-space shift keying and RIS-spatial modulation schemes to implement the index modulation at the receiver side. In addition, the authors of [62] proposed to jointly encode the information in the transmitted signal and RIS state for improving spectral efficiency. On the other hand, another type of the RIS-assisted system, in which the RIS needs to deliver its own information to the receiver, was considered in [63]. In this case, the RIS is equipped with sensors, such as temperature and/or humidity sensors, and has a requirement on conveying its locally acquired information to an intended receiver. Besides the information from sensors, there are other potential sources of RIS information, including the RIS control signaling, RIS maintenance data, estimated CSI at the RIS, etc., [50, 64]. One option for the information transfer of the RIS is to
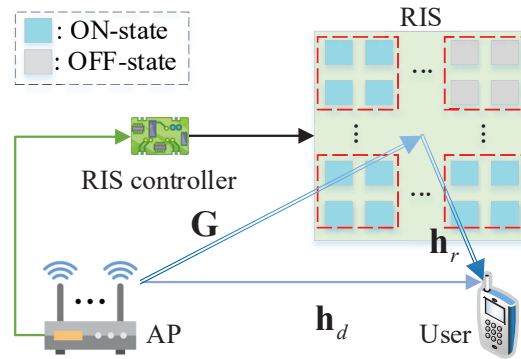
Figure 3.1: The considered RIS-enhanced downlink MISO wireless communication system. The RIS adopts reflection pattern modulation via the combination of its elements' ON/OFF states to simultaneously enhance the communication between the AP and receiving user, and convey its own information data.

install multiple transmit RF chains, which, however, incurs much higher hardware cost and energy consumption as compared to the fully-passive RIS case [40]. Alternatively, when no transmit RF chain is installed at the RIS, its own information should be implicitly delivered via appropriately designed reflection state, which is a more cost-effective, yet challenging, solution. To this end, a passive beamforming and information transfer (PBIT) scheme was proposed in [65] and [66] to convey the RIS's information besides implementing fully-passive beamforming. However, in this scheme, the number of activated (ON-state) elements[1] varies over time, which can cause significant fluctuation in the reflected signal power, resulting in a relatively high outage probability.

In this chapter, we consider an RIS-enhanced multiple-input single-output (MISO) wireless communication system as shown in Fig. 3.1, where a multi-antenna AP communicates with a single-antenna user with the aid of an RIS. The RIS is explored as an efficient and inexpensive dual-use technology of passive beamforming and information transfer by proposing the concept of reflection pattern modulation (RPM). The core RPM idea is to activate a subset of RIS unit cell elements in order to reflect a sharp beam towards the intended destination, while exploiting the indices of the activated (ON-state) elements to implicitly convey the RIS's information. The main contributions of this chapter are summarized as follows:

- We present an RIS-based RPM (RIS-RPM) scheme for PBIT, in which the RIS plays the following two roles: 1) it boosts the received signal power at the intended receiver; 2) it transfers via reflection its locally acquired information data. We consider the practical scenario where the ON/OFF state information of the RIS is statistically known by the AP, for which an optimization problem is formulated to maximize the *average* received signal power by jointly optimizing the active beamforming at the AP and passive beamforming at the RIS. The formulated problem, however, is non-convex and thus difficult to solve optimally. As such, we propose an efficient algorithm based on the alternating optimization technique to find a high-quality suboptimal solution. Moreover, we formulate and solve an optimization problem to maximize the *instantaneous* received signal power by designing the active and passive beamforming based on the RIS's instantaneous

---

[1]The ON/OFF switching of the RIS elements can be implemented by the switchers.

ON/OFF state information, which serves as a performance upper bound for the practical beamforming design based on the RIS's statistical ON/OFF state information.

- The asymptotic outage probability of the proposed RIS-RPM scheme over Rayleigh fading channels is derived in closed-form. It is shown that the RIS is able to drastically increase the diversity gain by properly designing the phase shifts of its elements. Moreover, we analyze the achievable rate of the RIS-RPM scheme for the practical case where the transmitted symbol is drawn from a finite constellation input. Simulation results validate the effectiveness of the proposed RIS-RPM scheme as well as the proposed optimization algorithm, and unveil the effect of various system parameters on the achievable rate performance. It is shown that the RIS-RPM scheme outperforms the conventional RIS-assisted system with full-ON RIS reflection [48] in terms of achievable rate performance, despite the loss in the received signal strength. Moreover, the RIS-RPM scheme is shown to have the potential to strike an attractive trade-off between the received signal power and achievable rate performance by varying the number of activated (ON-state) RIS elements.

The rest of this chapter is organized as follows. In Section 3.2, we introduce the system model associated with channel acquisition. In Sections 3.3 and 3.4, we formulate two optimization problems for designing active and passive beamforming with efficient solutions. Performance analysis in terms of outage probability and achievable rate is presented in Section 3.5. Simulation results are presented in Section 3.6 and conclusions are drawn in Section 3.7.

Notation: Upper and lower case boldface letters denote matrices and column vectors, respectively. $(\cdot)^{\dagger}$, $(\cdot)^{T}$, $(\cdot)^{H}$, and $(\cdot)^{-1}$ represent conjugation, transpose, Hermitian transpose, and inversion operations, respectively. $[\mathbf{X}]_{\iota,\jmath}$ denotes the $(\iota,\jmath)$-th entry of matrix $\mathbf{X}$, and $[\mathbf{x}]_{\iota}$ denotes the ι-th entry of vector $\mathbf{x}$. $\mathbf{I}_N$ denotes an $N \times N$ identity matrix, and $\mathbf{1}_N$ and $\mathbf{0}_N$ denote N-dimensional all-one and all-zero column vectors, respectively. $\|\mathbf{x}\|_0$ and $\|\mathbf{x}\|$ denote the zero norm and Euclidean norm of vector $\mathbf{x}$, respectively. $\mathbb{C}^{n \times m}$ denotes the set of $n \times m$ complex-valued matrices. $\mathrm{diag}(\mathbf{x})$ denotes a diagonal matrix with each diagonal entry being the corresponding entry in vector $\mathbf{x}$. $\mathrm{tr}(\mathbf{X})$ and $\mathrm{rank}(\mathbf{X})$ denote the trace and rank of matrix $\mathbf{X}$, respectively. $|\cdot|$ and $\angle(\cdot)$ denote the modulus and phase of a complex number, respectively. $p(\cdot)$ denotes the probability density function (PDF). $\lceil \cdot \rceil$ represents the ceiling function that returns the least integer greater than or equal to the argument. $\binom{n}{k}$ denotes the binomial coefficient. $\mathbb{A} \setminus \mathbb{B}$ is the complement set of $\mathbb{B}$ with respect to $\mathbb{A}$. $\mathcal{CN}(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ denotes the distribution of a circularly symmetric complex Gaussian random vector with covariance matrix $\boldsymbol{\Sigma}$ and mean $\boldsymbol{\mu}$. $\mathbb{E}_X\{\cdot\}$ denotes the expectation over random variable X. Let $\mathbf{X} \succeq 0$ denote that a Hermitian matrix $\mathbf{X}$ is positive semi-definite.

3.2 System Description and Channel Estimation

As illustrated in Fig. 3.1, we consider a MISO wireless communication system, where an RIS composed of L unit cell elements is deployed to enhance the communication link from an AP equipped with N transmit antennas to a single-antenna user. An RIS

Table 3.1: An Example of the Proposed RPM Scheme with an RIS of $L = 4$ Unit Cell Elements ($K = 3$ ON-state Elements). The Index of the OFF-state Element is Used to Implicitly Convey the RIS's Information Bits.

Information bits of RIS	00	01	10	11
Index of OFF-state element	{4}	{3}	{2}	{1}

controller is attached to the RIS, which is responsible for reconfiguring the phase shifts and/or reflection amplitudes of the unit cell elements [67] and exchanging information with the AP for the realization of the simultaneous passive beamforming and reflection modulation. We assume that the RIS controller is equipped with a number of sensors in order to monitor/collect environmental data, which is typically low-rate bursty data for the user. In this chapter, we consider quasi-static block fading channels for the AP-user, AP-RIS, and RIS-user links, where all the channels remain unchanged over the coherence block equal to the duration of T_c symbol sampling periods, and are independent and identically distributed (i.i.d.) across different coherence blocks. This assumption is valid since RISs are practically used to support low-mobility users in their vicinity.

3.2.1 System Model

For each symbol duration, only K out of L $(K \leq L)$ unit cell elements are turned ON to reflect incident signals on purpose so that the indices of those ON-state elements can implicitly convey the RIS information. With the proposed RPM scheme, there are in total $\binom{L}{K}$ combinations for determining the indices of K ON-state elements at the RIS, such that $\log_2 \binom{L}{K}$ information bits can be conveyed. An example of the proposed RPM scheme with an RIS of $L = 4$ unit cell elements ($K = 3$ ON-state elements) is illustrated in Table 3.1. However, the number of unit cell elements at the RIS can be practically large, leading to an enormous calculation and storage overhead in the combination mapping as the number of K-combinations increases exponentially with L. To address this issue, we adopt the RIS-elements grouping method, where a total number of L unit cell elements are divided into G groups, each of which consists of $\bar{L} = L/G$ adjacent elements sharing a common reflection coefficient. For notational convenience, we assume that $\bar{L}$ is an integer that we refer to as the grouping size. As such, the information of the RIS is conveyed through the indices of the ON-state groups. Specifically, for each symbol duration, $\bar{K}$ out of G $(\bar{K} \leq G)$ groups are randomly turned ON for reflecting the incident signals, and the remaining $(G-\bar{K})$ groups are deliberately turned OFF for realizing the proposed RPM scheme. Therefore, a total number of $\log_2 \binom{G}{\bar{K}}$ information bits can be conveyed. Let $\mathbb{I} \triangleq \{i_1, i_2, \ldots, i_{\bar{K}}\}$ denote the indices of the $\bar{K}$ ON-state groups, where $i_k \in \mathbb{G} \triangleq \{1, 2, \ldots, G\}$ for $k = 1, 2, \ldots, \bar{K}$ with $i_1 < i_2 < \ldots < i_{\bar{K}}$. Given the indices of the ON-state groups, the common reflection coefficient

for the g-th group can be characterized by

$$\theta_g = \begin{cases} \beta_g e^{-j\phi_g}, & g \in \mathbb{I}, \\ 0, & g \in \mathbb{G} \setminus \mathbb{I}, \end{cases} \quad g = 1, 2, \ldots, G \tag{3.1}$$

where $\beta_g \in [0,1]$ and $\phi_g \in (0, 2\pi]$ represent the common reflection amplitude and phase shift for the g-th group, respectively. To enhance the reflected signal power and ease the hardware design, the reflection amplitudes of the ON-state groups are set to be the maximum value, i.e., $\beta_g = 1$, $\forall g \in \mathbb{I}$. Let $\boldsymbol{\theta} \triangleq [\theta_1, \theta_2, \ldots, \theta_G]^T$ denote the RIS reflection vector after grouping, which characterizes the equivalent interaction of the RIS with the incident signals. According to (4.3), we have $\|\boldsymbol{\theta}\|_0 = \bar{K}$.

Let $\mathbf{G} \triangleq [\boldsymbol{\Delta}_1, \boldsymbol{\Delta}_2, \ldots, \boldsymbol{\Delta}_G]^H \in \mathbb{C}^{L \times N}$, $\mathbf{h}_r^H \triangleq [\mathbf{r}_1^H, \mathbf{r}_2^H, \ldots, \mathbf{r}_G^H] \in \mathbb{C}^{1 \times L}$, and $\mathbf{h}_d^H \in \mathbb{C}^{1 \times N}$ denote the baseband channels from the AP to RIS, from the RIS to user, and from the AP to user, respectively, where $\boldsymbol{\Delta}_g^H \in \mathbb{C}^{\bar{L} \times N}$ and $\mathbf{r}_g^H \in \mathbb{C}^{1 \times \bar{L}}$ denote the corresponding channels associated with the RIS elements belonging to the g-th group, respectively. Moreover, due to the severe path loss and high attenuation, the signals reflected by the RIS more than once have negligible power and hence can be ignored. Accordingly, the received signal at the user is given by

$$y = \sqrt{P_t} \left(\sum_{g=1}^G \mathbf{r}_g^H \theta_g \boldsymbol{\Delta}_g^H + \mathbf{h}_d^H \right) \mathbf{w} x + n \tag{3.2}$$

where P_t is the maximum transmit power of the AP, $\mathbf{w} \in \mathbb{C}^{N \times 1}$ stands for the active beamforming at the AP, x is the transmitted symbol, which is drawn from an M-ary constellation $\mathbb{A}$ with normalized power, θ_g is the common reflection coefficient for the g-th group, and $n \sim \mathcal{CN}(0, \sigma^2)$ is the additive white Gaussian noise (AWGN) with σ^2 being the noise power. Since RIS elements in the same group share a common reflection coefficient, by denoting $\bar{\mathbf{h}}_g^H = \mathbf{r}_g^H \boldsymbol{\Delta}_g^H$ as the equivalent cascaded channel of the AP-RIS-user link associated with the g-th group without the effect of RIS reflection, (4.7) can be rewritten as

$$y = \sqrt{P_t} \left(\boldsymbol{\theta}^T \mathbf{H} + \mathbf{h}_d^H \right) \mathbf{w} x + n \tag{3.3}$$

where $\mathbf{H} = \left[\bar{\mathbf{h}}_1, \bar{\mathbf{h}}_2, \ldots, \bar{\mathbf{h}}_G \right]^H \in \mathbb{C}^{G \times N}$ denotes the cascaded AP-RIS-user channel matrix without the effect of RIS reflection. From (4.49), it can be observed that the information delivered to the user consists of two parts: the first part is from the AP *explicitly* expressed as x and the second part is from the RIS *implicitly* embedded in $\boldsymbol{\theta}$.

3.2.2 Channel Estimation

To boost the received signal power by jointly optimizing the active beamforming at the AP and passive beamforming at the RIS, the knowledge of $\mathbf{H}$ and $\mathbf{h}_d$ is required. By assuming channel reciprocity and using time-division duplex (TDD) protocol, the CSI of $\mathbf{H}$ and $\mathbf{h}_d$ can be obtained at the AP. Specifically, during the channel training, $(G + 1)$ consecutive pilot symbols are sent by the user. For the duration of the i-th $(i = 0, 1, \ldots, G)$ pilot symbol, the received pilot signal vector at the AP can be expressed

as

$$\mathbf{y}_p^{(i)} = \sqrt{P_p} \left(\sum_{g=1}^{G} \mathbf{\Delta}_g \psi_g^{(i)} \mathbf{r}_g + \mathbf{h}_d \right) x_p^{(i)} + \mathbf{n}^{(i)} \tag{3.4}$$

where P_p is the transmit power of the user, $x_p^{(i)}$ is the i-th pilot symbol with normalized power, $\psi_g^{(i)}$ is the common phase shift for the g-th group during the transmission of the i-th pilot symbol with $|\psi_g^{(i)}| = 1$, and $\mathbf{n}^{(i)} \sim \mathcal{CN}(\mathbf{0}_N, \sigma^2 \mathbf{I}_N)$ is the AWGN vector. By letting $\boldsymbol{\psi}^{(i)} \triangleq \left[\psi_1^{(i)}, \psi_2^{(i)}, \ldots, \psi_G^{(i)} \right]^T$ denote the RIS phase-shift state during the transmission of the i-th pilot symbol, (3.4) can be rewritten as

$$\mathbf{y}_p^{(i)} = \sqrt{P_p} \left(\mathbf{H}^H \boldsymbol{\psi}^{(i)} + \mathbf{h}_d \right) x_p^{(i)} + \mathbf{n}^{(i)} = \sqrt{P_p} \tilde{\mathbf{H}} \tilde{\boldsymbol{\psi}}^{(i)} x_p^{(i)} + \mathbf{n}^{(i)} \tag{3.5}$$

where $\tilde{\mathbf{H}} = \begin{bmatrix} \mathbf{h}_d & \mathbf{H}^H \end{bmatrix} \in \mathbb{C}^{N \times (G+1)}$ and $\tilde{\boldsymbol{\psi}}^{(i)} = \begin{bmatrix} 1 \\ \boldsymbol{\psi}^{(i)} \end{bmatrix}$. By stacking $(G+1)$ consecutive received pilot signals, we have

$$\mathbf{Y}_p = \left[\mathbf{y}_p^{(0)}, \mathbf{y}_p^{(1)}, \ldots, \mathbf{y}_p^{(G)} \right] = \sqrt{P_p} \tilde{\mathbf{H}} \boldsymbol{\Psi} \text{diag}\left(\mathbf{x}_p \right) + \boldsymbol{\Omega} \tag{3.6}$$

where $\boldsymbol{\Psi} = \left[\tilde{\boldsymbol{\psi}}^{(0)}, \tilde{\boldsymbol{\psi}}^{(1)}, \ldots, \tilde{\boldsymbol{\psi}}^{(G)} \right]$ is the RIS phase-shift pattern, $\mathbf{x}_p = \left[x_p^{(0)}, x_p^{(1)}, \ldots, x_p^{(G)} \right]^T$ denotes the pilot sequence, and $\boldsymbol{\Omega} = \left[\mathbf{n}^{(0)}, \mathbf{n}^{(1)}, \ldots, \mathbf{n}^{(G)} \right]$ is the AWGN matrix. By using the DFT-based phase-shift pattern [57], the CSI of $\mathbf{H}$ and $\mathbf{h}_d$ can be estimated as

$$\hat{\tilde{\mathbf{H}}} = \begin{bmatrix} \hat{\mathbf{h}}_d & \hat{\mathbf{H}}^H \end{bmatrix} = \frac{1}{\sqrt{P_p}(G+1)} \mathbf{Y}_p \text{diag}\left(\mathbf{x}_p \right)^{-1} \mathbf{F}_{G+1}^H \tag{3.7}$$

where $[\mathbf{F}_{G+1}]_{\iota,\jmath} = e^{-j\frac{2\pi \iota \jmath}{G+1}}$, $\iota, \jmath = 0, 1, \ldots, G$.

After acquiring the CSI of $\mathbf{H}$ and $\mathbf{h}_d$, the AP jointly optimizes the active beamforming $\mathbf{w}$ and phase shifts of all RIS-elements groups to boost the received signal power, and then informs the RIS controller the optimized phase shifts to be implemented at the corresponding unit cell elements. In the following section, we formulate an optimization problem to maximize the *average* received signal power by jointly designing the active beamforming at the AP and passive beamforming at the RIS based on the statistical ON/OFF state information of the RIS.

3.3 Beamforming Design Based on Statistical ON/OFF State Information

3.3.1 Problem Formulation

We consider the practical assumption that the ON/OFF state information of the RIS is statistically known by the AP. Our objective is to minimize the outage probability of the combined AP-user channel by jointly optimizing the active beamforming at the AP

and passive beamforming at the RIS under the constraints of the maximum transmit power of the AP and unit-modulus reflection of the RIS. Specifically, given the indices of the ON-state groups $\mathbb{I}$, let $\mathbf{s} \triangleq [s_1, s_2, \ldots, s_G]^T$ denote the ON/OFF states of all RIS-elements groups in a vector form, with each entry given by

$$s_g = \begin{cases} 1, & g \in \mathbb{I}, \\ 0, & g \in \mathbb{G} \setminus \mathbb{I}, \end{cases} \tag{3.8}$$

where $s_g = 1$ and $s_g = 0$ represent the ON and OFF states, respectively. Based on (4.5), the RIS reflection vector $\boldsymbol{\theta}$ can be rewritten as $\boldsymbol{\theta} = \boldsymbol{\Phi}\mathbf{s}$, in which $\boldsymbol{\Phi} = \mathrm{diag}\,(\boldsymbol{\varphi})$ is the diagonal phase-shift matrix of the RIS with the g-th diagonal entry given by $\varphi_g = e^{-j\phi_g}$, $\phi_g \in (0, 2\pi]$, $g = \{1, 2, \ldots, G\}$. The combined channel from the AP to user is given as $\boldsymbol{\theta}^T \mathbf{H} + \mathbf{h}_d^H = \mathbf{s}^T \boldsymbol{\Phi}\mathbf{H} + \mathbf{h}_d^H$, and thus the achievable rate of the combined AP-user channel conditioned on $\mathbf{H}$ and $\mathbf{h}_d$ can be expressed as $\log_2 \left(1 + \frac{P_t}{\sigma^2} \left| \left(\mathbf{s}^T \boldsymbol{\Phi}\mathbf{H} + \mathbf{h}_d^H \right) \mathbf{w} \right|^2 \right)$ bits/second/Hertz. The corresponding outage probability for a fixed rate R is given by

$$\begin{aligned} p_{\mathsf{out}}(R) &= \mathbb{P} \left\{ \log_2 \left(1 + \frac{P_t}{\sigma^2} \left| \left(\mathbf{s}^T \boldsymbol{\Phi}\mathbf{H} + \mathbf{h}_d^H \right) \mathbf{w} \right|^2 \right) < R \right\} \\ &= \mathbb{P} \left\{ P_t \left| \left(\mathbf{s}^T \boldsymbol{\Phi}\mathbf{H} + \mathbf{h}_d^H \right) \mathbf{w} \right|^2 < (2^R - 1)\sigma^2 \right\}. \end{aligned} \tag{3.9}$$

The optimization problem for minimizing the outage probability in (3.9) under the constraints of the maximum transmit power of AP and unit-modulus reflection of RIS can be formulated as

$$\text{(P0):} \quad \min_{\mathbf{w}, \boldsymbol{\Phi}} \quad t \tag{3.10}$$

$$\text{s.t.} \quad \mathbb{P} \left\{ P_t \left| \left(\mathbf{s}^T \boldsymbol{\Phi}\mathbf{H} + \mathbf{h}_d^H \right) \mathbf{w} \right|^2 < (2^R - 1)\sigma^2 \right\} < t \tag{3.11}$$

$$\|\mathbf{w}\|^2 \leq 1 \tag{3.12}$$

$$|\varphi_g| = 1, \quad g = 1, 2, \ldots, G. \tag{3.13}$$

However, this problem is non-convex and difficult to solve due to the robust outage probability and unit-modulus constraint. On the other hand, it can be observed from (9) that given any ON/OFF-state vector $\mathbf{s}$, we should optimize the active beamforming at the AP and passive beamforming at the RIS to maximize the received signal power for minimizing the outage probability. However, since the instantaneous ON/OFF state information of the RIS is unavailable to the AP, we instead optimize the active beamforming and passive beamforming to maximize the *average* received signal power. Note that this is effective in improving the performance in terms of outage probability, as will be shown in our simulations. Let $\mathbf{A} = \mathbb{E}_{\mathbf{s}} \left\{ \mathbf{s}\mathbf{s}^T \right\}$ and $\mathbf{a} = \mathbb{E}_{\mathbf{s}} \left\{ \mathbf{s} \right\}$ denote the covariance matrix and mean vector of $\mathbf{s}$, respectively. Based on (4.5) and assuming that all possible index combinations are equiprobable, the elements of $\mathbf{A}$ and $\mathbf{a}$ can be derived as

follows.

$$[\mathbf{A}]_{i,j} = \begin{cases} \dfrac{\bar{K}}{G}, & i = j \\[3mm] \dfrac{\bar{K}(\bar{K}-1)}{G(G-1)}, & i \neq j \end{cases} \quad i,j = 1,2,\ldots,G \tag{3.14}$$

and

$$[\mathbf{a}]_i = \frac{\bar{K}}{G}. \tag{3.15}$$

Thus, the average received signal power normalized by the transmit power is given by

$$\mathbb{E}_{\mathbf{s}}\left\{\left|\left(\mathbf{s}^T\boldsymbol{\Phi}\mathbf{H} + \mathbf{h}_d^H\right)\mathbf{w}\right|^2\right\}$$

$$= \mathbf{w}^H\left(\mathbf{H}^H\boldsymbol{\Phi}^H\mathbb{E}_{\mathbf{s}}\{\mathbf{s}\mathbf{s}^T\}\boldsymbol{\Phi}\mathbf{H} + \mathbf{H}^H\boldsymbol{\Phi}^H\mathbb{E}_{\mathbf{s}}\{\mathbf{s}\}\mathbf{h}_d^H + \mathbf{h}_d\mathbb{E}_{\mathbf{s}}\{\mathbf{s}^T\}\boldsymbol{\Phi}\mathbf{H} + \mathbf{h}_d\mathbf{h}_d^H\right)\mathbf{w}$$

$$= \mathbf{w}^H\begin{bmatrix}\boldsymbol{\Phi}\mathbf{H} \\ \mathbf{h}_d^H\end{bmatrix}^H\underbrace{\begin{bmatrix}\mathbf{A} & \mathbf{a} \\ \mathbf{a}^T & 1\end{bmatrix}}_{\tilde{\mathbf{A}}}\begin{bmatrix}\boldsymbol{\Phi}\mathbf{H} \\ \mathbf{h}_d^H\end{bmatrix}\mathbf{w}. \tag{3.16}$$

Accordingly, the corresponding optimization problem based on the estimated CSI of $\hat{\mathbf{H}}$ and $\hat{\mathbf{h}}_d$ is formulated as:

$$\text{(P1):} \quad \max_{\mathbf{w},\boldsymbol{\Phi}} \quad \mathbf{w}^H\begin{bmatrix}\boldsymbol{\Phi}\hat{\mathbf{H}} \\ \hat{\mathbf{h}}_d^H\end{bmatrix}^H\tilde{\mathbf{A}}\begin{bmatrix}\boldsymbol{\Phi}\hat{\mathbf{H}} \\ \hat{\mathbf{h}}_d^H\end{bmatrix}\mathbf{w} \tag{3.17}$$

$$\text{s.t.} \quad \|\mathbf{w}\|^2 \leq 1 \tag{3.18}$$

$$|\varphi_g| = 1, \quad g = 1,2,\ldots,G. \tag{3.19}$$

It can be readily verified that problem (P1) is a non-convex problem as well, since the objective function of (3.17) is non-concave with respect to both $\mathbf{w}$ and $\boldsymbol{\Phi}$, and the constraint in (3.19) is not convex. Moreover, due to the mutual coupling between $\mathbf{w}$ and $\boldsymbol{\Phi}$ in the objective function of (3.17), problem (P1) becomes even more difficult to solve. To circumvent the above difficulties, we develop an alternating optimization algorithm to find a high-quality suboptimal solution to problem (P1) in the following, which iteratively optimizes one of $\mathbf{w}$ and $\boldsymbol{\Phi}$ with the other being fixed at each time for decoupling the original problem.

3.3.2 Joint Beamforming Design

For any given RIS phase-shift matrix $\boldsymbol{\Phi}$, problem (P1) can be rewritten as

$$\text{(P1.1):} \quad \max_{\mathbf{w}} \quad \mathbf{w}^H\tilde{\mathbf{R}}\mathbf{w} \tag{3.20}$$

$$\text{s.t.} \quad \|\mathbf{w}\|^2 \leq 1 \tag{3.21}$$

Algorithm 1 Alternating Optimization Algorithm for Solving Problem (P1)

Input: $\hat{\mathbf{H}}$, $\hat{\mathbf{h}}_d$, threshold ϵ, and the maximum iteration number I

1: Initialize the diagonal phase-shift matrix $\boldsymbol{\Phi}^{(1)} := \mathbf{I}_G$ and set the iteration number $n := 1$
2: **repeat**
3:    Substitute $\boldsymbol{\Phi}^{(n)}$ into (22) to get $\tilde{\mathbf{R}}^{(n)}$, then find the eigenvector corresponding to the maximum eigenvalue of $\tilde{\mathbf{R}}^{(n)}$ to obtain the active beamforming $\mathbf{w}^{(n)}$
4:    Given $\mathbf{w}^{(n)}$, solve problem (P1.4) via a convex optimization solver and the Gaussian randomization of (32) to obtain the RIS phase-shift matrix $\boldsymbol{\Phi}^{(n+1)}$
5:    Update $n := n + 1$
6: **until** The fractional increase of (17) is less than ϵ or $n > I$

Output: $\mathbf{w}^*$ and $\boldsymbol{\Phi}^*$

Figure 3.2: Alternating Optimization Algorithm for Solving Problem (P1)

where

$$\tilde{\mathbf{R}} = \begin{bmatrix} \boldsymbol{\Phi}\hat{\mathbf{H}} \\ \hat{\mathbf{h}}_d^H \end{bmatrix}^H \tilde{\mathbf{A}} \begin{bmatrix} \boldsymbol{\Phi}\hat{\mathbf{H}} \\ \hat{\mathbf{h}}_d^H \end{bmatrix}. \tag{3.22}$$

Since $\tilde{\mathbf{R}}$ is Hermitian, for any non-zero $\mathbf{w}$, we have the following inequality

$$\mathbf{w}^H \tilde{\mathbf{R}} \mathbf{w} \leq \lambda_{\max}(\tilde{\mathbf{R}}) \|\mathbf{w}\|^2 \tag{3.23}$$

where $\lambda_{\max}(\tilde{\mathbf{R}})$ denote the maximum eigenvalue of $\tilde{\mathbf{R}}$. Let $\mathbf{v}_{\max}$ denote the eigenvector corresponding to the maximum eigenvalue of $\tilde{\mathbf{R}}$. Then, it can be readily verified that the optimal solution to problem (P1.1) is given by $\mathbf{w}^* = \mathbf{v}_{\max}/\|\mathbf{v}_{\max}\|$.

Next, we optimize φ based on the given active beamforming $\mathbf{w}^*$. Specifically, for given $\mathbf{w}^*$, by letting $\boldsymbol{\Lambda} = \text{diag}\left(\hat{\mathbf{H}}\mathbf{w}^*\right)$ and $g_d = \hat{\mathbf{h}}_d^H \mathbf{w}^*$, problem (P1) can be rewritten as follows (omitted irrelevant terms for brevity).

$$(\text{P1.2}): \quad \max_{\varphi} \quad \varphi^H \boldsymbol{\Lambda}^H \mathbf{A} \boldsymbol{\Lambda} \varphi + g_d \varphi^H \boldsymbol{\Lambda}^H \mathbf{a} + g_d^H \mathbf{a}^T \boldsymbol{\Lambda} \varphi \tag{3.24}$$

$$\text{s.t.} \quad |\varphi_g| = 1, \quad g = 1, 2, \ldots, G. \tag{3.25}$$

From (3.24) and (3.25), we see that problem (P1.2) is a non-convex quadratically constrained quadratic program (QCQP), which can be reformulated as a homogeneous QCQP by introducing an auxiliary variable t [68], i.e.,

$$(\text{P1.3}): \quad \max_{\tilde{\varphi}} \quad \tilde{\varphi}^H \boldsymbol{\Xi} \tilde{\varphi} \tag{3.26}$$

$$\text{s.t.} \quad |\varphi_g| = 1, \quad g = 1, 2, \ldots, G \tag{3.27}$$

where

$$
\mathbf{\Xi} = \begin{bmatrix} \mathbf{\Lambda}^H \mathbf{A} \mathbf{\Lambda} & g_d \mathbf{\Lambda}^H \mathbf{a} \\ g_d^H \mathbf{a}^T \mathbf{\Lambda} & 0 \end{bmatrix}, \quad \tilde{\boldsymbol{\varphi}} = \begin{bmatrix} \boldsymbol{\varphi} \\ t \end{bmatrix}. \tag{3.28}
$$

The objective function of (3.26) can be rewritten as $\tilde{\boldsymbol{\varphi}}^H \mathbf{\Xi} \tilde{\boldsymbol{\varphi}} = \operatorname{tr}(\mathbf{\Xi} \mathbf{Q})$ with $\mathbf{Q} = \tilde{\boldsymbol{\varphi}} \tilde{\boldsymbol{\varphi}}^H$. Note that $\mathbf{Q}$ is a positive semidefinite matrix with $\operatorname{rank}(\mathbf{Q}) = 1$. However, as the rank-one constraint is non-convex, we apply the semidefinite relaxation (SDR) method to relax this constraint and reformulate problem (P1.3) as

$$
\text{(P1.4):} \quad \max_{\mathbf{Q}} \quad \operatorname{tr}(\mathbf{\Xi} \mathbf{Q}) \tag{3.29}
$$

$$
\text{s.t.} \quad [\mathbf{Q}]_{g,g} = 1, \ g = 1, 2, \ldots, G+1 \tag{3.30}
$$

$$
\mathbf{Q} \succeq 0 \tag{3.31}
$$

which is a standard convex semidefinite programming (SDP) problem and can be well solved via existing convex optimization solvers such as CVX [69]. It is worth pointing out that after the relaxation, the optimal solution $\mathbf{Q}^*$ to problem (P1.4) may not be a rank-one solution. Therefore, we retrieve $\tilde{\boldsymbol{\varphi}}^*$ from $\mathbf{Q}^*$ as follows.

$$
\tilde{\boldsymbol{\varphi}}^* = \begin{cases} \mathbf{U} \mathbf{D}^{1/2} \mathbf{1}_{G+1}, & \operatorname{rank}(\mathbf{Q}^*) = 1 \\ \mathbf{U} \mathbf{D}^{1/2} \boldsymbol{\varkappa}, & \operatorname{rank}(\mathbf{Q}^*) \neq 1 \end{cases} \tag{3.32}
$$

where $\mathbf{Q}^* = \mathbf{U} \mathbf{D} \mathbf{U}^H$ is the eigenvalue decomposition of $\mathbf{Q}^*$ and $\boldsymbol{\varkappa} \sim \mathcal{CN}(\mathbf{0}_{G+1}, \mathbf{I}_{G+1})$ is a random vector. Finally, the suboptimal solution $\boldsymbol{\varphi}^*$ to problem (P1.2) is given by

$$
\varphi_g^* = \frac{[\tilde{\boldsymbol{\varphi}}^*]_g \big/ [\tilde{\boldsymbol{\varphi}}^*]_{G+1}}{\left| [\tilde{\boldsymbol{\varphi}}^*]_g \big/ [\tilde{\boldsymbol{\varphi}}^*]_{G+1} \right|}, \quad g = 1, 2, \ldots, G. \tag{3.33}
$$

The algorithm proceeds by iteratively solving subproblems (P1.1) and (P1.4) in an alternating manner until the convergence criterion is met, i.e., the fractional increase of (3.17) is less than a small positive number ϵ, or the maximum number of iterations has been carried out in practice. Algorithm 3.2 summarizes the above procedures. The convergence of the proposed algorithm can be guaranteed by the fact that the objective value of problem (P1) is non-decreasing over iterations and upper-bounded by a finite value due to the limited transmit power. On the other hand, problem (P1.1) involving the eigenvalue decomposition of an $N \times N$ matrix can be solved with a complexity of $\mathcal{O}(N^3)$, and the SDP problem (P1.4) can be solved with a worst-case complexity of $\mathcal{O}((G+1)^{4.5})$ [70]. Given the number of iterations I, the total complexity for solving problem (P1) is thus given by $\mathcal{O}\left((N^3 + (G+1)^{4.5})I\right)$. After getting the optimized phase-shift vector $\boldsymbol{\varphi}^*$, the RIS performs passive beamforming with $\boldsymbol{\varphi}_{\mathbb{I}}^*$ only (while the RIS elements belonging to the remaining $(G - \bar{K})$ groups are set to be OFF, i.e., $\boldsymbol{\varphi}_{\mathbb{G}\backslash\mathbb{I}}^* = \mathbf{0}_{G-\bar{K}}$) according to the selection of $\mathbb{I}$ by RIS.

3.4 Beamforming Design Based on Instantaneous ON/OFF State Information

In this section, we characterize the upper bound on the received signal power, which serves to compare the above beamforming design based on the statistical ON/OFF state information of the RIS. We assume that the AP knows the ON/OFF state information of the RIS exactly in real time. It is worth mentioning that this assumption is similar to that in [62]. Under this assumption, we formulate an optimization problem to maximize the received signal power by jointly designing the active beamforming at the AP and passive beamforming at the RIS based on instantaneous ON/OFF state information of the RIS.

Given the indices of the ON-state groups $\mathbb{I}$, the corresponding optimization problem based on the estimated CSI can be formulated as follows (with P_t omitted for brevity).

$$\text{(P2):} \quad \max_{\mathbf{w}, \boldsymbol{\theta}} \quad \left| \left(\boldsymbol{\theta}^T \hat{\mathbf{H}} + \hat{\mathbf{h}}_d^H \right) \mathbf{w} \right|^2 \tag{3.34}$$

$$\text{s.t.} \quad \|\mathbf{w}\|^2 \leq 1, \tag{3.35}$$

$$\left| \theta_g \right| = 1, \quad \forall \, g \in \mathbb{I}, \tag{3.36}$$

$$\theta_g = 0, \quad \forall \, g \in \mathbb{G} \setminus \mathbb{I}. \tag{3.37}$$

By eliminating the constraint of (3.37), problem (P2) is equivalent to

$$\text{(P2.1):} \quad \max_{\mathbf{w}, \boldsymbol{\theta}_{\mathbb{I}}} \quad \left| \left(\boldsymbol{\theta}_{\mathbb{I}}^T \hat{\mathbf{H}}_{\mathbb{I}} + \hat{\mathbf{h}}_d^H \right) \mathbf{w} \right|^2 \tag{3.38}$$

$$\text{s.t.} \quad \|\mathbf{w}\|^2 \leq 1, \tag{3.39}$$

$$\left| \theta_{i_k} \right| = 1, \; i_k \in \mathbb{I}, \; k = 1, 2, \ldots, \bar{K} \tag{3.40}$$

where $\boldsymbol{\theta}_{\mathbb{I}}$ is the sub-vector consisting of the $\bar{K}$ entries of $\boldsymbol{\theta}$ indexed by $\mathbb{I}$, and $\hat{\mathbf{H}}_{\mathbb{I}}$ is the sub-matrix consisting of the $\bar{K}$ rows of $\hat{\mathbf{H}}$ indexed by $\mathbb{I}$. It can be readily verified that problem (P2.1) is non-convex as well, since the objective function of (3.38) is non-concave with respect to both $\mathbf{w}$ and $\boldsymbol{\theta}_{\mathbb{I}}$, as well as the constraint of (3.40) is not convex. Apparently, problem (P2.1) can be solved suboptimally by leveraging the alternating optimization technique similarly to Algorithm 3.2. Specifically, based on the alternating optimization technique, one of $\mathbf{w}$ and $\boldsymbol{\theta}_{\mathbb{I}}$ is optimized with the other being fixed in each iteration. For given active beamforming $\mathbf{w}$, problem (P2.1) can be reformulated as

$$\text{(P2.2):} \quad \max_{\boldsymbol{\theta}_{\mathbb{I}}} \quad \left| \sum_{k=1}^{\bar{K}} \theta_{i_k} \hat{\bar{\mathbf{h}}}_{i_k}^H \mathbf{w} + \hat{\mathbf{h}}_d^H \mathbf{w} \right|^2 \tag{3.41}$$

$$\text{s.t.} \quad \left| \theta_{i_k} \right| = 1, \; i_k \in \mathbb{I}, \; k = 1, 2, \ldots, \bar{K}. \tag{3.42}$$

By exploiting the triangle inequality, we have

$$\left| \sum_{k=1}^{\bar{K}} \theta_{i_k} \hat{\bar{\mathbf{h}}}_{i_k}^H \mathbf{w} + \hat{\mathbf{h}}_d^H \mathbf{w} \right| \leq \sum_{k=1}^{\bar{K}} \left| \theta_{i_k} \hat{\bar{\mathbf{h}}}_{i_k}^H \mathbf{w} \right| + \left| \hat{\mathbf{h}}_d^H \mathbf{w} \right| \tag{3.43}$$

with equality if and only if $\angle \left(\theta_{i_k} \hat{\mathbf{h}}_{i_k}^H \mathbf{w} \right) = \angle \left(\hat{\mathbf{h}}_d^H \mathbf{w} \right)$ for all $i_k \in \mathbb{I}$. Therefore, the optimal phase shift for the k-th ($k = 1, 2, \ldots, \bar{K}$) ON-state group is given by

$$\phi_{i_k}^* = \angle \left(\hat{\mathbf{h}}_{i_k}^H \mathbf{w} \right) - \angle \left(\hat{\mathbf{h}}_d^H \mathbf{w} \right), \quad i_k \in \mathbb{I}. \tag{3.44}$$

For given $\boldsymbol{\theta}_{\mathbb{I}}^*$ in (3.44), it can be readily obtained that the optimal active beamforming $\mathbf{w}^*$ is given by

$$\mathbf{w}^* = \mathbf{w}_{\mathsf{MRT}} \triangleq \frac{\left(\boldsymbol{\theta}_{\mathbb{I}}^T \hat{\mathbf{H}}_{\mathbb{I}} + \hat{\mathbf{h}}_d^H \right)^H}{\left\| \boldsymbol{\theta}_{\mathbb{I}}^T \hat{\mathbf{H}}_{\mathbb{I}} + \hat{\mathbf{h}}_d^H \right\|} \tag{3.45}$$

which is the well-known maximum-ratio transmission (MRT). $\boldsymbol{\theta}_{\mathbb{I}}$ and $\mathbf{w}$ are iteratively optimized according to (3.44) and (3.45) in an alternating manner until the convergence criterion is met. Note that the above solution is guaranteed to converge since the objective value of problem (P2) is non-decreasing over iterations and the optimal objective value of problem (P2) is finite.

3.5 Performance Analysis

In this section, we investigate the proposed RIS-RPM scheme in terms of the outage probability and achievable rate.

3.5.1 Outage Probability

For ease of exposition, we assume $N = 1$ with $\mathbf{G} \equiv \mathbf{g}$ and $\mathbf{h}_d^H \equiv h_d^\dagger$, such that the active beamforming vector $\mathbf{w}$ can be dropped. Moreover, we consider the Rician fading channel model for all the channels involved, where each channel coefficient equals the superposition of a determined line-of-sight (LoS) component and a non-LoS component (characterized by a complex Gaussian random variable). Let κ_{AR}, κ_{Ru}, and κ_{Au} denote the Rician factors of the AP-RIS, RIS-user, and AP-user links, respectively. In particular, the RIS is generally installed on the walls/ceilings to establish a LoS link with the AP to boost the signal strength in its vicinity, while the user is usually in a relatively rich scattering environment. Therefore, we assume $\kappa_{\mathsf{AR}} = \infty$, $\kappa_{\mathsf{Ru}} = 0$, and $\kappa_{\mathsf{Au}} = 0$, so that the AP-RIS channel has only a fixed LoS component while the AP-user and RIS-user channels can be well characterized by Rayleigh fading. Let σ_g^2 denote the power in the LoS component of the AP-RIS channel and assume $\mathbf{h}_r^H \sim \mathcal{CN}(\mathbf{0}_L, \sigma_h^2 \mathbf{I}_L)$ as well as $h_d^\dagger \sim \mathcal{CN}(0, \sigma_d^2)$. As such, we have $\mathbf{H} \equiv \mathbf{h} \sim \mathcal{CN}(\mathbf{0}_G, \sigma_r^2 \mathbf{I}_G)$ with $\sigma_r^2 = \bar{L} \sigma_h^2 \sigma_g^2$. For ease of notation, we assume $\sigma_d^2 = \sigma_r^2 = 1$. Moreover, we resort to a unified definition of signal-to-noise ratio (SNR) as $\gamma = P_t / \sigma^2$ to draw essential insights. Then the outage probability in (3.9) can be simplified as

$$p_{\mathsf{out}}(R) = \mathbb{P} \left\{ \left| \mathbf{s}^T \boldsymbol{\Phi} \mathbf{h} + h_d^\dagger \right|^2 < \frac{2^R - 1}{\gamma} \right\}. \tag{3.46}$$

Let $\phi_0 = \angle(h_d^\dagger)$ and $\boldsymbol{\chi} \triangleq [\chi_0, \chi_1, \ldots, \chi_G]^T = \begin{bmatrix} h_d^\dagger \\ \boldsymbol{\Phi}\mathbf{h} \end{bmatrix}$. First, we assume perfect CSI

available at the AP for setting the phase shifts of RIS elements. The optimal phase shifts for $N = 1$ are solutions that arrange the signals reflected by RIS elements to align in phase with the signal over the direct link at the user, regardless of the ON/OFF-state information of the RIS. Thus, we have $\boldsymbol{\chi} = e^{j\phi_0}\left[|\chi_0|, |\chi_1|, \ldots, |\chi_G|\right]^T$, where $\{|\chi_g|\}_{g=0}^G$

are i.i.d. Rayleigh random variables with parameter $\sqrt{1/2}$. Let $X = \left|\mathbf{s}^T\boldsymbol{\Phi}\mathbf{h} + h_d^\dagger\right|^2$,

which is the square of the sum of $(\bar{K}+1)$ independent Rayleigh random variables and has a Gamma distribution with parameters

$$k_x = \frac{\mathbb{E}\{X\}^2}{\mathbb{E}\{X^2\} - \mathbb{E}\{X\}^2}, \quad \theta_x = \frac{\mathbb{E}\{X^2\} - \mathbb{E}\{X\}^2}{\mathbb{E}\{X\}} \tag{3.47}$$

where

$$\mathbb{E}\{X\} = (\bar{K}+1)(1 + \frac{\pi}{4}\bar{K}) \tag{3.48}$$

$$\mathbb{E}\{X^2\} = 2(\bar{K}+1) + (\frac{3\pi}{2} + 3)(\bar{K}+1)\bar{K} + \frac{3\pi}{2}(\bar{K}+1)\bar{K}(\bar{K}-1)$$

$$+ \frac{\pi^2}{16}(\bar{K}+1)\bar{K}(\bar{K}-1)(\bar{K}-2). \tag{3.49}$$

Its probability density function is

$$p_X(x) = x^{k_x-1}\frac{e^{-x/\theta_x}}{(\theta_x)^{k_x}\Gamma(k_x)}, \quad x \geq 0. \tag{3.50}$$

Approximating e^{-x/θ_x} by 1 for $x \to 0$, we have

$$\mathbb{P}\{X < \delta\} \approx \frac{\delta^{k_x}}{(\theta_x)^{k_x}\Gamma(k_x+1)} \tag{3.51}$$

for a small positive value $\delta \to 0$. Hence at high SNR the outage probability can be approximated by

$$p_{\text{out}}(R) \approx p_{\text{out}}^{\text{U}}(R) \triangleq \frac{(2^R - 1)^{k_x}}{(\theta_x)^{k_x}\Gamma(k_x+1)}\gamma^{-k_x}. \tag{3.52}$$

From (3.52), we see a diversity gain of k_x. It can be readily verified that k_x linearly increases with $\bar{K}$ and satisfies $1 \leq k_x \leq \bar{K}+1$ with equality if and only if $\bar{K} = 0$. As shown in Fig. 2, $p_{\text{out}}^{\text{U}}(R)$ tracks very well the trend of $p_{\text{out}}(R)$ in the high SNR region, and the increase of $\bar{K}$ leads to an increase in the diversity gain.

Next, we consider the benchmark case with unit phase shifts at the RIS, i.e., $\boldsymbol{\Phi} = \mathbf{I}_G$, which does not require any CSI for setting the phase shifts and thus can dispense with the channel acquisition. In this case, X is the square of the sum of $\bar{K}+1$ independent $\mathcal{CN}(0, 1)$ random variables and follows Gamma distribution with parameters $k_x = 1$ and $\theta_x = \bar{K}+1$. Hence at high SNR the outage probability with the unit phase shift design can be approximated by

$$p_{\text{out}}^{\text{unit}}(R) \approx \frac{2^R - 1}{(\bar{K}+1)\gamma}. \tag{3.53}$$

Comparing (3.52) with (3.53), we can see an increase in the diversity gain due to the properly designed RIS phase shifts.
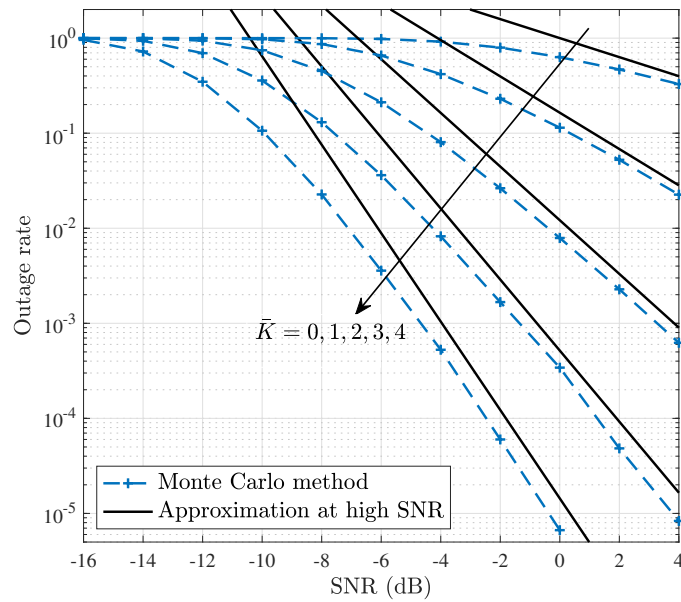
Figure 3.3: Outage rate of the proposed RIS-RPM scheme in the case of $N = 1$ assuming $\sigma_d^2 = \sigma_r^2 = 1$ and perfect CSI available at the AP, where $G = 4$, $R = 1$ and $\bar{K}$ varies from 0 to G.

3.5.2 Achievable Rate

For practical implementation, constellation $\mathbb{A}$ is typically a finite and discrete complex signal set of cardinality M with normalized power, i.e., $\mathbb{A} \triangleq \{a_m\}_{m=1}^M$ with $\mathbb{E}\{|a_m|^2\} = 1$, where the constellation points are independent and equiprobable. Let $\mathbb{S} \triangleq \{\mathbb{S}_j\}_{j=1}^J$ denote the index set of all the possible combinations of the $\bar{K}$ ON-state groups with the cardinality of $J = \binom{G}{\bar{K}}$, where $\mathbb{S}_j$ is the j-th element of set $\mathbb{S}$ representing the j-th index combination realization. Assume that all the combination realizations in $\mathbb{S}$ are independent and equiprobable.

Beamforming Design Based on Statistical ON/OFF State Information

First, we focus on the RIS-RPM scheme with the practical beamforming design based on statistical ON/OFF state information of the RIS. We rewrite the system model of (4.49) as

$$y = \sqrt{P_t} \left(\mathbf{s}^T \mathbf{\Phi H w} + \mathbf{h}_d^H \mathbf{w} \right) x + n. \tag{3.54}$$

Let $\tilde{\mathbf{g}}_r = \mathbf{\Phi H w}$, whose entries are the effective channels of the cascaded AP-RIS-user links perceived by the user associated with the corresponding RIS-elements groups. Let $\tilde{g}_d = \mathbf{h}_d^H \mathbf{w}$ denote the effective channel of the direct AP-user link perceived by the user. Before recovering the information to be sent by the AP and RIS, the user generally has to acquire the knowledge of $\tilde{\mathbf{g}}_r$ and $\tilde{g}_d$. Note that the knowledge of $\tilde{\mathbf{g}}_r$ can be obtained at the user based on downlink pilot training by using $\mathbf{\Phi}$ as the RIS reflection pattern, while the knowledge of $\tilde{g}_d$ can be obtained by applying conventional channel estimation methods with all RIS-elements groups turned OFF. Therefore, we assume

$$\bar{R}_{\text{RIS-RPM}} = \log_2 J + \log_2 M - \log_2 e$$

$$- \frac{1}{J}\frac{1}{M}\sum_{j=1}^{J}\sum_{m=1}^{M}\mathbb{E}_{\tilde{\mathbf{g}}_r,\tilde{g}_d,v}\left\{\log_2\sum_{j'=1}^{J}\sum_{m'=1}^{M}e^{-\frac{\left|\frac{v}{\sqrt{P_t}}+\left(\sum\limits_{k\in\mathbb{S}_j}[\tilde{\mathbf{g}}_r]_k+\tilde{g}_d\right)a_m-\left(\sum\limits_{k\in\mathbb{S}_{j'}}[\tilde{\mathbf{g}}_r]_k+\tilde{g}_d\right)a_{m'}\right|^2}{\sigma^2/P_t}}\right\}$$

(3.56)

that the knowledge of $\tilde{\mathbf{g}}_r$ and $\tilde{g}_d$ is available at the user side, and the achievable rate of the RIS-RPM scheme with the practical beamforming design is given by

$$\bar{R}_{\text{RIS-RPM}} = \mathbb{E}_{\tilde{\mathbf{g}}_r,\tilde{g}_d}\left\{\mathsf{I}\left(\mathbb{I},x;y|\tilde{\mathbf{g}}_r,\tilde{g}_d\right)\right\}$$

(3.55)

where $\mathsf{I}(X,Y;Z)$ denotes the mutual information between the random vector (X,Y) and random variable Z.

Proposition 1. *The achievable rate of the RIS-RPM scheme with the practical beam-forming design is given by (3.56), which is shown at the top of the page, where v is a complex Gaussian random variable following $\mathcal{CN}(0,\sigma^2)$.*

Proof. According to the definition of mutual information, (3.55) can be derived as

$$\bar{R}_{\text{RIS-RPM}} = \mathsf{H}\left(\mathbb{I},x\right) - \mathbb{E}_{\tilde{\mathbf{g}}_r,\tilde{g}_d}\left\{\mathsf{H}\left(\mathbb{I},x|y,\tilde{\mathbf{g}}_r,\tilde{g}_d\right)\right\}$$

(3.57)

where $\mathsf{H}(\cdot)$ and $\mathsf{H}(\cdot|\cdot)$ denote the marginal entropy and conditional entropy, respectively. Due to the independence between the selection of $\mathbb{I}$ and symbol modulation at the AP, we have $\mathsf{H}(\mathbb{I},x) = \log_2 J + \log_2 M$. The last term at the right hand side of (3.57) can be expressed according to the definition of conditional entropy as

$$\mathsf{H}\left(\mathbb{I},x|y,\tilde{\mathbf{g}}_r,\tilde{g}_d\right) = \frac{1}{J}\frac{1}{M}\sum_{j=1}^{J}\sum_{m=1}^{M}\int_y p\left(y|\mathbb{I}=\mathbb{S}_j,x=a_m,\tilde{\mathbf{g}}_r,\tilde{g}_d\right)$$

$$\times \log_2\frac{p(y)}{\frac{1}{J}\frac{1}{M}p\left(y|\mathbb{I}=\mathbb{S}_j,x=a_m,\tilde{\mathbf{g}}_r,\tilde{g}_d\right)}dy$$

(3.58)

where

$$p\left(y|\mathbb{I}=\mathbb{S}_j,x=a_m,\tilde{\mathbf{g}}_r,\tilde{g}_d\right) = \frac{1}{\pi\sigma^2}e^{-\frac{\left|y-\sqrt{P_t}\left(\sum\limits_{k\in\mathbb{S}_j}[\tilde{\mathbf{g}}_r]_k+\tilde{g}_d\right)a_m\right|^2}{\sigma^2}}$$

(3.59)

and

$$p(y) = \frac{1}{J}\frac{1}{M}\frac{1}{\pi\sigma^2}\sum_{j'=1}^{J}\sum_{m'=1}^{M}e^{-\frac{\left|y-\sqrt{P_t}\left(\sum\limits_{k\in\mathbb{S}_{j'}}[\tilde{\mathbf{g}}_r]_k+\tilde{g}_d\right)a_{m'}\right|^2}{\sigma^2}}.$$

(3.60)

$$\bar{R}_{\text{RIS-RPM}}^{\text{UB}} = \log_2 J + \log_2 M - \log_2 e$$
$$- \frac{1}{J}\frac{1}{M}\sum_{j=1}^{J}\sum_{m=1}^{M}\mathbb{E}_{\mathbf{H},\mathbf{h}_d,v}\left\{\log_2\sum_{j'=1}^{J}\sum_{m'=1}^{M}e^{-\frac{\left|v+\sqrt{P_t}f(\mathbb{S}_j,\mathbf{H},\mathbf{h}_d)a_m-\sqrt{P_t}f(\mathbb{S}_{j'},\mathbf{H},\mathbf{h}_d)a_{m'}\right|^2}{\sigma^2}}\right\}$$

$$(3.64)$$

Replacing y with $v \triangleq y - \sqrt{P_t}\left(\sum_{k\in\mathbb{S}_j}[\tilde{\mathbf{g}}_r]_k + \tilde{g}_d\right)a_m$ yields

$$\mathsf{H}\left(\mathbb{I},x|y,\tilde{\mathbf{g}}_r,\tilde{g}_d\right) = -\log_2\pi\sigma^2 + \frac{1}{J}\frac{1}{M}\sum_{j=1}^{J}\sum_{m=1}^{M}\int_v p(v)$$

$$\times\log_2\sum_{j'=1}^{J}\sum_{m'=1}^{M}e^{-\frac{\left|\frac{v}{\sqrt{P_t}}+\left(\sum_{k\in\mathbb{S}_j}[\tilde{\mathbf{g}}_r]_k+\tilde{g}_d\right)a_m-\left(\sum_{k\in\mathbb{S}_{j'}}[\tilde{\mathbf{g}}_r]_k+\tilde{g}_d\right)a_{m'}\right|^2}{\sigma^2/P_t}}dv + \frac{1}{J}\frac{1}{M}\sum_{i=1}^{J}\sum_{j=1}^{M}\int_v p(v)\log_2\frac{1}{p(v)}dv$$

$$(3.61)$$

where $p(v) = \frac{1}{\pi\sigma^2}\exp(-\frac{|v|^2}{\sigma^2})$ is the PDF of a complex Gaussian random variable with zero mean and variance σ^2. Since the last term at the right hand side of (3.61) is the differential entropy of a $\mathcal{CN}(0,\sigma^2)$ random variable, which is equal to $\log_2\pi\sigma^2 e$, we finally obtain the expression of $\bar{R}_{\text{RIS-RPM}}$ as (3.56). □

Beamforming Design Based on Instantaneous ON/OFF State Information

Next, we characterize the upper bound on the achievable rate of the RIS-RPM scheme, where the active and passive beamforming vectors are optimized based on the instantaneous ON/OFF state information of the RIS. In this case, beamforming $\mathbf{w}$ and $\boldsymbol{\theta}_\mathbb{I}$ highly depend on the selection of $\mathbb{I}$. Define

$$f(\mathbb{I},\mathbf{H},\mathbf{h}_d) \triangleq \left(\boldsymbol{\theta}_\mathbb{I}^T\mathbf{H}_\mathbb{I} + \mathbf{h}_d^H\right)\mathbf{w} \qquad (3.62)$$

which is the effective channel perceived by the user and can be obtained by downlink channel training. By assuming that the knowledge of $f(\mathbb{I},\mathbf{H},\mathbf{h}_d)$ is available at the user side, the achievable rate of the RIS-RPM scheme with the beamforming design presented in Section 3.4 is given by

$$\bar{R}_{\text{RIS-RPM}}^{\text{UB}} = \mathbb{E}_{\mathbf{H},\mathbf{h}_d}\left\{\mathsf{I}\left(\mathbb{I},x;y|\mathbf{H},\mathbf{h}_d\right)\right\}. \qquad (3.63)$$

Proposition 2. *The achievable rate of the RIS-RPM scheme with the beamforming design presented in Section 3.4 is given by (3.64), which is shown at the top of the next page, where v is a complex Gaussian random variable following $\mathcal{CN}(0,\sigma^2)$.*

Proof. The proof is similar to (3.56) and omitted for brevity. □

3.6   Simulation Results and Discussions

In this section, simulation results are presented to evaluate the performance of our proposed schemes. We consider a three dimensional coordinate system, where the centers of the AP and RIS are located at $(0, 0, 0)$ and $(0, d_0, 0)$, respectively, and the user is located at $(0, d_y, d_z)$. For the AP, we consider a uniform linear array of $N = 4$ antennas with an antenna spacing of half-wavelength, located in x-axis. For the RIS, we consider a uniform square array of $L = 12 \times 12 = 144$ elements with an element spacing of half-wavelength, deployed in x-z plane. The fading channel model for all the channels involved is given by

$$h = \sqrt{\frac{\kappa \varpi}{\kappa + 1}} e^{-j 2\pi d / \vartheta} + \sqrt{\frac{1}{\kappa + 1}} \mathcal{CN}(0, \varpi)$$

where κ is the Rician factor, ϖ is the path loss, d is the signal propagation distance, and ϑ is the signal wavelength. We resort to the simplified path loss model, i.e., $\varpi = C_0 d^{-\alpha}$, where C_0 is the path loss at a reference distance of 1 meter (m), and α is the path loss exponent. The Rician factors of the AP-user, AP-RIS, and RIS-user links are set as $\kappa_{\mathsf{Au}} = 0$, $\kappa_{\mathsf{AR}} = \infty$, and $\kappa_{\mathsf{Ru}} = 0$, respectively, as in Section 3.5.1. The path loss exponents of the AP-user, AP-RIS, and RIS-user links are set as 3.8, 2.2, and 2.4, respectively. The noise power is set as $\sigma^2 = -80$ dBm. Other parameters are set as follows: $d_0 = 50$ m, $d_z = 2$ m, $\vartheta = 0.1$ m, $T_c = 150$ symbol sampling periods, $C_0 = 30$ dB, and quadrature phase-shift keying (QPSK) for symbol modulation at the AP. We consider the Zadoff-Chu sequence as the pilot sequence and set $P_p = 10$ dBm for channel estimation. The maximum iteration number in Algorithm 3.2 is set as $I = 5$ and the threshold is set as $\epsilon = 10^{-4}$. The values of G and $\bar{K}$ as well as the AP's transmit power level will be specified later to study their effects on the system performance. In the following simulations, the results are obtained by averaging over more than 1000 independent channel realizations.

3.6.1   Performance of Algorithm 1

To evaluate the effectiveness of Algorithm 3.2, we consider the following schemes with fixed $G = 4$ and $P_t = 20$ dBm: 1) Benchmark scheme without information transfer (IT) [48] where all RIS elements are turned ON and problem (P2) with $\mathbb{I} = \mathbb{G}$ is solved to obtain the active and passive beamforming vectors; 2) Conventional MISO scheme without RIS in which $\mathbf{w} = \hat{\mathbf{h}}_d / \lVert \hat{\mathbf{h}}_d \rVert$; 3) Random phase shift scheme where the phase shifts of diagonal entries in $\mathbf{\Phi}$ are randomly drawn from $(0, 2\pi]$ and then problem (P1.1) is solved to obtain the active beamforming vector; 4) Upper bound that solves problem (P2); 5) PBIT scheme where elements in the ON/OFF-state vector $\mathbf{s}$ are independently drawn from the set $\{0, 1\}$ with equal probability. For the PBIT scheme, the active and passive beamforming vectors are obtained by solving problem (P1) with $\mathbf{A} = \frac{1}{4}(\mathbf{1}_G \cdot \mathbf{1}_G^T + \mathbf{I}_G)$ and $\mathbf{a} = \frac{1}{2} \cdot \mathbf{1}_G$. The number of ON-state groups in Algorithm 3.2 is set as $\bar{K} = 3$. Note that schemes 1) and 2) can be recognized as two special cases of Algorithm 1 with $\bar{K} = G$ and $\bar{K} = 0$, respectively. Firstly, it can be observed that when the user locates in the vicinity of the RIS, all the RIS-assisted schemes significantly enhance the average received power at the user as compared to the scheme without RIS. Secondly, as shown in Fig. 3.4 and Fig. 3.5 , Algorithm 3.2 significantly outperforms the random
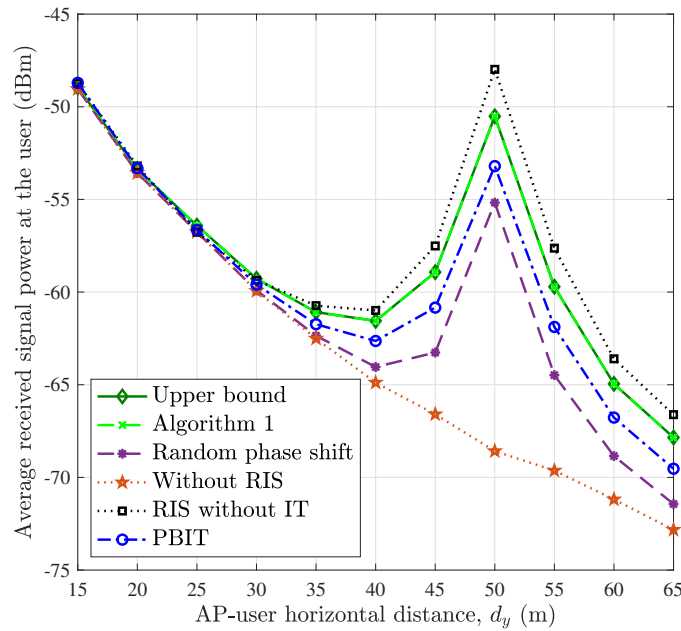
Figure 3.4: Average received signal power at the user versus AP-user horizontal distance d_y, where $G = 4$, $P_t = 20$ dBm, and $\bar{K} = 3$ in Algorithm 3.2.Perfect CSI at the AP

phase shift scheme. Moreover, as expected in Section 3.4, the average received power achieved by Algorithm 1 is upper-bounded by the beamforming design based on the instantaneous ON/OFF state information of the RIS. On the other hand, it is worth pointing out that Algorithm 3.2 incurs small received signal power loss as compared to the scheme without IT. This is expected since in Algorithm 3.2, one RIS-elements group is turned OFF deliberately to convey additional information of the RIS, while all RIS-elements groups are used for enhancing the reflected signal power in the latter scheme. The impact of channel estimation error on the performance of active and passive beamforming is also shown in Fig. 3.4 and Fig. 3.5. We can observe that as compared to the upper bound, Algorithm 3.2 with perfect CSI at the AP achieves nearly the same performance, while with estimated CSI it performs worse when the user is far away from both the AP and RIS. This can be explained by the fact that given the same transmit power at the user, received training signal power is lower when the user moves far away from both the AP and RIS, and thus the channel estimation accuracy is reduced.

3.6.2   Outage Rate Performance

In Fig. 3.6, we consider the PBIT scheme where elements in $\mathbf{s}$ are independently drawn from the set $\{0, 1\}$ with equal probability. Under the PBIT scheme, the average number of ON-state groups at the RIS is $\mathbb{E}\{\mathbf{s}^T \mathbf{s}\} = G/2$. The outage rates of our proposed RIS-RPM scheme with $\bar{K} = G - 1$ and $\bar{K} = G/2$ are plotted. Two benchmark schemes are considered: 1) Benchmark scheme without IT; 2) Conventional MISO scheme without RIS. Fig. 3.6 shows the outage rate performance of different schemes versus the AP's transmit power, where $G = 6$, $R = 1$ and $d_y = 45$ m. We can observe that the outage rate performance of the proposed RIS-RPM scheme with $\bar{K} = G/2 = 3$ outperforms
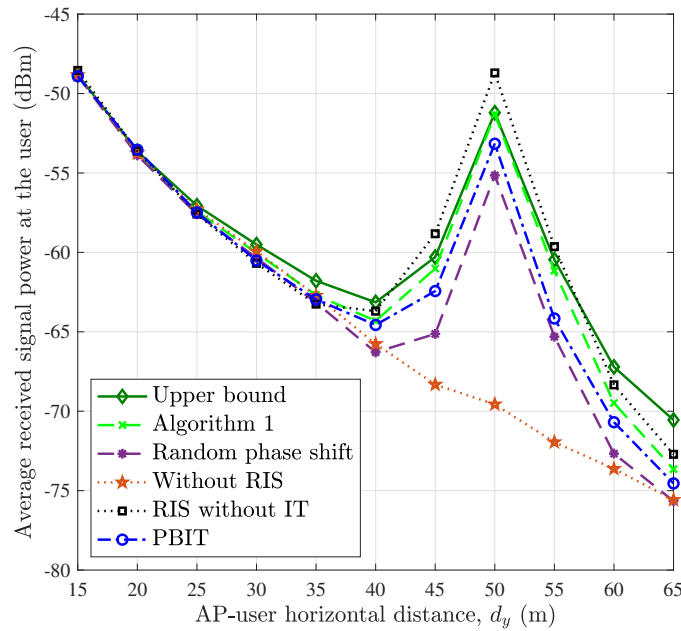
Figure 3.5: Average received signal power at the user versus AP-user horizontal distance d_y, where $G = 4$, $P_t = 20$ dBm, and $\bar{K} = 3$ in Algorithm 3.2.Estimated CSI at the AP, $P_p = 10$ dBm
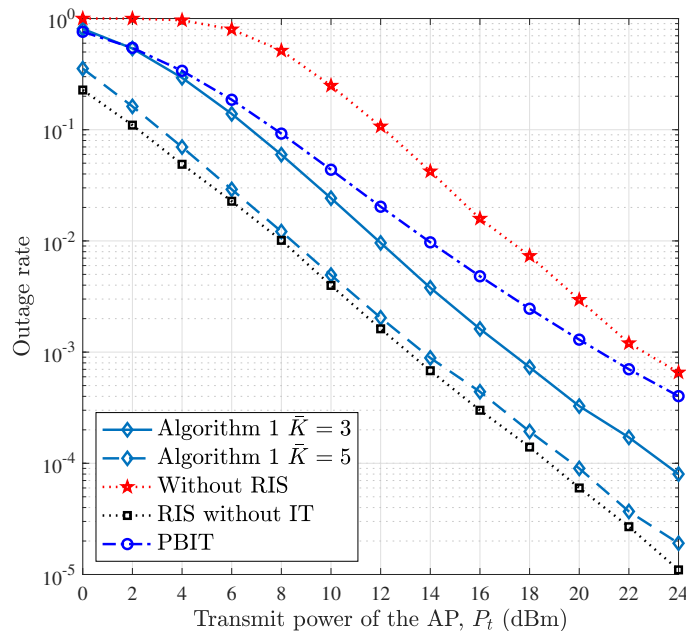
Figure 3.6: Performance comparison between the proposed RIS-RPM scheme and the PBIT counterpart, where $G = 6$, $R = 1$, and $d_y = 45$ m.

that of the PBIT counterpart. This can be understood by the fact that different from the PBIT counterpart that suffers from large fluctuation in the reflected signal power due to the varying number of ON-state elements, the RIS-RPM scheme keeps the number of ON-state elements at each time constant to reduce such power fluctuation, thus

Figure 3.7: Effect of AP-user horizontal distance d_y on the achievable rate, where $G = 4$, and $P_t = 0$ dBm.

showing better outage rate performance. Moreover, by increasing the number of ON-state groups $\bar{K}$, the outage rate performance of the RIS-RPM scheme can be further improved.

3.6.3   Achievable Rate Performance

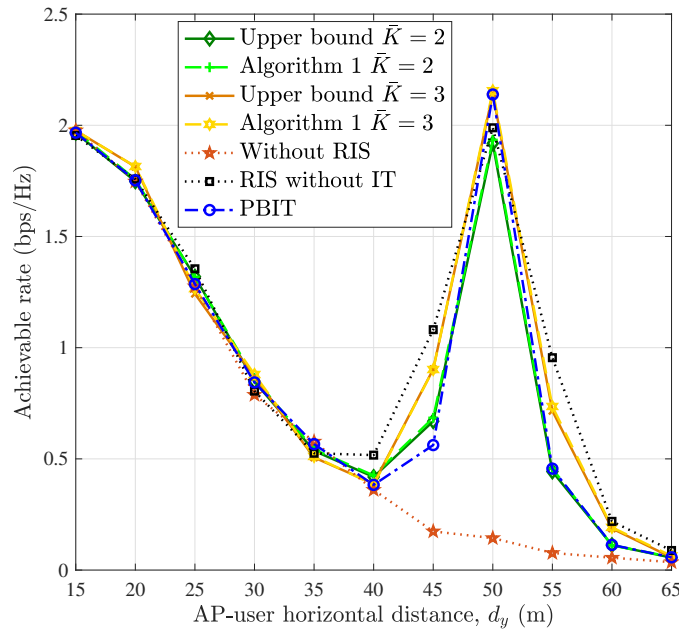To evaluate the achievable rate performance of the RIS-RPM scheme, the following schemes are considered: 1) Benchmark scheme without IT; 2) Conventional MISO scheme without RIS; 3) Upper bound that computes the achievable rate of (3.64); 4) PBIT scheme where elements in s are independently drawn from the set $\{0, 1\}$ with equal probability. In the sequel, the active and passive beamforming vectors are obtained based on the estimated CSI.

Effect of AP-User Horizontal Distance d_y

In Fig. 3.9, we evaluate the effect of d_y on the achievable rate, where $G = 4$, $\bar{K} = \{2, 3\}$ and $P_t = \{0 \text{ dBm}, 20 \text{ dBm}\}$. One can observe from Fig. 3.7 that, when the transmit power of the AP is very low, the achievable rate of the scheme without RIS decreases as the user moves away from the AP, and approaches zero. In contrast, the achievable rates of those schemes assisted by the RIS increase drastically as the user moves toward the RIS, and decrease as the user moves away from both the AP and RIS. This is because when the user is close to either the AP or RIS, it is able to receive stronger transmitted/reflected signals from the AP/RIS. This phenomenon implies that the cell-edge user can benefit from an RIS deployed in its neighborhood, i.e., the rate of the cell-edge user can be enhanced by deploying an RIS, instead of improving the AP's transmit power or deploying an expensive AP/relay. On the other hand, we observe
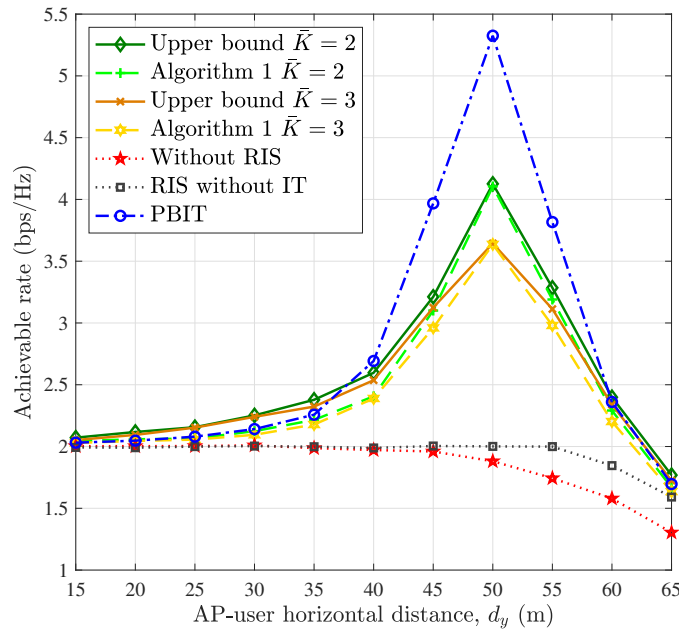
Figure 3.8: Effect of AP-user horizontal distance d_y on the achievable rate, where $G = 4$, and $P_t = 20$ dBm.

from Fig. 3.7 that the RIS-RPM scheme with $\bar{K} = 3$ has the potential to outperform the one without IT despite that the received signal power achieved by the RIS-RPM scheme is lower than that achieved by the one without IT as shown in Fig. 3.5. For the RIS-RPM scheme with $\bar{K} = 2$, since a large number of RIS elements are turned OFF deliberately for information transfer, the additional information from the RIS cannot compensate for the information reduction caused by the loss of received signal power, thus leading to a smaller achievable rate than the scheme without IT. Moreover, it can be observed from Fig. 3.8 that when the AP's transmit power is high, the RIS-RPM scheme exhibits significantly superior rate performance over the scheme without IT, which is attributed to the additional information delivered by the RIS. This implies that the RIS-RPM scheme provides a mechanism to enable a flexible tradeoff between the received signal power and achievable rate performance by varying the number of OFF-state groups at the RIS. Moreover, when the transmit power of the AP is high, the PBIT scheme achieves the maximum achievable rate, as expected.

Effect of Number of ON-State RIS-elements Groups $\bar{K}$

We compare the achievable rate of the RIS-RPM scheme versus $\bar{K}$ in Fig. 3.9, where $G = 9$, $d_y = 45$ m, and $P_t = \{10\,\text{dBm}, 30\,\text{dBm}\}$ are considered. One can observe that there exists an optimal $\bar{K}$, which varies with different AP's transmit power levels. For $P_t = 30$ dBm, the optimal $\bar{K}$ is 5 whereas the optimal $\bar{K}$ for $P_t = 10$ dBm is 6. Generally, it is expected that when the AP's transmit power is high enough, the optimal $\bar{K}$ is more likely to be $\lceil G/2 \rceil$, since the entropy of the RIS is maximized. Furthermore, when the AP's transmit power is very high, due to the assumption of finite-alphabet input, the achievable rate at different $\bar{K}$ values will be the sum of the corresponding uncoded transmitted information rates of the AP and RIS. In contrast, when the AP's transmit
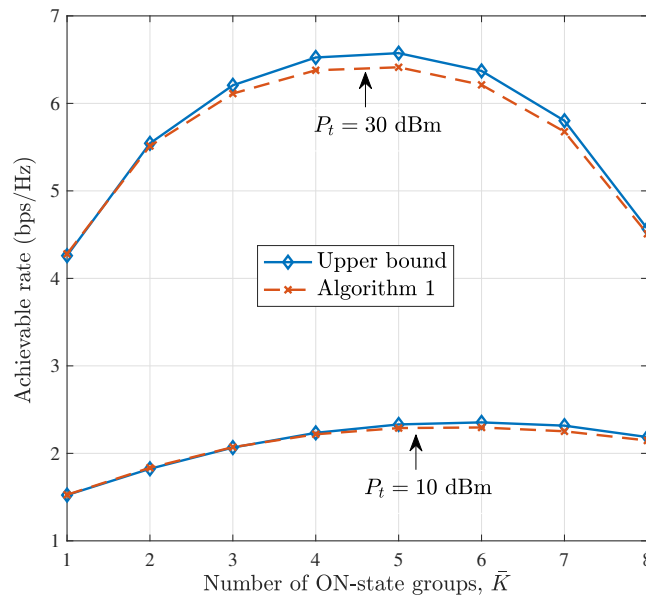
Figure 3.9: Effect of $\bar{K}$ on the achievable rate, where $G = 9$, $d_y = 45$ m, and P_t equal to 10 dBm and 30 dBm are considered.

Table 3.2: Grouping Ratio and Channel Estimation Overhead Ratio

G	2	4	6
ρ	1/72	1/36	1/24
ξ	3/150	5/150	7/150
$L_x \times L_z$	12×6	6×6	6×4

Each group consists of $\bar{L} = L_x \times L_z$ elements with L_x elements along x-axis and L_z elements along z-axis.

power becomes very low, the optimal $\bar{K}$ is more likely to be $(G - 1)$, since the reflected signal power is maximized while the RIS still can convey its information through RPM.

Effect of RIS-elements Grouping Ratio

The RIS-elements grouping ratio is defined by $\rho \triangleq 1/\bar{L}$. Let $\xi \triangleq (G + 1)/T_c$ denote the ratio of time overhead for channel estimation to the coherence time normalized to the symbol sampling period. In Fig. 3.10, we examine the effect of ρ on the achievable rate, where $d_y = 45$ m, $G = \{2, 4, 6\}$, and $\bar{K} = G/2 = \{1, 2, 3\}$ for the proposed RIS-RPM scheme. Note that the average number of ON-state RIS elements keeps constant for different schemes. As can be seen, the RIS-RPM scheme with large G can achieve better achievable rate performance. This phenomenon can be explained by the fact that with large grouping ratio, not only the degrees of freedom for RIS reflection design increases, achieving high passive beamforming gain, but also more additional information can be conveyed through the index combination of RIS-elements groups, both improving the sum achievable rate. However, as shown in Table 3.2, the pilot overhead ratio ξ increases with the grouping ratio, which results in more time for channel estima-
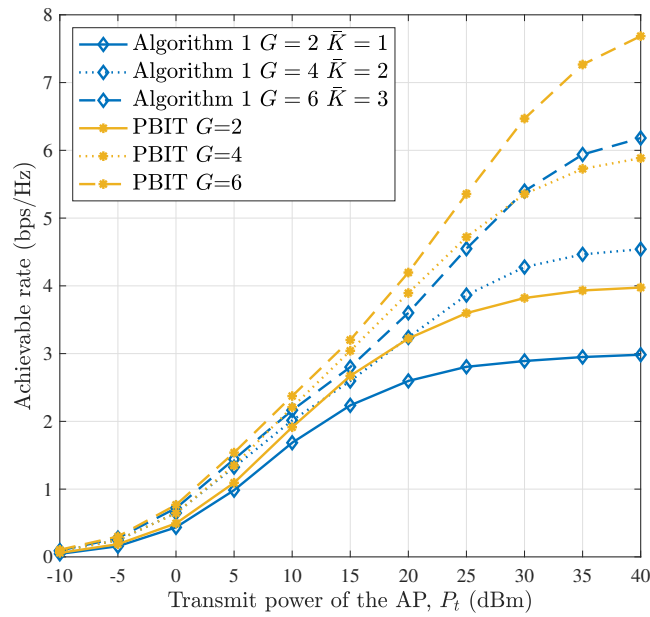
Figure 3.10: Effect of RIS-elements grouping ratio on the achievable rate, where $d_y = 45$ m and the average number of ON-state RIS elements at each time is $\bar{L} \times \bar{K} = 72$.

tion and less time for data transmission, thus reducing the average achievable rate. On the other hand, we can observe that a large AP's transmit power level is needed for the PBIT scheme to be competitive.

3.7 Conclusions

In this section, we considered an RIS-enhanced MISO wireless communication system and proposed the RPM scheme for the dual-use of passive beamforming and information transfer of the RIS. A practical beamforming design based on the RIS's statistical ON/OFF state information was proposed to maximize the *average* received signal power at the user, for which an efficient algorithm based on the alternating optimization technique was proposed to obtain a high-quality solution. Next, we formulated an optimization problem to maximize the *instantaneous* received signal power by designing active and passive beamforming based on the RIS's instantaneous ON/OFF state information, which characterized the upper bound on the received signal power of the RIS-RPM scheme. Moreover, the asymptotic outage probability of the RIS-RPM scheme over Rayleigh fading channels was derived in closed-form. In particular, the RIS was shown to be able to increase the diversity gain by properly designing the phase shifts of its elements. The achievable rate of the RIS-RPM scheme has been analyzed for the case where the transmitted symbol was drawn from a finite constellation. Finally, simulation results corroborated the effectiveness of Algorithm 3.2 as well as the RIS-RPM scheme and revealed the effect of different system parameters on the achievable rate performance of the RIS-RPM scheme. It was shown that the RIS-RPM scheme was able to improve the achievable rate performance despite the loss in received signal power as compared to the conventional RIS-assisted system with full-ON reflection.

Chapter 4

Reconfigurable Intelligent Surface-Aided Quadrature Reflection Modulation for Simultaneous Passive Beamforming and Information Transfer

Reflection modulation based on RIS is considered to be a promising information transfer mechanism without requiring any additional radio frequency chains. However, existing reflection modulation schemes consider manipulating the ON/OFF states of RIS elements, which suffers from power loss. In this section, we propose a new scheme, called RIS-aided quadrature reflection modulation (RIS-QRM), for harvesting the reflection power based on an RIS-aided downlink MISO wireless system. To this end, RIS-QRM partitions RIS elements into two subsets for reflecting impinging signals into two orthogonal directions as well as passive beamforming and encodes its local data onto the element partition options. A closed-form expression for the unconditional pairwise error probability of RIS-QRM in Rician fading is derived assuming the maximum-likelihood detection. Moreover, we propose a low-complexity detection method with compromised performance for decoupling the joint search of the constellation symbol and the RIS element partition. Computer simulation results corroborate the effectiveness of RIS-QRM and validate the analytical results. It is shown that RIS-QRM is able to improve the error performance of the additional bits delivered by the RIS without deteriorating that of the bits carried on the constellation symbol, as compared to ON/OFF-based schemes.

4.1   Introduction

By properly adjusting the phase shifts of all the elements of a RIS, the reflected signals can be added constructively at the receiver to improve the SNR or destructively to mitigate interference/eavesdropping, which is similar to the functionality of traditional beamforming and is referred to as passive beamforming. RISs resemble multi-antenna amplify-and-forward relays with a full-duplex protocol but without signal amplification. In particular, RISs do not require any costly RF chains for signal reception and re-

transmission, nor complex signal processing techniques for self-interference cancellation; thereby greatly reducing hardware cost and energy consumption [46, 71]. These appealing advantages have driven a variety of wireless communications empowered by RISs, such as RIS-aided MIMO systems [72–74], RIS-aided energy transfer [52], RIS-aided orthogonal frequency division multiplexing (OFDM) systems over frequency-selective channels [59], and so on. The authors of [75] realized an RIS-based 8-phase shift-keying (8-PSK) transmitter that uses an RIS with 256 reconfigurable elements instead of an RF chain. In [76], the authors realized an RIS-based 16-quadrature amplitude modulation (16-QAM) transmitter and an RIS-based space-down-conversion receiver for a 2×2 MIMO transmission. In particular, a theoretical performance comparison between orthogonal multiple access (OMA) and non-orthogonal multiple access (NOMA) in RIS-aided downlink communications was investigated in [77]. Moreover, since the proper configuration of RIS elements needs channel state information (CSI) of all links involved in RIS-aided systems, various RIS channel estimation methods have been proposed, e.g., [49, 53, 55, 78, 79].

4.1.1 Prior Works

Spatial modulation (SM), which uses only a single RF chain to transmit a constellation signal and conveys additional information by the ON/OFF states of transmit antennas, has also received much attention as it is capable of striking an attractive tradeoff between energy efficiency and spectral efficiency [80, 81]. A well-known variant of SM, called quadrature SM, improves its overall throughput by extending conventional SM to in-phase and quadrature dimensions [82, 83]. A recent generalization is to implement SM by an RIS. In [61], the author proposed to carry out SM on receive antenna indices by an RIS that steers a beam towards a particular receive antenna according to the input information bits. Nonetheless, this work still focused on using the RIS to improve communication spectral efficiency. In particular, the authors of [62] demonstrated that a fixed RIS configuration which maximizes the SNR at the receiver is a suboptimal scheme in terms of achievable rate, and proposed to jointly encode information in the transmitted signal and in the RIS configuration [84, 85].

On the other hand, a new scenario where an RIS needs to upload its local data to a base station (BS) was considered in [86] and [66], for which a passive beamforming and information transfer (PBIT) scheme was proposed to improve the communication performance from the user to the BS using passive beamforming at the RIS and simultaneously deliver additional information from the RIS to the BS by the ON/OFF states of RIS elements. There are many prospective sources of RIS local data [64]. For example, if an RIS is embedded into a wall/ceiling with sensors for environment monitoring or into an object (e.g., a smart t-shirt) for health monitoring, it needs to report the sensed data to an intended receiver, e.g., the wireless network or a mobile phone [63]. The PBIT scheme, however, suffers from a relatively high link outage probability, since the number of activated RIS elements varies over time, causing significant fluctuations in the reflected signal power. To minimize the outage probability of the communication link between transceivers, the authors of [87] proposed to fix the number of RIS elements switched on at each time, which is referred to as the RIS-based reflection pattern modulation (RIS-RPM) scheme. Existing RIS-based SM schemes [66, 87, 88] that use the ON/OFF states of RIS elements to carry information, however, have two main draw-

backs. Firstly, since only a portion of the available RIS elements are switched on at each time, the aperture gain provided by the RIS is reduced, which limits the effective received signal power. Secondly, due to the inherent randomness of the RIS configuration dependent on the incoming random bits, joint optimization of the beamforming at the transmitter/receiver and the RIS configuration is a stochastic optimization problem that is difficult to solve in general.

4.1.2   Main Contributions

To address the above issues, we consider in this chapter an RIS-aided downlink MISO wireless communication system, for which a new reflection modulation scheme is proposed to maximize the SNR at the user by using passive beamforming at the RIS and simultaneously convey additional information from the RIS to the user in a passive and full-on reflection[1] manner. Our main contributions are summarized as follows:

- We propose a new reflection modulation scheme, referred to as RIS-QRM, for simultaneous passive beamforming and information transfer of RIS. To be specific, the reflecting elements of the RIS are randomly partitioned into two subsets according to its local data: the phase shifts of the first subset are tuned to make the signals reflected by the first subset in phase or aligned with the signal coming directly from the AP, while the phase shifts of the second subset are designed to make the reflected signals by the second subset orthogonal to those reflected by the first subset, where the RIS element partition carries the local data of RIS. Unlike ON/OFF-based reflection modulation schemes (e.g., [66] and [87]) in which only a portion of the available RIS elements reflect impinging waves at each time instance, the proposed RIS-QRM scheme leverages the large-scale aperture gain of RIS by keeping all the RIS elements activated. Moreover, the joint optimization of the active beamforming at the AP and the phase shifts of RIS elements to maximize the SNR no longer needs the knowledge of the real-time RIS configuration, getting rid of difficult stochastic optimization.

- We derive a closed-form expression for the unconditional pairwise error probability (PEP) of the RIS-QRM scheme for an RIS-aided single-input single-output (SISO) setting over Rician fading channels, by assuming the maximum-likelihood (ML) detection and that the AP-user direct link is blocked, which is used to calculate tight upper bounds on the average bit error probabilities (BEPs) of the conventional constellation modulation and the RIS-modulated data through the element partition. It is worth noting that from the perspective of practical implementation, adjacent RIS elements are usually grouped to have the same configuration for reducing complexity [55]. As a result, the number of effective channels is greatly reduced, and the framework for the calculation of the BEP developed in [61] may be not applicable to our case as the number of effective channels may be insufficient to satisfy the hypothesis of the central limit theorem. In addition, we propose a low-complexity detection method with lower performance for decoupling the joint detection of the constellation symbol and the RIS element partition.

---

[1]All of the available RIS elements reflect impinging signals at all time.

- Finally, extensive computer simulation results are presented to corroborate the effectiveness of the proposed RIS-QRM scheme and validate our analytical results. Particularly, in simultaneous passive beamforming and information transfer schemes, there exists a fundamental tradeoff between the error performance of the bits carried on the constellation symbol and that of the additional bits modulated by the RIS through the element partition. Moreover, it is shown that the RIS-QRM scheme not only improves the error performance of the bits carried on the constellation symbol, but also improves that of the additional bits carried on the element partition of the RIS, as compared to the PBIT scheme. Compared to the RIS-RPM scheme, the RIS-QRM scheme is able to improve the error performance of the additional bits carried on the element partition of the RIS without deteriorating that of the bits carried on the constellation symbol.

4.1.3 Organization and Notation

The rest of this chapter is organized as follows. In Section 4.2, we present the system model and the main idea of RIS-QRM. In Section 4.3, we derive closed-form expressions for the upper bounds on the BEPs of the constellation symbol and the RIS element partition. In Section 4.4, we generalize the RIS-QRM scheme for application to the MISO case where there is a direct channel between the transceivers and propose a low-complexity detector. Computer simulation results are provided in Section 4.5. Finally, Section 4.6 concludes the chapter.

In this chapter, italic letters, uppercase and lowercase boldface letters (e.g., x, $\mathbf{X}$ and $\mathbf{x}$) denote scalars, matrices and column vectors, respectively. Uppercase blackboard bold letters (e.g., $\mathbb{A}$) denote discrete and finite sets. Superscripts †, ^{T}, ^{H}, and $^{-1}$ represent complex conjugate, transpose, Hermitian transpose, and inversion operations, respectively. $\mathbf{I}_n$, $\mathbf{1}_{n \times m}$, and $\mathbf{0}_{n \times m}$ denote an identity matrix of size $n \times n$, an all-one matrix of size $n \times m$, and an all-zero matrix of size $n \times m$, respectively. $\mathbb{C}^{n \times m}$ denotes the space of $n \times m$ complex-valued matrices. For a complex number x, $|x|$ and $\angle x$ denote the absolute value and phase, respectively, and $\mathrm{Re}(x)$ and $\mathrm{Im}(x)$ denote the real and imaginary parts of x, respectively. For a complex-valued vector $\mathbf{x}$, $\|\mathbf{x}\|$ denotes the Euclidean norm, and $\mathrm{diag}(\mathbf{x})$ denotes a diagonal matrix with the elements in $\mathbf{x}$ on the main diagonal. $\lfloor \cdot \rfloor$ returns the largest integer less than or equal to the argument. $\binom{n}{m}$ denotes the binomial coefficient. $\langle \cdot, \cdot \rangle$ denotes the inner product. $\mathbb{A} \setminus \mathbb{B}$, $\mathbb{A} \cup \mathbb{B}$, and $\mathbb{A} \cap \mathbb{B}$ denote the relative complement of set $\mathbb{B}$ in set $\mathbb{A}$, the union and intersection of set $\mathbb{A}$ and set $\mathbb{B}$, respectively. $\mathbb{E}\{\cdot\}$ and $\mathrm{Var}\{\cdot\}$ denote the expectation and variance of a random variable, respectively. $\mathrm{tr}(\cdot)$ denotes the matrix trace. $\mathcal{N}(\mu, \sigma^2)$ represents the normal distribution with mean μ and variance σ^2. $\mathcal{CN}(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ represents the distribution of a circularly symmetric complex Gaussian random vector with mean vector $\boldsymbol{\mu}$ and covariance matrix $\boldsymbol{\Sigma}$.

4.2 System Model and Principle of RIS-QRM

4.2.1 System Model

As illustrated in Fig. 4.1, we consider an RIS-aided wireless communication system in the downlink, where an RIS composed of L_0 reflecting elements is deployed to as-
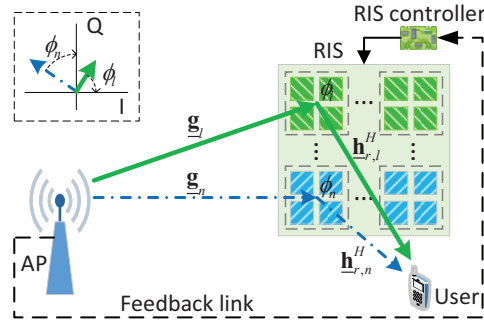
Figure 4.1: An illustration of the proposed RIS-QRM system, where the RIS focuses reflected signals towards the user while conveying additional information by the RIS partition. By adjusting the phase shifts at the RIS, the cascaded channel via green elements and that via blue elements are orthogonal.

sist the transmission from an AP to a user near the RIS. For ease of exposition, we assume in this section that both the AP and the user are equipped with a single antenna and the direct link between them is blocked due to the presence of obstacles. To reduce channel estimation overhead and implementation complexity, we recall the RIS-elements grouping method in [55], where the L_0 reflecting elements are divided into L groups, denoted by the set $\mathbb{L} \triangleq \{1, 2, \ldots, L\}$, each consisting of $L_1 = L_0/L$ (assumed to be an integer for notational convenience) adjacent elements to share the same reflection coefficient. Moreover, the RIS is connected to a smart controller, which is responsible for exchanging information with the AP via a separate feedback link and dynamically adjusting the reflection coefficient of each RIS element. Suppose that the RIS is equipped with sensors for monitoring/collecting environmental data (typically low-rate bursty data), so it needs to report sensed data to an intended receiver, e.g., the user.

In this chapter, we consider a quasi-static block fading channel model for all individual links, where the channel coefficients remain constant within each fading block and are independent and identically distributed (i.i.d.) across different fading blocks. Without loss of generality, we focus on the downlink communication in one particular fading block. Let $\mathbf{g} \in \mathbb{C}^{L_0 \times 1} \triangleq \left[\underline{\mathbf{g}}_1^H, \underline{\mathbf{g}}_2^H, \ldots, \underline{\mathbf{g}}_L^H\right]^H$ and $\mathbf{h}_r^H \in \mathbb{C}^{1 \times L_0} \triangleq \left[\underline{\mathbf{h}}_{r,1}^H, \underline{\mathbf{h}}_{r,2}^H, \ldots, \underline{\mathbf{h}}_{r,L}^H\right]$ denote the element-wise baseband channels of the AP-RIS and RIS-user links, respectively, where $\underline{\mathbf{g}}_l \in \mathbb{C}^{L_1 \times 1}$ and $\underline{\mathbf{h}}_{r,l}^H \in \mathbb{C}^{1 \times L_1}$ consist of the corresponding channel coefficients associated with the l-th group, $l = 1, 2, \ldots, L$. Let $\boldsymbol{\theta} \triangleq \left[\beta_1 e^{j\theta_1}, \beta_2 e^{j\theta_2}, \ldots, \beta_L e^{j\theta_L}\right]^H$ denote the reflection coefficients of the RIS, where $j = \sqrt{-1}$, $\beta_l \in [0, 1]$ and $\theta_l \in [0, 2\pi)$ respectively denote the common reflection amplitude and phase shift within the l-th group: all elements in this group share the same reflection coefficient $\beta_l e^{j\theta_l}$. In order to maximize the reflected power of each element and simplify the hardware design, the reflection amplitudes are fixed to the maximum value, i.e., $\beta_l = 1, \forall l = 1, 2, \ldots, L$. The baseband received signal at the user is given by

$$\underline{y} = \sum_{l=1}^{L} \underline{\mathbf{h}}_{r,l}^H e^{j\theta_l} \underline{\mathbf{g}}_l x + n \tag{4.1}$$

where x is the transmitted symbol drawn from an M-ary discrete signal constellation with transmit power equal to P_t, and $n \sim \mathcal{CN}(0, N_0)$ is the additive white Gaussian noise

(AWGN) with noise power equal to N_0. Let $\underline{h}_l = \mathbf{h}_{r,l}^H \underline{\mathbf{g}}_l$ denote the effective (aggregated) channel coefficient of the link between the AP and user via group l with the common phase shift equal to zero. By stacking $\underline{h}_l$ with $l = 1, \ldots, L$ into $\mathbf{h} = [\underline{h}_1, \underline{h}_2, \ldots, \underline{h}_L]^T$ as the AP-RIS-user cascaded channel, we have

$$y = \boldsymbol{\theta}^H \mathbf{h} x + n. \tag{4.2}$$

Since a channel estimation method has been proposed in [57] to acquire the CSI of $\mathbf{h}$, we assume in this chapter that the AP has perfect knowledge of $\mathbf{h}$ and consequently is able to optimize the reflection coefficients for performing passive beamforming at the RIS.

4.2.2   Principle of RIS-QRM

In order to implicitly convey the RIS information, we propose a mapping method that selects p $(1 \leq p < L)$ out of the L groups according to the incoming information bits, and then tunes the phase shifts of the selected groups to align the phases of the corresponding effective channels, while tuning the phase shifts of the remaining $L-p$ groups so that the effective channels associated with the remaining groups are orthogonal to the effective channels associated with the selected groups. In this manner, additional information is carried by the RIS partition. Let $\mathbb{I} = \{i_1, i_2, \ldots, i_p\}$ denote the indices of p selected groups, where $i_\partial \in \mathbb{L}$ for $\partial = 1, 2, \ldots, p$; accordingly, the remaining groups are indexed by $\mathbb{R} = \mathbb{L} \setminus \mathbb{I}$. With the proposed RIS-QRM scheme, the phase shifts of the RIS are given by

$$\phi_l = \begin{cases} \theta_l^*, & l \in \mathbb{I} \\ \theta_l^* + \pi/2, & l \in \mathbb{R} \end{cases}, \quad l = 1, 2, \ldots, L \tag{4.3}$$

where $\theta_l^* = -\angle \underline{h}_l$. The reflection coefficients of the RIS can be expressed as

$$\boldsymbol{\phi} \triangleq \left[e^{j\phi_1}, e^{j\phi_2}, \ldots, e^{j\phi_L} \right]^H = \mathsf{diag}(\boldsymbol{\theta}^*)\mathbf{s} \tag{4.4}$$

where $\boldsymbol{\theta}^* = \left[e^{j\theta_1^*}, e^{j\theta_2^*}, \ldots, e^{j\theta_L^*} \right]^H$ and $\mathbf{s} \triangleq [s_1, s_2, \ldots, s_L]^H$ with each entry given by

$$s_l = \begin{cases} 1, & l \in \mathbb{I} \\ e^{j\pi/2}, & l \in \mathbb{R} \end{cases}, \quad l = 1, 2, \ldots, L. \tag{4.5}$$

In (4.4), the vector $\mathbf{s}$ describes the RIS partition. In particular, given L and p, there are in total $\binom{L}{p}$ different RIS partition realizations. As such, a total of $b = \lfloor \log_2 \binom{L}{p} \rfloor$ information bits can be conveyed by the RIS partition, and therefore $\mathbb{I}$ has $B = 2^b$ possible realizations. A look-up table of size B providing a one-to-one mapping between the b bits and the p indices can be stored at the controller of the RIS and at the user. Table 4.1 provides an example of a look-up table with $L = 4$, $p = 2$, and $b = 2$.

Under the RIS-QRM scheme, the baseband received signal at the user is given by

$$y = \boldsymbol{\phi}^H \mathbf{h} x + n = \mathbf{s}^H \mathsf{diag}(\boldsymbol{\theta}^*)^H \mathbf{h} x + n. \tag{4.6}$$

Table 4.1: A Look-up Table Example for $L = 4$, $p = 2$, and $b = 2$.

Bits	Indices $\mathbb{I}$ / $\mathbb{R}$	Partition vector $\mathbf{s}$
00	$\{1,2\}$ / $\{3,4\}$	$\left[1, 1, e^{j\pi/2}, e^{j\pi/2}\right]^H$
01	$\{2,3\}$ / $\{1,4\}$	$\left[e^{j\pi/2}, 1, 1, e^{j\pi/2}\right]^H$
10	$\{1,3\}$ / $\{2,4\}$	$\left[1, e^{j\pi/2}, 1, e^{j\pi/2}\right]^H$
11	$\{1,4\}$ / $\{2,3\}$	$\left[1, e^{j\pi/2}, e^{j\pi/2}, 1\right]^H$

By letting $\mathbf{v} \triangleq [v_1, v_2, \ldots, v_L]^T = \text{diag}(\boldsymbol{\theta}^*)^H \underline{\mathbf{h}}$, we have

$$y = \mathbf{s}^H \mathbf{v} x + n = \left(\sum_{l \in \mathbb{I}} v_l + j \sum_{l \in \mathbb{R}} v_l \right) x + n. \tag{4.7}$$

It is worth pointing out that although the proposed RIS-QRM scheme has the same spectral efficiency as the RIS-RPM scheme (i.e., $b + \log_2 M$), the former improves the received power by reactivating the inactive elements of the latter, thus increasing the achievable rate in the low SNR regime. We assume that the CSI of $\mathbf{v}$ is perfectly known at the user, which can be obtained by downlink pilot training. By considering a joint search for all possible RIS partition realizations and signal constellation points, the ML detector of the RIS-QRM system is given by

$$(\hat{x}, \hat{\mathbf{s}}) = \arg \min_{x \in \mathbb{M}, \, \mathbf{s} \in \mathbb{S}} \left| y - \mathbf{s}^H \mathbf{v} x \right|^2 \tag{4.8}$$

where $\mathbb{M}$ and $\mathbb{S}$ are sets of all possible realizations of x and $\mathbf{s}$, respectively. It can be observed that the search complexity of the ML detector in (4.8) is $\sim \mathcal{O}(MB)$.

4.3 Analysis of BEP

In this section, we characterize the BEP performance of the proposed RIS-QRM scheme in Rician fading assuming the ML detection. A Rician fading channel can be expressed as the superposition of a deterministic line-of-sight (LoS) component and a non-LoS (NLoS) component (characterized by a complex Gaussian random variable). Thus, the AP-RIS channel $\mathbf{g}$ and the RIS-user channel $\mathbf{h}_r$ are given by

$$\mathbf{g} = \sqrt{\frac{\kappa_1}{\kappa_1 + 1}} \mathbf{g}^{\text{LoS}} + \sqrt{\frac{1}{\kappa_1 + 1}} \mathbf{g}^{\text{NLoS}} \tag{4.9}$$

$$\mathbf{h}_r = \sqrt{\frac{\kappa_2}{\kappa_2 + 1}} \mathbf{h}_r^{\text{LoS}} + \sqrt{\frac{1}{\kappa_2 + 1}} \mathbf{h}_r^{\text{NLoS}} \tag{4.10}$$

where κ_1 and κ_2 are the Rician factors of the AP-RIS and RIS-user channels, respectively, $\mathbf{g}^{\text{LoS}}$ and $\mathbf{g}^{\text{NLoS}} \sim \mathcal{CN}(\mathbf{0}_{L_0 \times 1}, 2\sigma_1^2 \mathbf{I}_{L_0})$ denote the deterministic LoS and non-LoS (Rayleigh fading) components of the AP-RIS channel, respectively, $\mathbf{h}_r^{\text{LoS}}$ and $\mathbf{h}_r^{\text{NLoS}} \sim \mathcal{CN}(\mathbf{0}_{L_0 \times 1}, 2\sigma_2^2 \mathbf{I}_{L_0})$ denote the deterministic LoS and Rayleigh fading components of the

RIS-user channel, respectively. For ease of exposition, we assume $\mathbf{g}^{\mathsf{LoS}} = \mu_1 \mathbf{1}_{L_0 \times 1}$ and $\mathbf{h}_r^{\mathsf{LoS}} = \mu_2 \mathbf{1}_{L_0 \times 1}$, where $\mu_1, \mu_2 \in \mathbb{C}$ are of channel power $|\mu_1|^2 = \bar{\nu}_1^2$ and $|\mu_2|^2 = \bar{\nu}_2^2$, respectively. Note that $\kappa_1 = \bar{\nu}_1^2/2\sigma_1^2$ and $\kappa_2 = \bar{\nu}_2^2/2\sigma_2^2$, where $2\sigma_1^2$ and $2\sigma_2^2$ are the average powers in non-LoS components $\mathbf{g}^{\mathsf{NLoS}}$ and $\mathbf{h}_r^{\mathsf{NLoS}}$. Although the following analysis is based on the assumption that the AP-user direct link is blocked, it can be applied to the case with a direct link, by assuming that the direct and cascaded channels are independent random variables.

Without loss of generality, we assume that x is transmitted by the AP and that $\mathbf{s}$ is selected at the RIS. Based on (4.8), the conditional PEP of detecting $(x, \mathbf{s})$ in lieu of $(\hat{x}, \hat{\mathbf{s}})$ $(x, \hat{x} \in \mathbb{M}$ and $\mathbf{s}, \hat{\mathbf{s}} \in \mathbb{S})$ is given by

$$P\{x, \mathbf{s} \rightarrow \hat{x}, \hat{\mathbf{s}}|\mathbf{v}\} = P\left\{ \left| y - \hat{\mathbf{s}}^H \mathbf{v}\hat{x} \right|^2 < \left| y - \mathbf{s}^H \mathbf{v}x \right|^2 \right\}$$
$$= P\left\{ |\varpi + n|^2 < |n|^2 \right\} \tag{4.11}$$

where $\varpi = \mathbf{s}^H \mathbf{v}x - \hat{\mathbf{s}}^H \mathbf{v}\hat{x}$ denotes the Euclidean distance between $(x, \mathbf{s})$ and $(\hat{x}, \hat{\mathbf{s}})$. From the above, we see that the probability of detecting $(x, \mathbf{s})$ in lieu of $(\hat{x}, \hat{\mathbf{s}})$ is equal to the probability that n is closer to point ϖ than to the origin point. Let $\bar{n} = \langle n, \varpi \rangle / |\varpi|$ denote the projection of n onto the direction of ϖ, which follows a Gaussian distribution with zero mean and variance $N_0/2$. Then, we have

$$P\{x, \mathbf{s} \rightarrow \hat{x}, \hat{\mathbf{s}}|\mathbf{v}\} = P\left\{ \bar{n} > \frac{|\varpi|}{2} \right\} = Q\left(\sqrt{\frac{|\varpi|^2}{2N_0}} \right). \tag{4.12}$$

In order to obtain the unconditional PEP, the conditional PEP expression given in (4.12) needs to be averaged with respect to $\mathbf{v}$, i.e., $P\{x, \mathbf{s} \rightarrow \hat{x}, \hat{\mathbf{s}}\} = \mathbb{E}_{\mathbf{v}} \{P\{x, \mathbf{s} \rightarrow \hat{x}, \hat{\mathbf{s}}|\mathbf{v}\}\}$. Defining $Z = |\varpi|^2$ and considering the alternative form of the Q-function, we obtain the following unconditional PEP expression:

$$P\{x, \mathbf{s} \rightarrow \hat{x}, \hat{\mathbf{s}}\} = \int_0^\infty \frac{1}{\pi} \int_0^{\frac{\pi}{2}} \exp\left(\frac{-z}{4N_0 \sin^2 \varphi} \right) f_Z(z) d\varphi dz$$
$$= \frac{1}{\pi} \int_0^{\frac{\pi}{2}} \mathcal{M}_Z \left(-\frac{1}{4N_0 \sin^2 \varphi} \right) d\varphi \tag{4.13}$$

where $f_Z(\cdot)$ and $\mathcal{M}_Z(\cdot)$ are the probability density function (PDF) and moment generating function (MGF) of random variable Z, respectively.

Proposition 3. *Z can be approximated with a Gamma distribution with PDF*

$$f_Z(z) = z^{\bar{k}-1} \frac{e^{-z/\bar{\theta}}}{\bar{\theta}^{\bar{k}} \Gamma(\bar{k})}, \quad z \geq 0 \tag{4.14}$$

where $\bar{k}$ and $\bar{\theta}$ are given in (4.31) in the following proof.

Proof. Let ϖ_r and ϖ_i denote real and imaginary parts of ϖ, which are given by

$$\varpi_r = x_r \sum_{l \in \mathbb{I}} v_l - x_i \sum_{l \in \mathbb{L} \backslash \mathbb{I}} v_l - \hat{x}_r \sum_{l \in \hat{\mathbb{I}}} v_l + \hat{x}_i \sum_{l \in \mathbb{L} \backslash \hat{\mathbb{I}}} v_l \tag{4.15}$$

$$\varpi_i = x_i \sum_{l \in \mathbb{I}} v_l + x_r \sum_{l \in \mathbb{L} \backslash \mathbb{I}} v_l - \hat{x}_i \sum_{l \in \hat{\mathbb{I}}} v_l - \hat{x}_r \sum_{l \in \mathbb{L} \backslash \hat{\mathbb{I}}} v_l \tag{4.16}$$

where x_r and x_i denote real and imaginary parts of x, respectively, and $\hat{x}_r$ and $\hat{x}_i$ denote real and imaginary parts of $\hat{x}$, respectively. We define $\mathbb{A}_1 = \mathbb{I} \cap \hat{\mathbb{I}}$, $\mathbb{A}_2 = \mathbb{I} \setminus \mathbb{A}_1$, $\mathbb{A}_3 = \hat{\mathbb{I}} \setminus \mathbb{A}_1$, and $\mathbb{A}_4 = \mathbb{L} \setminus \{\mathbb{A}_1 \cup \mathbb{A}_2 \cup \mathbb{A}_3\}$. Note that $\mathbb{A}_1$, $\mathbb{A}_2$, $\mathbb{A}_3$, and $\mathbb{A}_4$ are disjoint to each other. Then, (4.15) and (4.16) are rewritten as

$$
\begin{aligned}
\varpi_r = & (x_r - \hat{x}_r) \sum_{l \in \mathbb{A}_1} v_l + (x_r + \hat{x}_i) \sum_{l \in \mathbb{A}_2} v_l \\
& + (-x_i - \hat{x}_r) \sum_{l \in \mathbb{A}_3} v_l + (-x_i + \hat{x}_i) \sum_{l \in \mathbb{A}_4} v_l
\end{aligned}
\tag{4.17}
$$

$$
\begin{aligned}
\varpi_i = & (x_i - \hat{x}_i) \sum_{l \in \mathbb{A}_1} v_l + (x_i - \hat{x}_r) \sum_{l \in \mathbb{A}_2} v_l \\
& + (x_r - \hat{x}_i) \sum_{l \in \mathbb{A}_3} v_l + (x_r - \hat{x}_r) \sum_{l \in \mathbb{A}_4} v_l.
\end{aligned}
\tag{4.18}
$$

It is clear that the real and imaginary parts of ϖ are different linear combinations of L i.i.d. random variables $v_1, \ldots, v_L$. Therefore, (4.17) and (4.18) are re-expressed in a generic form as

$$
\varpi_r = \bar{\mathbf{a}}^T \mathbf{v}, \quad \varpi_i = \bar{\mathbf{b}}^T \mathbf{v}
\tag{4.19}
$$

where $\bar{\mathbf{a}}^T = [\bar{a}_1, \bar{a}_2, \ldots, \bar{a}_L]$ and $\bar{\mathbf{b}}^T = [\bar{b}_1, \bar{b}_2, \ldots, \bar{b}_L]$ are the corresponding coefficient vectors.

Define $\underline{\mathbf{g}}_l \triangleq \left[\underline{g}_{l,1}, \underline{g}_{l,2}, \ldots, \underline{g}_{l,L_1} \right]^H$ and $\underline{\mathbf{h}}_{r,l}^H \triangleq \left[\underline{h}_{r,l,1}, \underline{h}_{r,l,2}, \ldots, \underline{h}_{r,l,L_1} \right]$. Under the assumption of Rician fading, elements in $\underline{\mathbf{g}}_l^H$ and $\underline{\mathbf{h}}_{r,l}$ can be described by

$$
\underline{g}_{l,m} = \mu_1 + \mathcal{CN}(0, 2\sigma_1^2)
\tag{4.20}
$$

$$
\underline{h}_{r,l,m} = \mu_2 + \mathcal{CN}(0, 2\sigma_2^2)
\tag{4.21}
$$

with $m = 1, 2, \ldots, L_1$. Defining $Y_m = \underline{h}_{r,l,m} \underline{g}_{l,m}^\dagger$, then we have $\mathbb{E}\{Y_m\} = \mu_1^\dagger \mu_2 \triangleq \mu$ and $\mathrm{Var}\{\mathrm{Re}(Y_m)\} = \mathrm{Var}\{\mathrm{Im}(Y_m)\} = \bar{\nu}_1^2 \sigma_2^2 + \sigma_1^2(\bar{\nu}_2^2 + 2\sigma_2^2)$. As $L_1 \gg 1$, we resort to the central limit theorem to approximate the distribution of v_l. Specifically, the real and imaginary parts of $\underline{h}_l = \sum_{m=1}^{L_1} Y_m$ are two independent Gaussian random variables distributed according to $\mathcal{N}(L_1 \mathrm{Re}(\mu), \sigma^2)$ and $\mathcal{N}(L_1 \mathrm{Im}(\mu), \sigma^2)$ with $\sigma^2 \triangleq L_1 \bar{\nu}_1^2 \sigma_2^2 + L_1 \sigma_1^2(\bar{\nu}_2^2 + 2\sigma_2^2)$. Then, $v_l = |\underline{h}_l|$ is a Rician random variable with PDF

$$
f_{v_l}(z) = \frac{z}{\sigma^2} \exp\left(-\frac{z^2 + \bar{\nu}^2}{2\sigma^2} \right) I_0\left(\frac{z\bar{\nu}}{\sigma^2} \right), z \geq 0
\tag{4.22}
$$

where $\bar{\nu} = L_1 \bar{\nu}_1 \bar{\nu}_2$ and $I_0(\cdot)$ is the modified Bessel function of the first kind with order zero. The first four moments of v_l are given by

$$
\mathbb{E}\{v_l\} = \sigma \sqrt{\frac{\pi}{2}} {}_1F_1\left(-\frac{1}{2}; 1; -\frac{\bar{\nu}^2}{2\sigma^2} \right)
\tag{4.23}
$$

$$
\mathbb{E}\{v_l^2\} = 2\sigma^2 + \bar{\nu}^2
\tag{4.24}
$$

$$
\mathbb{E}\{v_l^3\} = 3\sigma^3 \sqrt{\frac{\pi}{2}} {}_1F_1\left(-\frac{3}{2}; 1; -\frac{\bar{\nu}^2}{2\sigma^2} \right)
\tag{4.25}
$$

$$
\mathbb{E}\{v_l^4\} = 8\sigma^4 + 8\sigma^2\bar{\nu}^2 + \bar{\nu}^4
\tag{4.26}
$$

where $_1F_1(a; q; p)$ is the confluent hypergeometric function of the first kind.

Based on (4.19), Z can be expressed as

$$Z = \bar{\mathbf{a}}^T \mathbf{v}\mathbf{v}^T \bar{\mathbf{a}} + \bar{\mathbf{b}}^T \mathbf{v}\mathbf{v}^T \bar{\mathbf{b}} = \mathrm{tr}(\bar{\mathbf{a}}\bar{\mathbf{a}}^T \mathbf{v}\mathbf{v}^T + \bar{\mathbf{b}}\bar{\mathbf{b}}^T \mathbf{v}\mathbf{v}^T)$$
$$= \mathrm{tr}(\mathbf{C}\mathbf{V}) = \mathbf{v}^T \mathbf{C}\mathbf{v} \tag{4.27}$$

where $\mathbf{C} = \bar{\mathbf{a}}\bar{\mathbf{a}}^T + \bar{\mathbf{b}}\bar{\mathbf{b}}^T$ and $\mathbf{V} = \mathbf{v}\mathbf{v}^T$. The first moment of Z is given by

$$\mathbb{E}\{Z\} = \mathrm{tr}(\mathbf{C} \cdot \mathbb{E}\{\mathbf{V}\}) \tag{4.28}$$

and its second moment is given by (4.30), shown at the top of the next page,

$$\mathbb{E}\{Z^2\} = \mathbb{E}\left\{\left(\sum_{\iota=1}^{L}\sum_{\jmath=1}^{L} v_\iota \mathbf{C}_{\iota\jmath} v_\jmath\right)^2\right\} = \sum_{\iota=1}^{L}\sum_{\jmath=1}^{L}\sum_{\iota'=1}^{L}\sum_{\jmath'=1}^{L} \mathbf{C}_{\iota\jmath}\mathbf{C}_{\iota'\jmath'}\mathbb{E}\{v_\iota v_\jmath v_{\iota'} v_{\jmath'}\}$$
$$= \sum_{\iota\in\mathbb{L}}\mathbf{C}_{\iota\iota}^2\mathbb{E}\{v_\iota^4\} + \sum_{\iota\in\mathbb{L}}\sum_{\jmath\in\mathbb{L}\setminus\{\iota\}} 4\mathbf{C}_{\iota\iota}\mathbf{C}_{\iota\jmath}\mathbb{E}\{v_\iota^3\}\mathbb{E}\{v_\jmath\}$$
$$+ \sum_{\iota\in\mathbb{L}}\sum_{\iota'\in\mathbb{L}\setminus\{\iota\}}\sum_{\jmath\in\mathbb{L}\setminus\{\iota,\iota'\}} (2\mathbf{C}_{\iota\iota}\mathbf{C}_{\iota'\jmath} + 4\mathbf{C}_{\iota\iota'}\mathbf{C}_{\iota\jmath})\mathbb{E}\{v_\iota^2\}\mathbb{E}\{v_{\iota'}\}\mathbb{E}\{v_\jmath\} \tag{4.29}$$
$$+ \sum_{\iota\in\mathbb{L}}\sum_{\jmath\in\mathbb{I}\setminus\{\iota\}} (\mathbf{C}_{\iota\iota}\mathbf{C}_{\jmath\jmath} + 2\mathbf{C}_{\iota\jmath}^2)\mathbb{E}\{v_\iota^2\}\mathbb{E}\{v_\jmath^2\}$$
$$+ \sum_{\iota\in\mathbb{L}}\sum_{\iota'\in\mathbb{L}\setminus\{\iota\}}\sum_{\jmath\in\mathbb{L}\setminus\{\iota,\iota'\}}\sum_{\jmath'\in\mathbb{L}\setminus\{\iota,\iota',\jmath\}} \mathbf{C}_{\iota\iota'}\mathbf{C}_{\jmath\jmath'}\mathbb{E}\{v_\iota\}\mathbb{E}\{v_{\iota'}\}\mathbb{E}\{v_\jmath\}\mathbb{E}\{v_{\jmath'}\}$$

where $\mathbf{C}_{\iota\jmath}$ denotes the entry on the ι-th row and $\jmath$-th column of $\mathbf{C}$. Define $\bar{\mathbf{V}} = \mathbb{E}\{\mathbf{V}\}$. Based on (4.23) and (4.24), the elements of $\bar{\mathbf{V}}$ are given by

$$\bar{\mathbf{V}}_{\iota\jmath} = \begin{cases} 2\sigma^2 + \bar{\nu}^2, & \iota = \jmath \\ \sigma^2\dfrac{\pi}{2}\,_1F_1^2\left(-\dfrac{1}{2}; 1; -\dfrac{\bar{\nu}^2}{2\sigma^2}\right), & \iota \neq \jmath \end{cases} \tag{4.30}$$

where $_1F_1^2(\cdot; \cdot; \cdot)$ denotes the square of the confluent hypergeometric function $_1F_1(\cdot; \cdot; \cdot)$, and $\iota, \jmath = 1, 2, \ldots, L$. According to the Gamma second order moment matching approximation method [89], a Gamma distribution with the same first and second moments has parameters

$$\bar{k} = \frac{\mathbb{E}\{Z\}^2}{\mathbb{E}\{Z^2\} - \mathbb{E}\{Z\}^2}, \quad \bar{\theta} = \frac{\mathbb{E}\{Z^2\} - \mathbb{E}\{Z\}^2}{\mathbb{E}\{Z\}}. \tag{4.31}$$

$\square$

Based on the PDF of Z in (4.14), the MGF of Z can be written as

$$\mathcal{M}_Z(\tau) = \frac{1}{(\bar{\theta})^{\bar{k}}\Gamma(\bar{k})}\int_0^\infty e^{(\tau-1/\bar{\theta})z} z^{\bar{k}-1}dz = \left(1 - \tau\bar{\theta}\right)^{-\bar{k}} \tag{4.32}$$

where the closed-form expression is obtained by using formula [90, eq. (3.326.2)] to the integral. By substituting (4.32) into (4.13), the unconditional PEP can be derived in closed-form as follows:

$$
\begin{aligned}
P\{x, \mathbf{s} \rightarrow \hat{x}, \hat{\mathbf{s}}\} &= \frac{1}{\pi} \int_0^{\frac{\pi}{2}} \left(1 + \frac{\bar{\theta}}{2N_0\left(1 - \cos 2\varphi\right)}\right)^{-\bar{k}} d\varphi \\
&\overset{(a)}{=} \frac{1}{2\pi} \int_0^{\pi} \left(1 + \frac{\bar{\theta}}{2N_0\left(1 - \cos \psi\right)}\right)^{-\bar{k}} d\psi \\
&\overset{(b)}{=} \frac{1}{2\pi} \int_{\frac{1}{2}}^{\infty} \left(1 + \frac{\bar{\theta}}{2N_0}t\right)^{-\bar{k}} t^{-1}(2t - 1)^{-\frac{1}{2}} dt \\
&\overset{(c)}{=} \frac{1}{2\pi} \left(\frac{4N_0}{\bar{\theta}}\right)^{\bar{k}} \left(\frac{4N_0}{\bar{\theta}} + 1\right)^{\frac{1}{2} - \bar{k}} \\
&\quad \times \mathsf{B}\left(\frac{1}{2} + \bar{k}, \frac{1}{2}\right) {}_2F_1\left(1, \frac{1}{2}; \frac{1}{2}; -\frac{4N_0}{\bar{\theta}}\right) \\
&= \frac{1}{4}\left(\frac{\bar{\theta}}{\pi N_0}\right)^{\frac{1}{2}} \left(1 + \frac{\bar{\theta}}{4N_0}\right)^{-\frac{1}{2} - \bar{k}} \frac{\Gamma\left(\frac{1}{2} + \bar{k}\right)}{\Gamma\left(1 + \bar{k}\right)}
\end{aligned} \tag{4.33}
$$

where (a) holds by applying substitution $\psi = 2\varphi$, (b) holds by applying substitution $t = 1/(1 - \cos \psi)$, (c) holds by using formula [90, eq. (3.197.2)] to the integral, $\mathsf{B}(\cdot, \cdot)$ is the beta function [90, eq. (8.384.1)], and ${}_2F_1(-\varrho_1, \varrho_2; \varrho_2; -\varrho_3)$ is the Gauss hypergeometric function [90, eq. (9.121.1)]. Finally, the upper bounds on the BEPs for detecting x and s are given by

$$
P_b^{\mathsf{AP}} \leq \frac{1}{MB} \sum_x \sum_{\hat{x}} \sum_{\mathbf{s}} \sum_{\hat{\mathbf{s}}} P\{x, \mathbf{s} \rightarrow \hat{x}, \hat{\mathbf{s}}\} \frac{e(x, \hat{x})}{\log_2 M} \tag{4.34}
$$

$$
P_b^{\mathsf{RIS}} \leq \frac{1}{MB} \sum_{\mathbf{s}} \sum_{\hat{\mathbf{s}}} \sum_x \sum_{\hat{x}} P\{x, \mathbf{s} \rightarrow \hat{x}, \hat{\mathbf{s}}\} \frac{e(\mathbf{s}, \hat{\mathbf{s}})}{b} \tag{4.35}
$$

where $e(\cdot, \cdot)$ denotes the number of bits in error for the corresponding pairwise error event. The BEP that combines x and s is upper bounded by

$$
P_b \leq \sum_x \sum_{\mathbf{s}} \sum_{\hat{x}} \sum_{\hat{\mathbf{s}}} \frac{P\{x, \mathbf{s} \rightarrow \hat{x}, \hat{\mathbf{s}}\} e(x, \mathbf{s} \rightarrow \hat{x}, \hat{\mathbf{s}})}{MB(b + \log_2 M)}. \tag{4.36}
$$

In particular, let us compare the Euclidean distances of the proposed RIS-QRM scheme and the RIS-RPM counterpart. When the SNR approaches infinity, the probability of confusing x with $\hat{x}$ approaches zero so that it is valid to assume $x = \hat{x}$. Conditioned on the correct detection of the constellation symbol, the Euclidean distance between s and ŝ of RIS-QRM is given by

$$
D_{\mathsf{QRM}}^{\mathbf{s}} = \sqrt{2}|x| \left| \sum_{l \in \mathbb{A}_2} v_l - \sum_{l \in \mathbb{A}_3} v_l \right| \tag{4.37}
$$

and that of the RIS-RPM counterpart is given by

$$D_{\mathsf{RPM}}^{\mathbf{s}} = |x| \left| \sum_{l \in \mathbb{A}_2} v_l - \sum_{l \in \mathbb{A}_3} v_l \right|. \tag{4.38}$$

Therefore, the Euclidean distance of the RIS-QRM scheme is $\sqrt{2}$ times that of the RIS-RPM counterpart due to the quadrature component, as it will be confirmed by the simulations.

4.4 Extension to MISO Setting and Low-Complexity Detector

4.4.1 Extension to MISO Setting

In this section, we consider a more general case, where the AP is equipped with N transmit antennas and there exists a direct channel between the AP and the user. Let $\mathbf{h}_d^H \in \mathbb{C}^{1 \times N}$ and $\mathbf{G} \in \mathbb{C}^{L_0 \times N} \triangleq \left[\underline{\mathbf{G}}_1^H, \underline{\mathbf{G}}_2^H, \ldots, \underline{\mathbf{G}}_L^H \right]^H$ denote the baseband channels of the AP-user direct link and the AP-RIS link, respectively, where $\underline{\mathbf{G}}_l \in \mathbb{C}^{L_1 \times N}$ are element-wise channel coefficients associated with the l-th group for $l = 1, 2, \ldots, L$. We denote the AP-RIS-user cascaded channel matrix as $\underline{\mathbf{H}} = \left[\bar{\mathbf{h}}_1, \bar{\mathbf{h}}_2, \ldots, \bar{\mathbf{h}}_L \right]^H$, where $\bar{\mathbf{h}}_l^H = \underline{\mathbf{h}}_{r,l}^H \underline{\mathbf{G}}_l$ with $l = 1, 2, \ldots, L$ denotes the effective channel from the AP to user via group l with the common phase shift equal to zero. We assume that the knowledge of $\mathbf{h}_d^H$ and $\underline{\mathbf{H}}$ is perfectly known at the AP.

Due to the presence of the AP-user direct channel as well as multiple antennas at the AP, two questions need to be answered:

- *Q1*: How to design the active beamforming of the AP to cooperate with the phase shifts at the RIS?

- *Q2*: How to design the reflection coefficients at the RIS to implicitly convey additional information?

To answer the first question, we consider the RIS reflection coefficients without information transfer (i.e., $\boldsymbol{\theta}$), and get the baseband received signal at the user, which is given by

$$\underline{y} = \left(\boldsymbol{\theta}^H \underline{\mathbf{H}} + \mathbf{h}_d^H \right) \mathbf{w} x + n \tag{4.39}$$

where $\mathbf{w} \in \mathbb{C}^{N \times 1}$ denotes the active beamforming vector of the AP with $\|\mathbf{w}\| = 1$. Hence, the received SNR is given by

$$\mathsf{SNR} = \frac{P_t}{N_0} \left| \left(\boldsymbol{\theta}^H \underline{\mathbf{H}} + \mathbf{h}_d^H \right) \mathbf{w} \right|^2. \tag{4.40}$$

In order to fully reap the three-dimensional beamforming gain, the active beamforming vector of the AP usually needs to be jointly optimized with the phase shifts of the RIS based on the knowledge of $\mathbf{h}_d^H$ and $\underline{\mathbf{H}}$. In this chapter, our objective is to maximize the

received SNR by jointly optimizing the active beamforming at the AP and the phase shifts of the RIS, subject to the unit-modulus reflection constraint on each RIS element. Note that this is effective in improving the error performance of the bits carried on x as well as $\mathbf{s}$, as will be shown in the simulations. The optimization problem is formulated as (irrelevant terms are omitted for simplicity)

$$\text{(P1):} \quad \max_{\mathbf{w},\boldsymbol{\theta}} \quad \left|\left(\boldsymbol{\theta}^H \underline{\mathbf{H}} + \mathbf{h}_d^H\right)\mathbf{w}\right|^2 \tag{4.41}$$

$$\text{s.t.} \quad \|\mathbf{w}\|^2 = 1 \tag{4.42}$$

$$0 \leqslant \theta_l < 2\pi, \; l = 1, 2, \ldots, L. \tag{4.43}$$

It is easily verified that this problem is always feasible, since there exists at least one solution that satisfies the constraints in (4.42) and (4.43), e.g., $\boldsymbol{\theta} = \mathbf{1}_{L\times 1}$ and $\mathbf{w} = \left(\mathbf{1}_{1\times L}\underline{\mathbf{H}} + \mathbf{h}_d^H\right)^H \big/ \left\|\mathbf{1}_{1\times L}\underline{\mathbf{H}} + \mathbf{h}_d^H\right\|$. However, it is non-convex since the objective function of (4.41) is not jointly concave with respect to $\mathbf{w}$ and $\boldsymbol{\theta}$, and the constraints in (4.42) and (4.43) are both non-convex. Problem (P1) can be sub-optimally solved by leveraging the alternating optimization (AO) technique. An AO algorithm has been provided in [91], which iteratively optimizes the phase shifts of the RIS and the active beamforming of the AP in an alternating manner until the convergence is achieved. Specifically, for a given phase-shift vector $\boldsymbol{\theta}$, the maximum-ratio transmission (MRT) solution is optimal, i.e.,

$$\mathbf{w}^* = \frac{\left(\boldsymbol{\theta}^H \underline{\mathbf{H}} + \mathbf{h}_d^H\right)^H}{\left\|\boldsymbol{\theta}^H \underline{\mathbf{H}} + \mathbf{h}_d^H\right\|}. \tag{4.44}$$

For a given active beamforming vector $\mathbf{w}^*$ in (4.44), problem (P1) is reduced to

$$\text{(P2):} \quad \max_{\boldsymbol{\theta}} \quad \left|\left(\boldsymbol{\theta}^H \underline{\mathbf{H}} + \mathbf{h}_d^H\right)\mathbf{w}^*\right|^2 \tag{4.45}$$

$$\text{s.t.} \quad 0 \leqslant \theta_l < 2\pi, \; l = 1, 2, \ldots, L. \tag{4.46}$$

The optimal solution to problem (P2) is given by $\underline{\boldsymbol{\theta}}^* = \left[e^{j\underline{\theta}_1^*}, \ldots, e^{j\underline{\theta}_L^*}\right]^H$, where the l-th phase shift is given by

$$\underline{\theta}_l^* = \angle(\mathbf{h}_d^H \mathbf{w}^*) - \angle(\bar{\mathbf{h}}_l^H \mathbf{w}^*), \quad l = 1, 2, \ldots, L. \tag{4.47}$$

The vectors $\mathbf{w}$ and $\boldsymbol{\theta}$ are iteratively optimized in an alternating manner, until the fractional increase of (4.41) is less than a sufficiently small value $\epsilon > 0$ or the maximum number of iterations is reached. By a slight abuse of notation, we denote the outputs of the AO algorithm as $\mathbf{w}^*$ and $\underline{\boldsymbol{\theta}}^*$. After obtaining $\mathbf{w}^*$ and $\underline{\boldsymbol{\theta}}^*$, the AP reports the optimal phase shifts (i.e., $\underline{\boldsymbol{\theta}}^*$) to the RIS controller via the feedback link.

We note that the preferred phase shifts in (4.47) co-phase the signals from the AP to the user via the RIS with the signal over the AP-user direct link in order to achieve coherent signal combination. To implicitly convey additional information, we extend the principle of the RIS-QRM described in Section 4.2.2, by arranging the signals reflected by the RIS to be aligned either in phase or in quadrature with the signal over the direct link. Specifically, the phase shifts of the RIS-QRM scheme are given by

$$\underline{\phi}_l = \begin{cases} \underline{\theta}_l^*, & l \in \mathbb{I} \\ \underline{\theta}_l^* + \pi/2, & l \in \mathbb{R} \end{cases}, \quad l = 1, 2, \ldots, L \tag{4.48}$$

where $\underline{\theta}_l^*$ is given in (4.47), so that, at the user, the signals reflected by the p selected groups are aligned in phase with the signal over the direct link and the signals reflected by the $L - p$ remaining groups are orthogonal to the signal over the direct link. Based on the RIS phase shifts in (4.48), the baseband received signal at the user can be expressed as

$$y = \left(\mathbf{s}^H \mathsf{diag}(\underline{\boldsymbol{\theta}}^*)^H \underline{\mathbf{H}} \mathbf{w}^* + \mathbf{h}_d^H \mathbf{w}^* \right) x + n \tag{4.49}$$

$$= \tilde{\mathbf{s}}^H \tilde{\mathbf{v}} x + n \tag{4.50}$$

where $\tilde{\mathbf{s}}^H = \begin{bmatrix} 1 & \mathbf{s}^H \end{bmatrix}$ and $\tilde{\mathbf{v}} \triangleq [\tilde{v}_0, \tilde{v}_1, \ldots, \tilde{v}_L]^T = \begin{bmatrix} \mathbf{h}_d^H \mathbf{w}^* \\ \mathsf{diag}(\underline{\boldsymbol{\theta}}^*)^H \underline{\mathbf{H}} \mathbf{w}^* \end{bmatrix}$. Letting $\phi_0 \triangleq \angle(\mathbf{h}_d^H \mathbf{w}^*)$, we have $\tilde{\mathbf{v}} = \left[e^{j\phi_0} |\tilde{v}_0|, e^{j\phi_0} |\tilde{v}_1|, \ldots, e^{j\phi_0} |\tilde{v}_L| \right]^T$, and therefore

$$y = e^{j\phi_0} \left(\sum_{l \in \{0, \mathbb{I}\}} |\tilde{v}_l| + j \sum_{l \in \mathbb{R}} |\tilde{v}_l| \right) x + n. \tag{4.51}$$

It is worth pointing out that by dividing the RIS into multiple small-sized RISs, each of which is paired with one single receive antenna, the RIS-QRM scheme can be readily extended to single-input multiple-output (SIMO) and multiple-input multiple-output (MIMO) cases. Specifically, when the user is equipped with N_r receive antennas, the RIS is equally divided into N_r small-sized RISs. Without loss of generality, we assume that the k-th small-sized RIS and the k-th receive antenna are paired, $k = 1, \ldots, N_r$. Then, each small-sized RIS can independently perform the QRM scheme by steering the signal impinging on it towards its paired receive antenna. Furthermore, by considering a time division multiple access (TDMA) method, the proposed scheme can be applied to the downlink of a multi-user MISO system. Specifically, the AP and RIS sequentially communicate with multiple users over consecutive time resource blocks.

4.4.2 Low-Complexity (LC) Detector

In this subsection, we propose a low-complexity detector to separately retrieve x and $\mathbf{s}$ from y. We assume that the effective channel vector $\tilde{\mathbf{v}}$ is perfectly known at the user. Define $\lambda \triangleq \sum_{l=0}^{L} |\tilde{v}_l|$, which is a known constant and is used to estimate the transmitted constellation symbol. To be specific, we define

$$r(\tilde{x}) \triangleq \mathsf{Re} \left(e^{-j\phi_0} y \frac{\langle 1 - j, \tilde{x} \rangle}{|\tilde{x}|} \right), \tilde{x} \in \mathbb{M} \tag{4.52}$$

where $\tilde{x}$ is one of M possible constellation points.

By denoting $\underline{y} = \left(\sum_{l \in \{0, \mathbb{I}\}} |\tilde{v}_l| + j \sum_{l \in \mathbb{R}} |\tilde{v}_l| \right) x$ as the noise-free received signal, we obtain

$$r(\tilde{x}) = \mathsf{Re} \left(\underline{y} \frac{\langle 1 - j, \tilde{x} \rangle}{|\tilde{x}|} \right) + \tilde{n} \tag{4.53}$$

where $\tilde{n} = \text{Re}\left(e^{-j\phi_0}n\langle 1 - j, \tilde{x}\rangle/|\tilde{x}|\right) \sim \mathcal{N}(0, N_0)$. From (4.53), we see that $r(\tilde{x})$ with different $\tilde{x}$ values are independent but non-identically distributed Gaussian random variables. In particular, if $\tilde{x} = x$ (x is the transmitted constellation symbol), the mean of $r(\tilde{x})$ is given by $\lambda|\tilde{x}|$. Otherwise, we have

$$\text{Re}\left(\underline{y}\frac{\langle 1 - j, \tilde{x}\rangle}{|\tilde{x}|}\right) \neq \lambda|\tilde{x}|. \tag{4.54}$$

Therefore, the decision rule for x is to choose the symbol achieving the maximum posterior probability, i.e.,

$$\hat{x}^0 = \arg\max_{\tilde{x}\in\mathbb{M}} \frac{1}{\sqrt{2\pi N_0}} \exp\left(-\frac{\left(r(\tilde{x}) - \lambda|\tilde{x}|\right)^2}{2N_0}\right). \tag{4.55}$$

Equivalently, x is estimated as

$$\hat{x}^0 = \arg\min_{\tilde{x}\in\mathbb{M}} \left|\text{Re}\left(e^{-j\phi_0}y\frac{\langle 1-j,\tilde{x}\rangle}{|\tilde{x}|}\right) - \lambda|\tilde{x}|\right|. \tag{4.56}$$

In this manner, x can be detected without knowing $\mathbf{s}$. To better understand how (4.56) works, hereafter we give an example with binary phase shift keying (BPSK). Consider BPSK, i.e., $\tilde{x} = \pm P_t$. Without loss of generality, we assume that $x = P_t$ is transmitted. Given the observation y, the LC detector calculates (4.52) for all the constellation points, i.e., $r(P_t) = \lambda P_t + \tilde{n}$ and $r(-P_t) = -\lambda P_t - \tilde{n}$, where $\tilde{n} = \text{Re}\left(e^{-j\phi_0}n(1-j)\right) \sim \mathcal{N}(0, N_0)$. Obviously, it is more likely that $|r(P_t)-\lambda P_t| = |\tilde{n}| < |r(-P_t)-\lambda P_t| = |2\lambda P_t+\tilde{n}|$, and when it is true, according to (4.56), we have $\hat{x}^0 = P_t$. By utilizing the estimate of x, the estimate of $\mathbf{s}$ is given by

$$\hat{\mathbf{s}}^0 = \arg\min_{\mathbf{s}\in\mathbb{S}} \left|y - [1 \quad \mathbf{s}^H]\tilde{\mathbf{v}}\hat{x}^0\right|. \tag{4.57}$$

Owing to the disjoint detection of x and $\mathbf{s}$, the search complexity is reduced to $\sim \mathcal{O}(M+B)$.

To further improve the error performance, an iterative algorithm can be applied to update the estimates of x and $\mathbf{s}$ based on the initial decisions (i.e., $\hat{x}^0$ and $\hat{s}^0$). Specifically, x and $\mathbf{s}$ can be estimated alternately in an iterative manner, until the given number of iterations, denoted by I, has been carried out. In iteration r, the estimates of x and $\mathbf{s}$ are given by

$$\hat{x}^r = \arg\min_{x\in\mathbb{M}} \left|y - \left[1 \quad \left(\hat{\mathbf{s}}^{r-1}\right)^H\right]\tilde{\mathbf{v}}x\right| \tag{4.58}$$

$$\hat{\mathbf{s}}^r = \arg\min_{\mathbf{s}\in\mathbb{S}} \left|y - [1 \quad \mathbf{s}^H]\tilde{\mathbf{v}}\hat{x}^r\right| \tag{4.59}$$

where $\hat{\mathbf{s}}^{r-1}$ is the estimate of $\mathbf{s}$ in iteration $(r-1)$, $r = 1, 2, \ldots, I$. Therefore, the search complexity of the LC detector is $\sim \mathcal{O}\left((M + B)(I + 1)\right)$. We will show in the following section that the optimal number of iterations in balancing the error performance and detection complexity is 1.
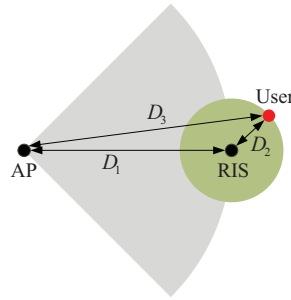
Figure 4.2: The simulation setup (top view), where $D_1 = 50$ m and $D_2 = 5$ m.

4.5 Simulation Results and Discussions

In this section, we present computer simulation results to numerically evaluate the performance of our proposed RIS-QRM scheme in terms of bit error rate (BER) and validate our analytical results in (4.34)-(4.36). The total number of RIS elements is set as $L_0 = 160$. The channel path loss is chosen according to the 3GPP Indoor (InH) model [92, Table B.1.2.1-1]. In particular, the channel path loss γ as a function of the distance D is

$$\gamma = \begin{cases} 42.3 + 16.9 \log_{10} D, & \text{if LoS} \\ 21 + 43.3 \log_{10} D, & \text{if NLoS} \end{cases} \tag{4.60}$$

where D is given in meters (m). We consider the simulation setup in Fig. 4.2. The distance between the AP and RIS is set as $D_1 = 50$ m and the user is placed in the vicinity of the RIS at a distance of 5 m, i.e., $D_2 = 5$ m. As such, the minimum and maximum distances between the AP and the user are given by 45 m and 55 m, i.e., 45 m $\leq D_3 \leq$ 55 m. We set $D_3 = 54$ m in the following simulations. As for the small-scale fading, we consider the Rician fading channel model for all individual links. The Rician factors of the AP-RIS, RIS-user, and AP-user links are denoted by κ_1, κ_2, and κ_3, respectively. We set $\kappa_1 = 1$ in Figs. 4.5-4.14, and show the effect of κ_1 on the error performance in Fig. 4.16. The Rician factor of the RIS-user channel is $\kappa_2 = 1$. We assume a non-LoS (Rayleigh fading) channel between the AP and the user, i.e., $\kappa_3 = 0$. The user's noise power level is $N_0 = -80$ dBm. For the AO algorithm in Section 4.4.1, the maximum number of iterations is 10 and $\epsilon = 10^{-4}$. The simulation results are averaged over 10, 000 independent fading channel realizations.

We first evaluate the effectiveness of the proposed RIS-QRM scheme. Two benchmark schemes are considered for comparison: 1) the RIS-RPM scheme where the elements in vector $\mathbf{s}$ indexed by $\mathbb{R}$ take the value of 0, i.e., $s_l = 0, \forall l \in \mathbb{R}$; 2) the PBIT scheme where the elements in vector $\mathbf{s}$ are independently drawn from the set $\{0, 1\}$ with equiprobability. In Figs. 4.5 and 4.8, we compare the error performance of different schemes versus the SNR, where the parameters are chosen as: $N = 1$, $L = 4$, 64-QAM for the PBIT scheme, and 256-QAM for the RIS-QRM and RIS-RPM schemes with $p = 1$ in Fig. 4.5; $N = 4$, $L = 2$, 128-QAM for the PBIT scheme, and 256-QAM for the RIS-QRM and RIS-RPM schemes with $p = 1$ in Fig. 4.8. Particularly, in Fig. 4.5, the AP-user direct link is blocked. The lower bound (LB) on the BER of x is obtained by assuming that $\mathbf{s}$ is perfectly known when detecting x. The LB on the BER of $\mathbf{s}$ is obtained by assuming that x is perfectly known when detecting $\mathbf{s}$. The BEP upper
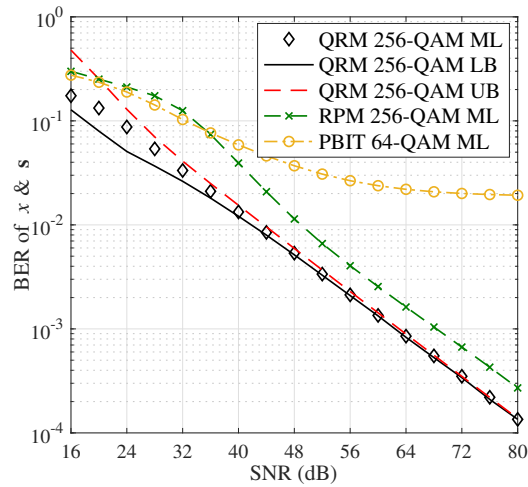
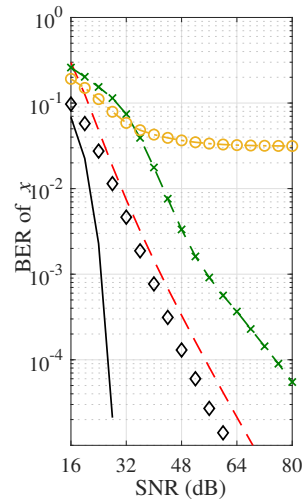Figure 4.3: BER of x and $\mathbf{s}$ versus SNR

Figure 4.4: BER of x versus SNR

bounds (UBs), given in (4.34)-(4.36), are also plotted as benchmarks. In Fig. 4.5, it is observed that the theoretical analysis results are consistent with the simulation results. Moreover, one can observe from Figs. 4.3 and 4.6 that the RIS-QRM scheme shows superior error performance as compared with the RIS-RPM and PBIT schemes. Compared to the RIS-RPM scheme, the proposed RIS-QRM not only improves the error performance of x (shown in Fig. 4.7), but also improves the error performance of $\mathbf{s}$ (shown in Fig. 4.8). This is attributed to the use of the quadrature component of the combined AP-user channel, which enhances the received SNR and increases the minimum Euclidean distance between all the possible realizations of $\mathbf{s}$. As shown in Figs. 4.7 and 4.8, the RIS-QRM scheme greatly outperforms the PBIT scheme as far as the error performance of x is concerned, which is attributed to the full reflection of the RIS at any transmission instance.

In Fig. 4.9, we compare the error performance of the RIS-QRM scheme by using different beamforming designs versus the SNR with $N = 4$, $L = 4$, $p = 3$, and 256-QAM. The following beamforming schemes are considered for comparison: 1) the AP-user
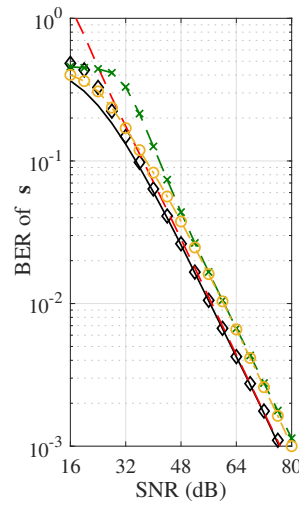
Figure 4.5: Error performance of different schemes versus SNR with $N = 1$, $L = 4$, and $p = 1$ for RIS-QRM and RIS-RPM schemes, where the AP-user direct link is blocked, i.e., $\gamma_3 = 0$.
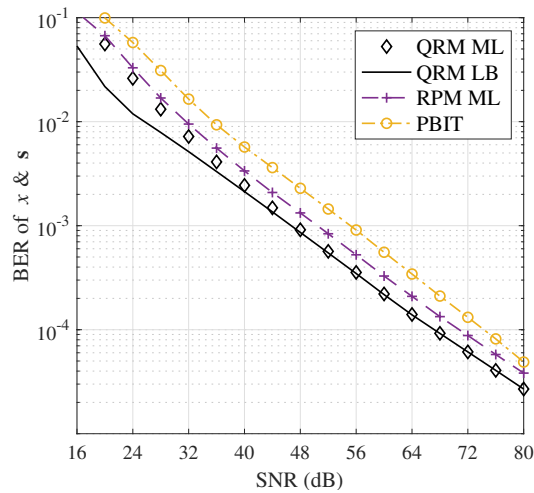

Figure 4.6: BER of x and $\mathbf{s}$ versus SNR

channel MRT scheme where the active beamforming at the AP is steered toward the AP-user direct path, i.e., $\mathbf{w} = \mathbf{h}_d/\|\mathbf{h}_d\|$; 2) the strongest path MRT scheme where the active beamforming at the AP is steered to the path of the largest channel power gain among the cascaded channels of all groups and the AP-user direct channel, i.e., $\mathbf{w} = \mathbf{h}_{\max}/\|\mathbf{h}_{\max}\|$ with $\mathbf{h}_{\max} = \arg\max\{\|\underline{\mathbf{h}}_1\|^2, \ldots, \|\underline{\mathbf{h}}_L\|^2, \|\mathbf{h}_d\|^2\}$; 3) the weakest path MRT scheme where the active beamforming at the AP is steered toward the path of the smallest channel power gain among the cascaded channels of all groups and the AP-user direct channel, i.e., $\mathbf{w} = \mathbf{h}_{\min}/\|\mathbf{h}_{\min}\|$ with $\mathbf{h}_{\min} = \arg\min\{\|\underline{\mathbf{h}}_1\|^2, \ldots, \|\underline{\mathbf{h}}_L\|^2, \|\mathbf{h}_d\|^2\}$. We can observe that the AO algorithm significantly outperforms the other schemes, which can be well understood by the fact that the AO algorithm can dynamically adjust the active beamforming to strike an optimal trade-off between the signal power to the RIS and that directly to the user. On the other hand, it is observed that the performance of the strongest path MRT scheme is better than that of the AP-user channel
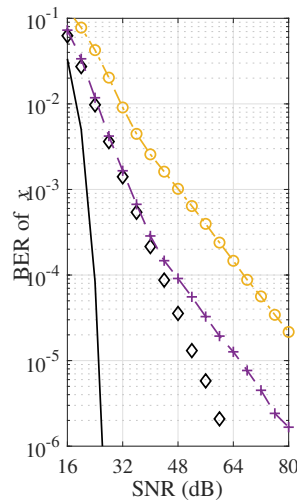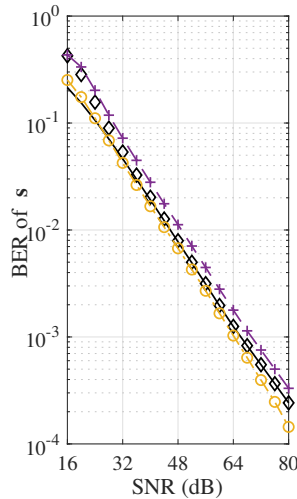
Figure 4.7: BER of x versus SNR

Figure 4.8: BER of $\mathbf{s}$ versus SNR

MRT scheme and the weakest path MRT scheme. This can be understood since the minimum Euclidean distance between all the possible realizations of $\mathbf{s}$ is maximized by maximizing the power of signal over the strongest path. In Fig. 4.12, we examine the performance of the LC detector with $I \in \{0, 1, 2\}$, $N = 4$, $L = 4$, $p = 3$, and 256-QAM. One can observe a significant BER improvement of the LC detector by increasing the value of I from 0 to 1. Compared to the LC detector with $I = 1$, the LC detector with $I = 2$ achieves nearly the same performance but with higher complexity, which implies that the optimal I in balancing the error performance and detection complexity is 1.

We disclose the impact of p on the error performance of the RIS-QRM and RIS-RPM schemes in Fig. 4.14, where p varies from 1 to 3 and other simulation parameters are $N = 4$, $L = 4$, and 256-QAM. For clarity, the lower bounds are omitted. From Fig. 4.13, we observe that the RIS-QRM scheme with $p = 1$ slightly outperforms that with $p = 2$ in terms of the error performance of x, which can be understood by the fact that $\mathbb{E}\left\{(a + b + c)^2 + d^2\right\} \geq \mathbb{E}\left\{(a + b)^2 + (c + d)^2\right\}$, where $a, b, c, d > 0$ are independent and identically distributed random variables. In other words, the more signals are in
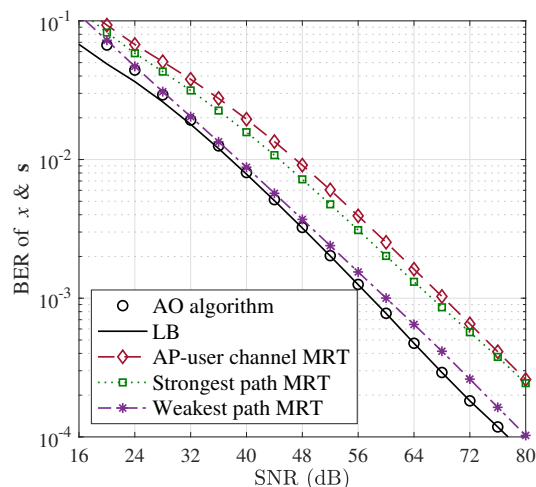
Figure 4.9: Error performance of RIS-QRM using different beamforming designs versus SNR with $N = 4$, $L = 4$, $p = 3$, and 256-QAM.
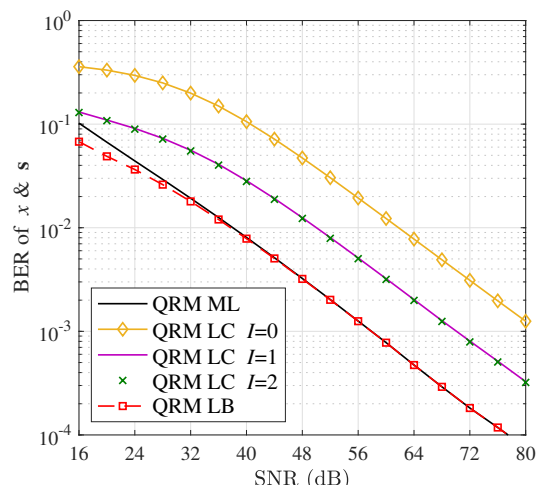

Figure 4.10: BER of x and s versus SNR

phase, the greater the average received power. Moreover, the RIS-QRM scheme with $p = 3$ outperforms that with $p = 1$ due to the AP-user direct channel, which makes the in-phase component of the combined AP-user channel more important. As far as the detection of s is concerned, it is observed from Fig. 4.14 that the error performance of the RIS-QRM scheme is almost insensitive to the value of p. We point out that in the case where there is no direct link between the AP and user, the optimal p can be 1 or $L - 1$, while in the case of a direct AP-user link, the optimal p is more likely to be $L - 1$. From Fig. 4.13, for $p = 3$, we observe a significant performance improvement of the RIS-QRM scheme over the RIS-RPM scheme in the high SNR regime. In the low SNR regime, on the other hand, the error performance of the RIS-QRM scheme is slightly worse than that of the RIS-RPM scheme. This can be understood by the fact that at low SNR, with $L - p = 1$, the received power contributed by the quadrature component of the combined AP-user channel is too small to compensate for the negative impact of the phase variations induced by the quadrature component of the combined channel on symbol detection. We reveal the impact of κ_1 on the error performance of the RIS-
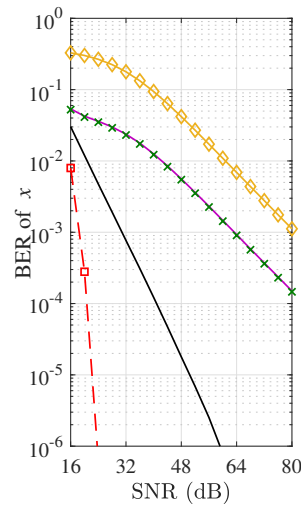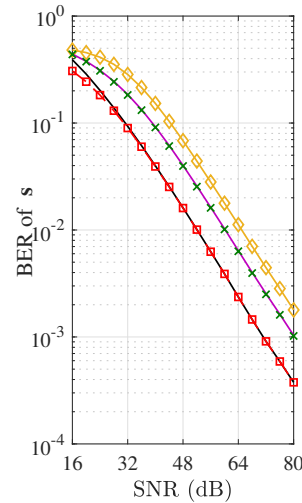
Figure 4.11: BER of x versus SNR

Figure 4.12: Performance of low-complexity detector versus SNR with $I \in \{0, 1, 2\}$, $N = 4$, $L = 4$, $p = 3$, and 256-QAM.

QRM and RIS-RPM schemes in Fig. 4.16, where the simulation parameters are $N = 4$, $L = 2$, $p = 1$, $\kappa_1 = \{-30, 30\}$ dB and 256-QAM. One interesting observation is that the effects of κ_1 on the error performance of x and on that of $\mathbf{s}$ are different. The error performance of x is improved when $\kappa_1 = 30$ dB, while the error performance of $\mathbf{s}$ is improved when $\kappa_1 = -30$ dB. This can be understood by the fact that the AP-RIS channel is a deterministic LoS channel when $\kappa_1 = 30$ dB and is a Rayleigh fading channel when $\kappa_1 = -30$ dB. The deterministic LoS channel facilitates the transmission of x while channel variations are used to convey additional information via the reflection modulation scheme. This observation implies that the appropriate deployment of an RIS needs to take into account the trade-off between the error performance of x and that of $\mathbf{s}$.
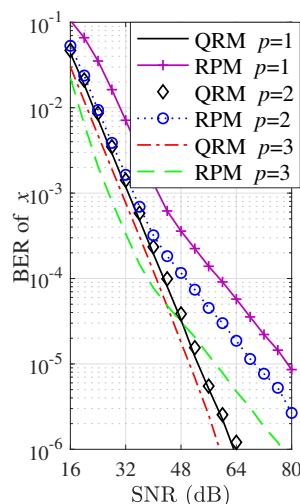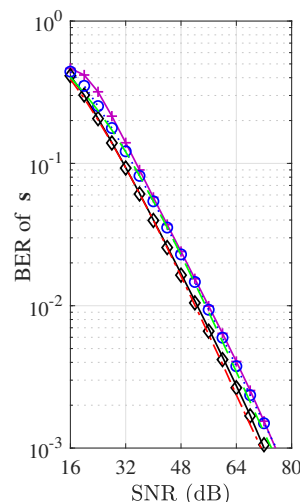
Figure 4.13: BER of x versus SNR

Figure 4.14: Effect of p on the error performance of RIS-QRM and RIS-RPM schemes with $N = 4$, $L = 4$, and 256-QAM.

4.6 Conclusions

In this section, we proposed the RIS-QRM scheme for simultaneous passive beam-forming and information transfer of RIS. In particular, for the SISO setting assuming that the direct link is blocked, a closed-form expression for the unconditional PEP of RIS-QRM in Rician fading was derived assuming the ML detection, which was used to calculate tight upper bounds on the average BEPs of the conventional constellation modulation and the RIS element partition. Moreover, we proposed a low-complexity detection method with lower performance for decoupling the joint detection of the con-stellation symbol and the RIS element partition. Computer simulation results have corroborated the effectiveness of RIS-QRM and validated the analytical results. In particular, RIS-QRM was shown to be able to improve the error performance of the additional bits delivered by the RIS without deteriorating that of the bits carried on the constellation symbol, as compared to ON/OFF-based reflection modulation schemes.
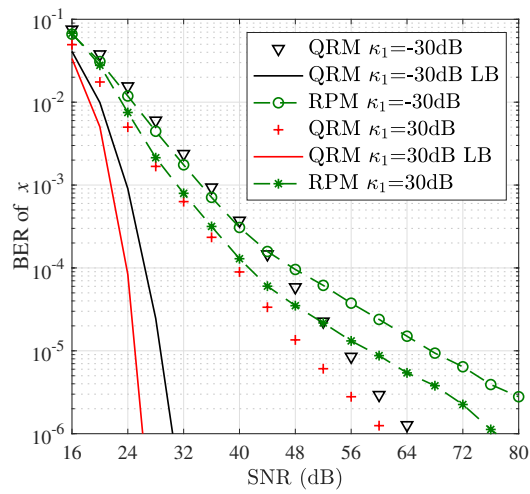
Figure 4.15: BER of x versus SNR

Figure 4.16: Effect of κ_1 on the error performance of RIS-QRM and RIS-RPM schemes with $N = 4$, $L = 2$, $p = 1$, and 256-QAM.

Chapter 5

Performance of Reconfigurable Intelligent Surface-Aided Links

5.1 Analytical Performance Assessment of Beamforming Efficiency in Reconfigurable Intelligent Surface-Aided Links

5.1.1 Introduction

In recent years, the dramatic increase in demand for high data rates has lead the research towards unallocated high frequency bands such as the millimeter-wave (mmWave) (30-100 GHz) and terahertz (THz) bands (0.1-10 THz) [93–97]. However, with the increase of frequency the transmitted signals are expected to be more vulnerable to the presence of obstacles (blockage), as compared to operation in the low GHz, for example [97–99]. Therefore, at such high frequencies, it is possible that the LoS link is blocked even at relatively short distances (in the order of a few meters), thus limiting the practicality of the mmWave and THz bands to mainly short range and LoS scenarios. In order to expand their use to non-line-of-sight (nLoS) scenarios, the use of RISs has been proposed for both indoor [100–103] and more general scenarios [26, 104–118].

In view of the constantly increasing volume of related works, it becomes clear that there are several parameters that need to be taken into account in order to assess the performance of a RIS-aided link. For example, the received power will strongly depend on the positioning of the RIS with respect to the transmitter and receiver, and will be affected by the RIS size and the properties of the transmitter beam. Therefore, there is a need for a simple analytical model that can clarify the crucial bounds imposed by the system design parameters on the link performance. In this work, in an attempt to provide a generalized approach, the efficiency of the RIS beamforming is studied in terms of the received power. The main contributions are summarized as follows.

- The system performance is qualitatively divided into two regimes of operation, with respect to the RIS size relative to the footprint of the transmitter beam on the RIS.

- For relatively small footprint, a simple analytical model is derived, providing insight on the interplay between crucial parameters, such as the positioning of the RIS

with respect to the transmitter and receiver, and the properties of the transmitter beam. The model clarifies why the received power does not increase monotonically with increasing gain, but rather acquires a maximum.

- For relatively large footprint, the impact of the RIS size on the maximum received power is demonstrated through simulations.

- The approach followed in this work provides simple guidelines for the algorithmic design of rules for optimized performance.
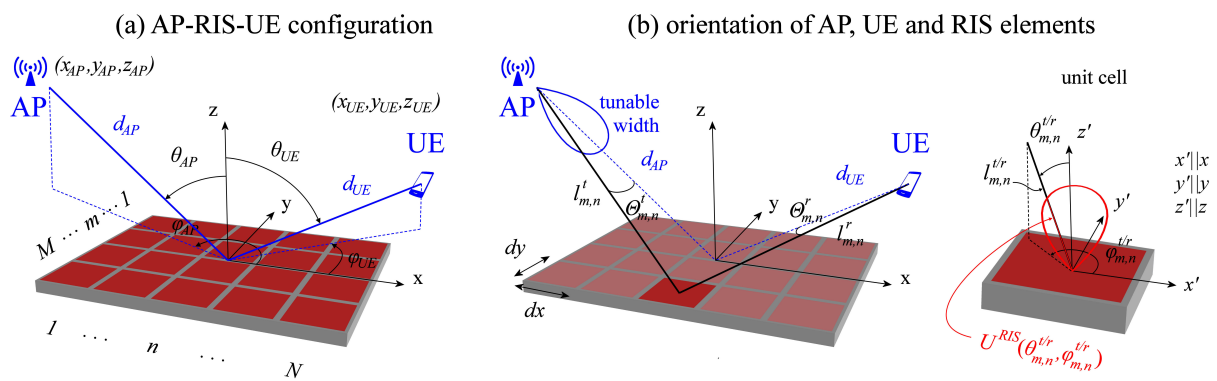
5.1.2 System Model

Figure 5.1: System model. (a) Schematic of a RIS-aided link, illustrating the positions and relative angles between the AP, RIS and UE. The RIS consists of $M \times N$ elements (here shown $4 \times 5 = 20$ elements). (b) Details on the orientation of the RIS elements and single unit cell with schematic representation of its radiation pattern. The angle $\Theta_{m,n}^t(\Theta_{m,n}^r)$ denotes the elevation angle from the transmitting (receiving) antenna to the unit cell. The angles $\theta_{m,n}^t$, $\phi_{m,n}^t$ ($\theta_{m,n}^r$, $\phi_{m,n}^r$) denote the elevation and azimuth angles from the unit cell to the transmitting (receiving) antenna. Along the designated direction, the normalized power radiation pattern of the unit cell, U^{RIS}, acquires the value $U^{RIS}(\theta_{m,n}^{t/r}, \phi_{m,n}^{t/r})$.

A RIS-aided system model is shown in Fig. 5.1, where the RIS is centered at the origin of the coordinate system and consists of $M \times N$ elements (unit cells), periodically arranged along the $x-$ and $y-$ axes, with periodicity dx and dy respectively. The AP is located at (x_{AP}, y_{AP}, z_{AP}) and is equipped with a highly directional antenna with tunable beam-width. The AP illuminates the RIS, which subsequently redirects the incident beam to the UE, located at (x_{UE}, y_{UE}, z_{UE}). In this work, misalignment between the AP and RIS or the RIS and UE is not considered, however the results can be directly extended to account for this possibility. The power received by the UE can be calculated using the power density, S_r, at the UE position and the effective aperture of the UE, A_r, as:

$$P_r = S_r A_r \equiv \frac{|E_r|^2}{2Z_o} A_r, \tag{5.1}$$

where E_r is the electric field at the position of the UE, Z_o is the characteristic impedance of air and $A_r = \frac{G_r\lambda^2}{4\pi}$, with G_r being the antenna gain of the UE and λ the free-space wavelength. The total electric field at the UE position is:

$$E_r = \sum_{m=1}^{M} \sum_{n=1}^{N} E_{m,n}, \tag{5.2}$$

where the contribution of the (m, n) RIS element to the total electric field is given by [26, 106]:

$$E_{m,n} = \sqrt{2Z_o P_t G_t A_{RIS} G_{RIS}} \frac{\sqrt{U^t U^{RIS,t} U^{RIS,r} U^r}}{4\pi l_{m,n}^t l_{m,n}^r}$$

$$|R_{m,n}| \exp\left(-j\phi_{m,n}\right) \exp\left(-j\frac{2\pi}{\lambda}(l_{m,n}^t + l_{m,n}^r)\right). \tag{5.3}$$

In Eq.(5.3) P_t, G_t are the transmitted power and antenna gain of the AP, respectively, A_{RIS}, G_{RIS} are the effective aperture and antenna gain of each RIS element, respectively, $l_{m,n}^t$, $l_{m,n}^r$ are the distances between the (m, n) RIS element and the AP and UE, respectively; $R_{m,n} = |R_{m,n}| \exp(-j\phi_{m,n})$ is the complex reflection coefficient introduced by the (m, n) RIS element. The normalized power radiation pattern of the AP, UE and the RIS element is denoted by U^t, U^r and U^{RIS}, respectively. In particular, $U^t \equiv U^t(\Theta_{m,n}^t, \Phi_{m,n}^t)$ is the value of U^t along the direction defined by the AP and the (m, n) element, as shown in Fig. 5.1(b), with the AP targeting the center of the RIS (origin of coordinate system). Similarly, $U^r \equiv U^r(\Theta_{m,n}^r, \Phi_{m,n}^r)$ is the value of U^r along the direction defined by the UE and the (m, n) element, with the UE also targeting the center of the RIS. Last, $U^{RIS,t/r}$ is the value of U^{RIS} along the direction defined by the positions of the (m, n) element and the AP/UE, as denoted with the superscript t/r, i.e. $U^{RIS,t/r} \equiv U^{RIS}(\theta_{m,n}^{t/r}, \phi_{m,n}^{t/r})$. Because the RIS elements are identical, they all share a common radiation pattern $U^{RIS} = \cos^2\theta_{\text{m,n}}^{\text{t/r}}$, with $\theta_{\text{m,n}}^{\text{t/r}} \in [0, \pi/2]$ and $\phi_{\text{m,n}}^{\text{t/r}} \in [0, 2\pi]$ defined with respect to the coordinate system (x', y', z') [see Fig. 5.1(b), right panel]. Note that, although U^{RIS} is common to all RIS elements, the angles $\theta_{m,n}^{t/r}, \phi_{m,n}^{t/r}$ change with (m, n) and, therefore, $U^{RIS,t/r}$ depends on the coordinates of each (m, n) element. Last, regarding the phase introduced by the RIS, it should be noted that a RIS design that aims to redirect a beam from the AP to the UE should provide the linear phase gradient given by [26, 106]:

$$\phi_{m,n} = \frac{2\pi}{\lambda}[(\sin\theta_{UE}\cos\phi_{UE} + \sin\theta_{AP}\cos\phi_{AP})x_m$$

$$+ (\sin\theta_{UE}\sin\phi_{UE} + \sin\theta_{AP}\sin\phi_{AP})y_n], \tag{5.4}$$

where $x_m = m \times d_x, m = 1, ..., M$ and $y_n = n \times d_y, n = 1, ..., N$ are the coordinates of the (m, n) RIS element.

5.1.3   RIS Illumination Regimes: RIS Size vs. AP Beam Footprint

In a conventional LoS link, the beam travels from the AP directly to the UE with increasing width –due to free-space propagation– and therefore with decreasing power

$$E = \frac{\sqrt{2Z_o \frac{2P_t}{\pi w_{RIS}^2}|R|^2}}{\sqrt{\left(1 - i\frac{z_B}{z_R}\right)\left(1 - i\frac{z_B}{z_R \cos^2\theta_B}\right)}}$$

$$\times \exp\left[-\frac{k_0}{2z_R}\left(\frac{x_B^2 + y_B^2}{1 - i\frac{z_B}{z_R}} - \sin^2\theta_B \frac{(x_B\cos\phi_B + y_B\sin\phi_B)^2}{\left(1 - i\frac{z_B}{z_R}\right)\left(1 + i\frac{z_R\cos^2\theta_B}{z_B}\right)}\right) - ik_0\Phi\right] \quad (5.5)$$

density, which can be further reduced by possible atmospheric absorption and scattering. As a result, for a certain emitted beam power, the narrower the beam-width at the UE, the higher the power density and the more the power that can be received by the UE. However, when an intermediate re-radiating element, such as a RIS, comes into play, this intuitive picture can change entirely. The reason is that the RIS is essentially a secondary antenna; it is driven by the incident field from the AP and is re-emitting a secondary beam (towards the UE), the propagation characteristics of which depend on both the incident power density on the RIS and the illuminated area.

For highly directional antennas, as considered here, the received power will be largely defined by the main lobe of the AP antenna, regardless of the exact radiation pattern, which will depend on the particular design specifications. Therefore, in order to be able to examine a simple analytical model that can capture the overall performance, beyond the antenna's particular characteristics, a convenient way to approximate the main lobe of a generalized AP antenna and to introduce tunability, is to consider a Gaussian radiation pattern (see e.g. [119]) of the form $G_t U^t \equiv G_t \exp(-\frac{G_t}{4}\sin^2\Theta_{m,n}^t)$, with $\Theta_{m,n}^t \in [0, \pi/2]$ and $\Phi_{m,n}^t \in [0, 2\pi]$ defined with respect to the coordinate system located at (x_{AP}, y_{AP}, z_{AP}) as shown in Fig. 5.1(b), with $m = 1...M, n = 1...N$. With the AP targeting the center of the RIS, the AP gain becomes maximum (equal to G_t) for the (m, n) element located at the origin (where $U_t = 1$). For highly directional beams, this form for the radiation pattern ensures that the total power radiated from the AP remains practically constant upon beam-width change, i.e. $\iint (P_t/4\pi d^2)G_t U^t dS \approx P_t$, where dS is the elementary surface element of a sphere of radius d centered at the AP.

In this notation, low G_t means a relatively wide AP beam that can illuminate the entire RIS; therefore, the incident power residing beyond the RIS surface cannot be captured by the RIS, leading effectively to power loss. Intuitively, to compensate for the lost power, G_t must be increased, in order to reduce the area of the AP beam-footprint on the RIS and achieve the utilization of the entire AP beam. Simultaneously, the power density of the beam will increase, as now the same incident power will be concentrated in a smaller region on the RIS, and more power is expected to be received by the UE, similarly to what would happen in a LoS link with increasing AP gain. However, a smaller AP footprint will lead the secondary beam (from the RIS to the UE) to undergo stronger spreading in air, subsequently reducing its power density and possibly leading to the opposite result.

From this discussion, it becomes apparent that, besides the incident power density, the size of the RIS illuminated area plays a decisive role on the overall performance of the RIS-aided link and, therefore, the trade-off between the RIS size, AP beam-
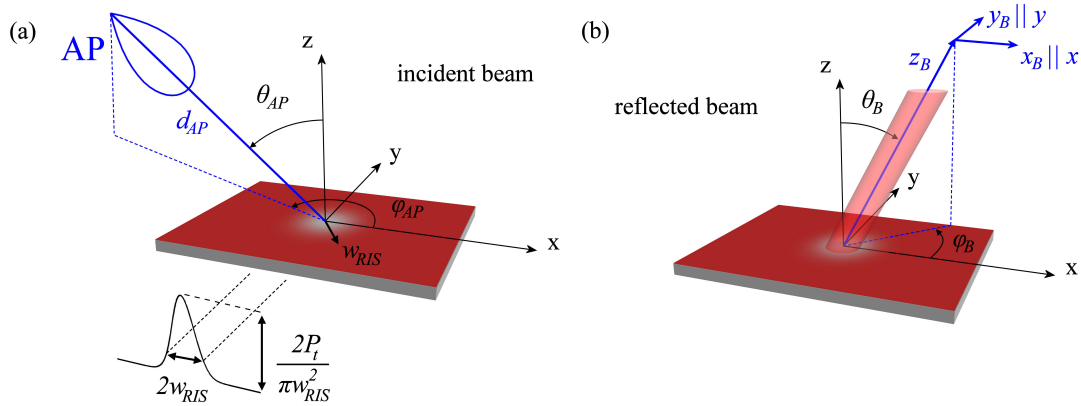
Figure 5.2: Schematic representation of RIS-aided beam steering. (a) The AP illuminates the RIS producing a Gaussian footprint of radius w_{RIS} and peak power density $2P_t/\pi w_{RIS}^2$ [see Eq.(5.6)]. (b) The incident beam is reflected by the RIS and propagates as a tilted Gaussian beam along the direction defined by the angles θ_B, ϕ_B. The non-orthogonal local beam coordinate system follows the beam as it propagates; z_B is the origin location along the propagation direction and the transverse coordinates, x_B and y_B, lie on a plane parallel to the x, y- plane.

width and AP power density, can lead to unexpected results with respect to the power received by the UE. To study this trade-off, two regimes of operation can be identified with respect to the RIS size. On the one hand, for relatively low G_t, the AP beam illuminates the entire RIS and, therefore, the footprint of the secondary beam does not depend on the AP beam width, rather it is determined by the actual area of the RIS (the effective aperture of the RIS is determined by the finite area of the RIS); we refer to this case as *finite RIS regime*. On the other hand, for relatively high G_t, the AP beam illuminates only a portion of the RIS and, therefore, the width of the secondary beam on the RIS is defined entirely by the AP beam footpint, i.e., the actual RIS size does not matter (the effective aperture of the RIS is determined by the AP beam footprint); in this case, the AP beam sees an effectively infinite RIS and we refer to this case as *infinite RIS regime*.

5.1.4 Infinite RIS Regime

When the entire AP beam is captured by the RIS, we can follow an alternative approach to model the RIS-aided link; we can leave aside the discrete nature of the RIS and treat it as a continuous surface, i.e. we can treat the beam as a continuous distribution. The advantage of this approach over the summation given by Eq.(5.2), is that the electric field at the position of the UE can be expressed in a simple closed form, enabling the derivation of analytical expressions that provide insight to the RIS operation. The starting point in this approach is to identify the properties of the AP footprint on the RIS. First, a beam of total power P_t illuminates the RIS, as illustrated schematically in Fig. 5.2(a). For a Gaussian beam, as considered here, the AP footprint on the RIS is also

Gaussian with radius w_{RIS} and power density:

$$S_{RIS}^{inc} = \frac{2P_t}{\pi w_{RIS}^2} \exp\left(-2\frac{x^2 + y^2}{w_{RIS}^2}\right).$$

(5.6)

Next, the AP footprint is reflected by the RIS and propagates as a tilted Gaussian beam along the direction defined by the angles θ_B, ϕ_B as shown in Fig. 5.2(b). The non-orthogonal local beam coordinate system follows the beam as it propagates; its origin is located at distance z_B from the center of the RIS along the beam propagation direction, and the transverse coordinates, x_B and y_B, lie on a plane parallel to the x, y- plane. Due to non-ideal reflection from the RIS, the power density of the reflected beam exactly at the RIS plane is $S_{RIS}^{inc}|R|^2$, where $R \equiv |R_{m,n}|$ is the amplitude of the common reflection coefficient of all RIS elements. Then, following a procedure as in [120], the electric field of the tilted beam produced by such a Gaussian footprint is given by Eq.(5.5), where z_R is the Rayleigh length, $k_0 = \frac{2\pi}{\lambda}$ is the free-space wavenumber, and:

$$\begin{aligned}
x_B &= x - z_B \sin\theta_B \cos\phi_B \\
y_B &= y - z_B \sin\theta_B \sin\phi_B \\
z_B &= z/\cos\theta_B.
\end{aligned}$$

(5.7)

The term Φ expresses the phase advance of the redirected beam and is given by: $\Phi = x_B \sin\theta_B \cos\phi_B + y_B \sin\theta_B \sin\phi_B + z_B$. For $\theta_B = \theta_{UE}$, $\phi_B = \phi_{UE}$ the RIS steers the beam towards the UE and, therefore, exactly at the UE position (x_{UE}, y_{UE}, z_{UE}) where $x_B = 0$, $y_B = 0$ and $z_B = d_{UE}$, the power density at the UE is simply given with the aid of Eq.(5.5) by:

$$S_r^{inf} = \frac{\frac{2P_t}{\pi w_{RIS}^2}|R|^2}{\sqrt{\left(1 + \frac{d_{UE}^2}{z_R^2}\right)\left(1 + \frac{d_{UE}^2}{z_R^2 \cos^4\theta_{UE}}\right)}},$$

(5.8)

where the superscript *inf* emphasizes the fact that the result of Eq. (5.8) derives from the continuous approach in the infinite RIS regime. Then, the power collected by the UE is $P_r^{inf} = S_r^{inf} A_r$, where A_r is the UE effective aperture introduced in Eq. (5.1).

The denominator of Eq. (5.8) implies that what is of importance is the RIS-UE distance (d_{UE}) relative to the Rayleigh length z_R, which is a function of w_{RIS} and the wavelength (see Appendix 5.1.8 for details):

$$z_R = \frac{k_0 w_{RIS}^2}{2} \equiv \frac{4k_0 d_{AP}^2}{G_t}.$$

(5.9)

In the limit $d_{UE}/z_R \ll 1$, Eq. (5.8) acquires a simple form:

$$d_{UE} \ll z_R : P_r^{inf} = A_r|R|^2\frac{2P_t}{\lambda}\frac{1}{z_R},$$

(5.10)

i.e. $P_r^{inf} \sim G_t$ [see Eq. (5.9)], as in typical LoS links. However, in the other extreme:

$$d_{UE} \gg z_R : P_r^{inf} = A_r|R|^2\frac{2P_t}{\lambda}\frac{\cos^2\theta_{UE}}{d_{UE}^2}z_R,$$

(5.11)

i.e. $P_r^{inf} \sim 1/G_t$, which means that P_r^{inf} decreases with increasing G_t, contrary to what might be intuitively expected, as in the previous case. The two limits imply that there should be a maximum P_r^{inf} for some value of G_t, which can be found by differentiating Eq. (5.8) with respect to z_R. The maximum power received by the UE is:

$$\max(P_r^{inf}) = A_r |R|^2 \frac{2P_t}{\lambda d_{UE}} \frac{\cos^2 \theta_{UE}}{1 + \cos^2 \theta_{UE}} \quad (5.12)$$

and occurs when:

$$z_R \cos \theta_{UE} = d_{UE}. \quad (5.13)$$

Interestingly, the maximum P_r^{inf} does not depend on the position of the AP. This is a direct consequence of the fact that the RIS size is much larger than the AP beam footprint and, hence, the latter can be tuned without limitation. Furthermore, because the Rayleigh length z_R is the distance denoting the transition from the near-field to the far-field of the beam, the result of Eq. (5.13) dictates that the maximum P_r^{inf} occurs when the UE resides in the vicinity of this transition. In essence, a change in G_t, the gain of the AP, modifies the relative distance z_R with respect to the position of the UE [see Eq. (5.9)]. By increasing G_t, w_{RIS} decreases, in turn reducing the distance z_R and effectively transferring the UE from the near- to the far-field of the beam. Therefore, relatively low G_t essentially means that $d_{UE} < z_R$ and increasing G_t will increase P_r^{inf} [see Eq.(5.10)] until the maximum P_r^{inf} is reached [Eq. (5.12)]. Further increase in G_t will lead to $d_{UE} > z_R$ and to gradual reduction of P_r^{inf} [see Eq.(5.11)]. The necessary AP gain to reach the maximum P_r^{inf} can be derived from Eq. (5.13) with the aid of Eq. (5.9) as:

$$G_t^{inf} = 4k_0 \cos \theta_{UE} \frac{d_{AP}^2}{d_{UE}}. \quad (5.14)$$

To demonstrate the above, a D-band in-door scenario is considered. The operating frequency is 150 GHz and the AP is placed at distance $d_{AP} = 1$ m from the RIS with $\theta_{AP} = 0^o$, while the UE is placed at $d_{UE} = 2$ m with $\theta_{UE} = 20^o$ [see Fig. 5.1]. The AP, RIS and UE are all on the plane $y = 0$, i.e. $\phi_{AP} = 180^o$ and $\phi_{UE} = 0^o$. The AP transmits a beam of tunable gain, G_t, and constant power $P_t = 1$ W, while the UE gain, G_r, is 20 dB. For a theoretically infinite RIS, the power received by the UE is found using Eq.(5.8) with the aid of Eq.(5.9) and is shown in Fig. 5.3(a), as a function of the AP gain, G_t. The equivalent numerical calculation using the discrete model of Eq.(5.3) is also shown, verifying the theoretically expected power levels. To account for the theoretically infinite RIS in the numerical model, a panel with 1200×1200 elements of size $dx = dy = \lambda/5$ has been considered. For the AP-RIS distance and AP gain levels used in this example, this size is adequate for the RIS to essentially capture the entire AP beam and operate as an effectively infinite panel. In the same Figure, Eq.(5.10) and Eq.(5.11) are shown as dashed lines, marking the asymptotic behavior of the communication link in the limit of relatively low and high AP gain. Essentially, the asymptotes serve as an upper threshold for the received power when the link resides in the infinite RIS regime. The vertical dashed line marks the AP gain that provides the maximum power at the UE and is extended to Fig. 5.3(b), to emphasize that this is in agreement with G_t^{inf} as predicted by Eq. (5.14), which results from the crossing
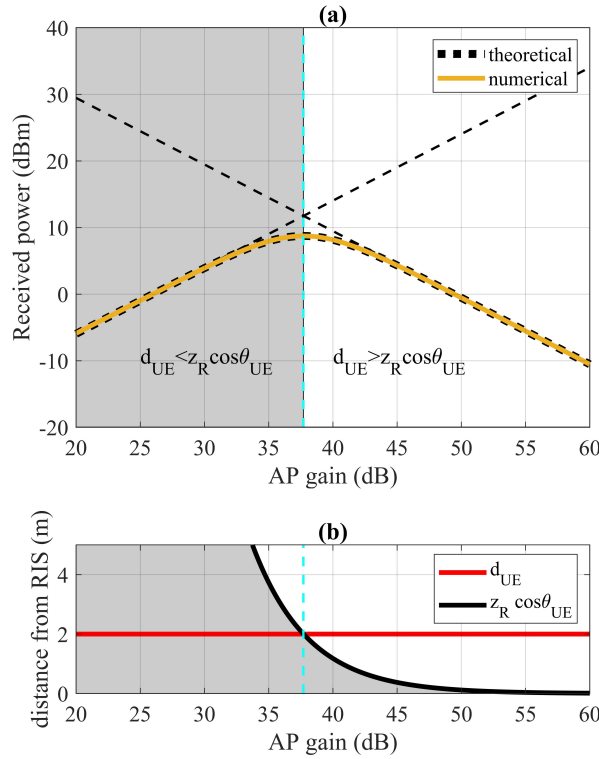
Figure 5.3: RIS-aided link in the infinite RIS regime. (a) Power received by the UE, P_r, as a function of the AP gain, G_t, for an infinitely large RIS (dashed line) and a RIS with $M = N = 1200$ (solid line). (b) Graphic solution of Eq. (5.13), marking G_t^{inf}, the AP gain that maximizes the received power. The transmitted power is $P_t = 1$ W, the UE gain $G_r = 20$ dB and the AP, RIS and UE lie on the same plane with $d_{AP} = 1$ m, $\theta_{AP} = 0^o$ and $d_{UE} = 2$ m, $\theta_{UE} = 20^o$.

of $z_R \cos \theta_{UE}$ with d_{UE} [graphic solution of Eq.(5.13)] and marks the transition from the near- to the far- field of the beam.

In Fig. 5.4 further examples for variable AP and UE positions are presented. In Fig. 5.4(a), the AP is at the position previously examined in Fig. 5.3 and the UE position is scanned along the direction $\theta_{UE} = 20^o$. For high AP gain ($d_{UE} \gg z_R$) the pathloss increases with increasing d_{UE} and the received power P_r^{inf} decreases, as intuitively expected and also predicted by Eq. (5.11). For low AP gain ($d_{UE} \ll z_R$), however, the received power is independent of the UE position, as also captured by Eq. (5.10); this is a direct consequence of the fact that, because the beam propagation is deeply in the near-field, the peak power of the secondary beam is practically constant and the pathloss is therefore similar for all examined cases. Furthermore, with increasing distance, d_{UE}, the maximum received power drops and moves to lower AP gain, as predicted by Eqs. (5.12),(5.14). Physically, this means larger AP footprint, i.e. weaker beam spreading and higher power density at the UE, thus compensating for the increased pathloss at the UE that is placed at larger distances from the RIS. In Fig. 5.4(b) the UE is at the position previously examined in Fig. 5.3 and the AP position is scanned along the direction $\theta_{AP} = 0^o$. In this case, the maximum received power is independent of the position of the AP [see Eq. (5.12)] and undergoes a shift towards higher AP gain with increasing d_{AP} [see Eq. (5.14)]. The reason is that, while an increasing AP-RIS distance leads to an increasing AP footprint and lower power density,

with higher AP gain, the same power levels can be restored at a fixed UE position, as demonstrated here. In essence, in the infinite RIS regime, there is always a combination of AP-RIS distance and AP gain that can lead to practically the same power at the UE. This does not hold always, as we will show next, when the finite RIS size comes into play.
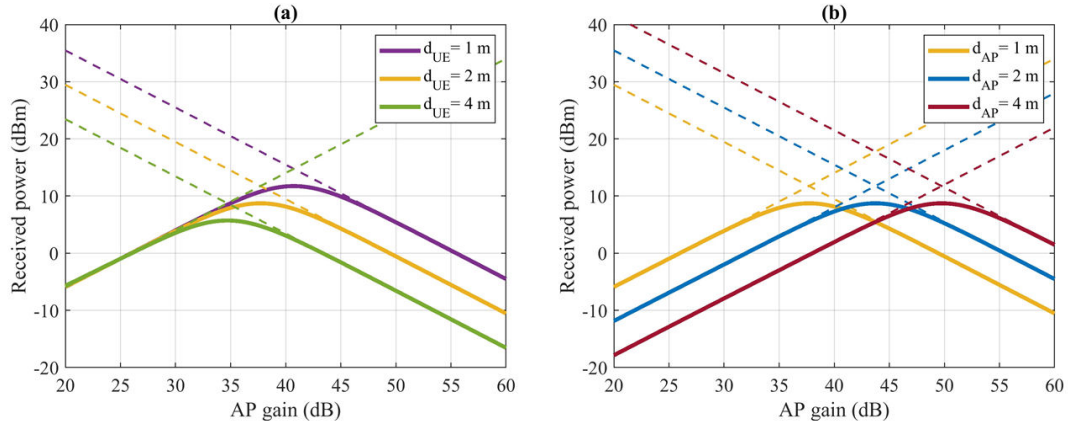

Figure 5.4: Received power, P_r, vs AP gain, G_t, for a RIS-aided link residing in the infinite RIS regime, with a RIS of $M = N = 1200$ elements. (a) Variable RIS-UE distance, d_{UE}, along the direction $\theta_{UE} = 20^o$ with fixed AP position ($d_{AP} =$1 m, $\theta_{AP} = 0^o$). (b) Variable RIS-AP distance, d_{AP}, along the direction $\theta_{AP} = 0^o$ with fixed UE position ($d_{UE} =$2 m, $\theta_{UE} = 20^o$). The AP, RIS and UE lie on the same plane, the transmitted power is $P_t = 1$ W and the UE gain is $G_{UE} = 20$ dB.
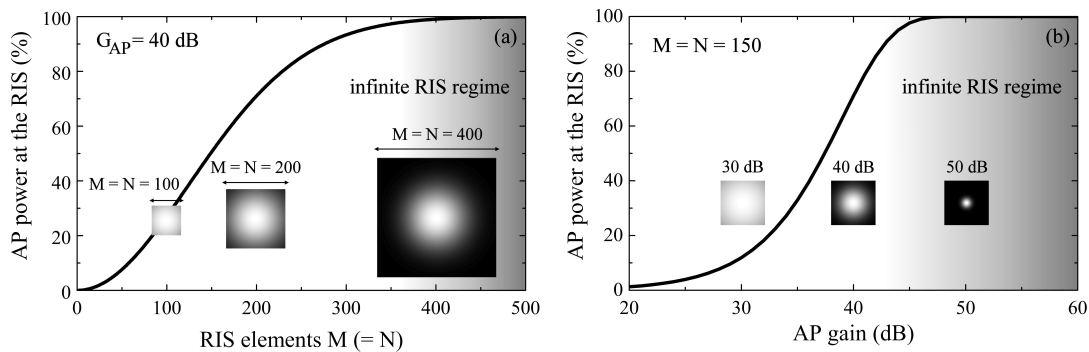

Figure 5.5: Effect of finite RIS size on the AP power received by the RIS. (a) Variable RIS size under constant AP gain ($G_{AP} = 40$ dB). (b) Variable AP gain under constant RIS size (M = N = 100 elements). The shading denotes that, with increasing AP gain, the communication link gradually enters the infinite RIS regime.

5.1.5   Impact of Finite RIS Size: From Pencil Beams to Wide Beams and Full RIS Coverage

Infinitely large RIS does not necessarily imply an actually infinite surface. In practice, it suffices to have a finite-size RIS that is large enough, so that the incident beam is captured entirely by the RIS surface and further increase in the RIS size does not

impose any change on the propagation properties of the re-directed beam. Therefore, a direct way to quantify the RIS size is via the percentage of the incident power that is collected by the RIS, as shown in Fig. 5.5. Clearly, by increasing either the RIS size with fixed AP gain [panel (a)] or the AP gain with fixed RIS size [panel (b)], the power received by the RIS increases until the entire beam is captured. At this limit, the total power captured by the RIS reaches 100% of the incident power and the communication link enters the regime of infinitely large RIS.
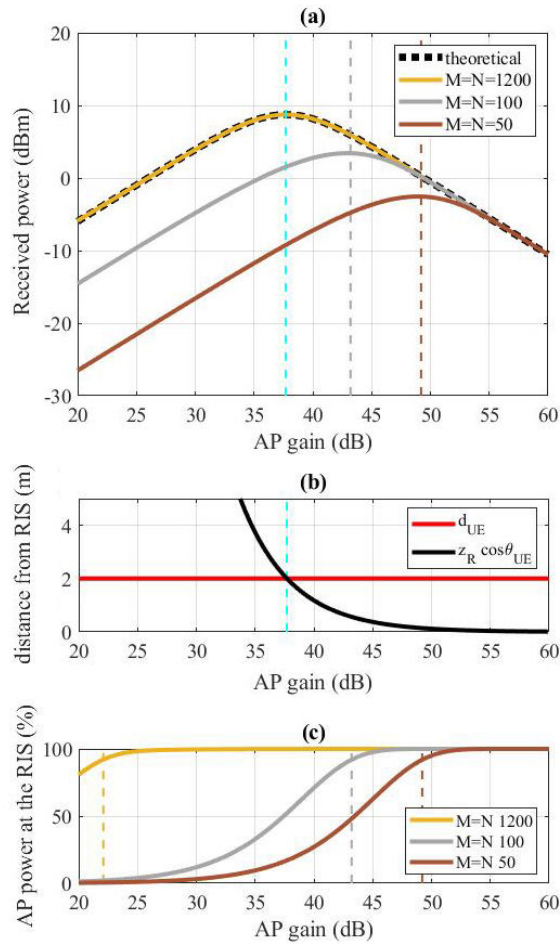

Figure 5.6: Impact of finite RIS size on received power. (a) Received power, P_r, (b) graphic solution of Eq. (5.13), and (c) power received by the RIS, vs AP gain, G_t, for a RIS with $M = N = 50$, 100 and 1200 elements. The transmitted power is $P_t = 1$ W, the UE gain $G_r = 20$ dB and the AP, RIS and UE lie on the same plane with $d_{AP} = 1$ m, $\theta_{AP} = 0^o$ and $d_{UE} = 2$ m, $\theta_{UE} = 20^o$.

To demonstrate the effect of the finite RIS size, the scenario of Fig. 5.3 is re-examined in Fig. 5.6 under variable RIS size. Fig. 5.6(a) shows the received power versus the AP gain, G_t, Fig. 5.6(b) shows graphically the solution of Eq. (5.13), which predicts the AP gain for maximum P_r in the infinite RIS regime and Fig. 5.6(c) shows the transition from the finite to the infinite RIS regime by means of the percentage of the P_t power that is collected by the RIS, as a function of the AP gain. Starting with a large RIS of $M = N = 1200$ elements, the system clearly resides in the infinite RIS regime, as can be identified in Fig. 5.6(c). As a result, the numerically calculated power

shown in Fig. 5.6(a) agrees with the theoretically expected for the infinite RIS, and the maximum received power marked with the cyan dashed line in Fig. 5.6(b) is located at the position predicted by Eqs. (5.12),(5.14). With decreasing RIS size the system enters the finite RIS regime as can be identified in Fig. 5.6(c), with implications on the power received by the UE. For low AP gain, the received power decreases with decreasing RIS size [Fig. 5.6(a)], as the amount of P_t captured by the RIS is reduced. For high AP gain, on the other hand, the received power is independent of the RIS size as the footprint of the AP beam on the RIS is always smaller than the RIS size (for the examples considered here) and, therefore, the propagation properties of the secondary beam are not affected by the RIS size. Furthermore, with decreasing RIS size, higher AP gain is required to reach the maximum P_r. This is a direct consequence of the fact that the finite RIS size truncates the incident power and, to compensate for the lost power, the AP footprint must become smaller, so that the entire incident power can be captured by the RIS. Interestingly, the AP gain that restores the power lost due to the finite RIS size is the gain that transfers the system to the infinite RIS regime, i.e. where the power received by the RIS approaches 100% of P_t. Further increase in the AP gain will result in no further power compensation, but to a smaller AP footprint and, therefore, stronger beam spreading and lower power density at the UE.

Last, the example of a UE with variable position -previously studied in Fig. 5.4(a) for an infinitely large RIS- is now examined in Fig. 5.7 for a small RIS consisting of 80×80 elements. Here, in contrast to the infinitely large RIS case, the received power drops with increasing AP-RIS distance, regardless of the AP gain and the maximum P_r occurs for the same AP gain [panel (a)]. In accordance with the behavior observed in Fig. 5.6, this is a direct consequence of the fact that, in all cases examined, the theoretically expected AP gain that marks the maximum P_r [panel (b)] is located at lower AP gain levels than the gain that marks the transition from the infinite to the finite RIS regime [panel (c)]. In other words, the power loss from the AP footprint on the RIS due to the finite (relatively small) RIS size dominates, leading to this more intuitive and typical performance.

5.1.6 Example Algorithm for the Optimization of the Received Power at a Moving UE

In a communication environment, where the AP beam properties and the RIS operation are tuned according to the UE requirements, there is a need for simple algorithmic description of the possible decisions that have to be made in order to maximize certain key performance indicators (KPIs). With the analysis presented in this work, the dimensionality of the multi-parameter communication space reduces to a few quantities, such as the size of the RIS and the gain of the AP. Then, depending of the KPIs of the particular link, decisions can be made based on these quantities.

For example, in a link where the power received by the UE must be maximized, the entire problem can be expressed in terms of two parameters only, namely G_t^{inf} [given by Eq. (5.14)] and the gain that marks the transition from the finite to the infinite RIS regime, which we denote here as $G_t^{M \times N}$ (the superscript emphasizes the dependence on the RIS size). From the examples of Fig. 5.6 and Fig. 5.7 it becomes apparent that if $G_t^{M \times N} > G_t^{inf}$, then $G_t \equiv G_t^{M \times N}$, while if $G_t^{M \times N} < G_t^{inf}$, then $G_t \equiv G_t^{inf}$. A simple algorithm is shown in the flow chart of Fig. 5.8(a), where $G_t^{M \times N}$ is calculated only once
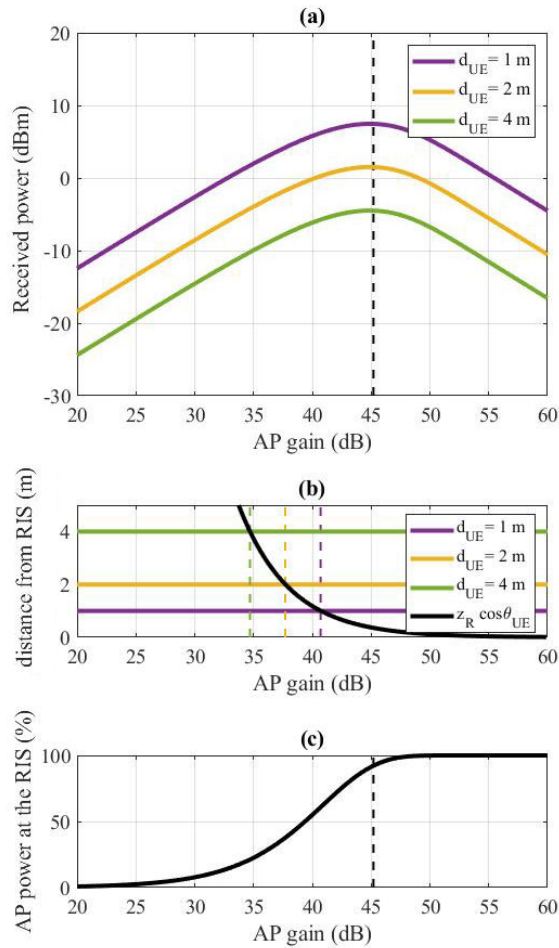
Figure 5.7: RIS-aided link residing in the finite RIS regime, with a RIS of $M = N = 80$ elements. (a) Received power, P_r, (b) graphic solution of Eq. (5.13), and (c) power received by the RIS, vs AP gain, G_t, for variable UE positions along the direction $\theta_{UE} = 20^o$. The transmitted power is $P_t = 1$ W, the UE gain is $G_{UE} = 20$ dB and the AP, RIS and UE lie on the same plane, with the AP position fixed ($d_{AP} = 1$ m, $\theta_{AP} = 0^o$).

in the beginning and G_t^{inf} is evaluated in real-time, as the UE moves. Another example is when the RIS can be separated in independent blocks, to serve multiple users. In this case $G_t^{M \times N}$ can be calculated a priory for the entire RIS and each block individually, and be compared with G_t^{inf} in a similar manner, as shown in the example of Fig. 5.8(a).

As an implementation of the algorithm shown in Fig. 5.8(a), a scenario of a moving UE is demonstrated in Fig. 5.8(b). As the UE moves, the algorithm decides on the AP antenna gain that optimizes the received power at the UE, leading to an overall stronger reception as compared to the power received under constant AP antenna gain. In this example, two cases of constant gain are shown (45 and 50 dB), and the optimized gain for each UE position is shown in the color bar. The size of the RIS is 500×500 elements, the AP is located at $\theta_{AP} = 0^o$ and $d_{AP} = 2$ m, while $\theta_{UE} = 20^o$ and d_{UE} ranges from 0.1 to 3 m.
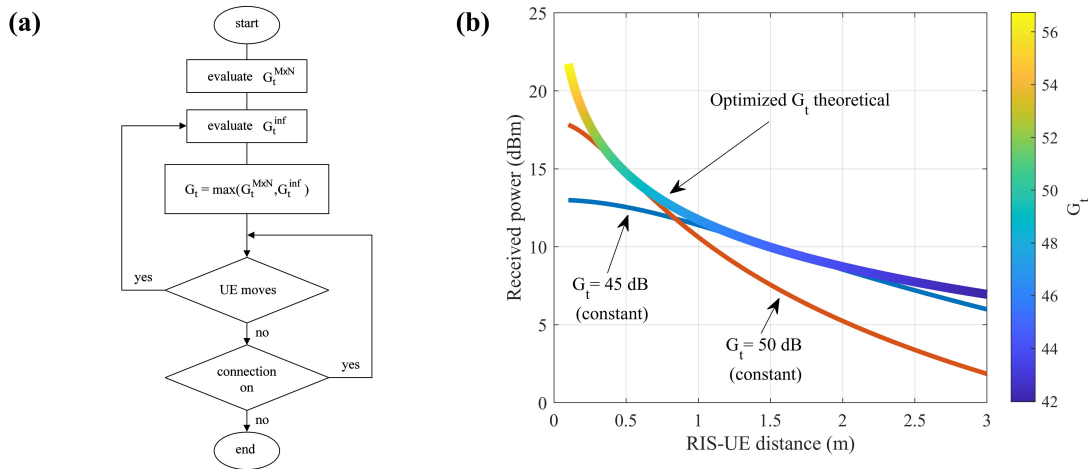
Figure 5.8: Example algorithm for the optimization of the received power at the UE. (a) Flow chart and (b) implementation of algorithm for a RIS of 500×500 elements, with the UE moving along the line with $\theta_{UE} = 20^o$ and the AP located at $\theta_{AP} = 0^o$ and $d_{AP} = 2$ m. G_t^{inf} depends on both the AP-RIS and RIS-UE distance and is evaluated when a UE movement is detected. $G_t^{M \times N}$ depends only on the AP-RIS distance and the RIS size and is, therefore, evaluated only once. For comparison, two examples with constant gain are also shown.

5.1.7 Conclusion

The successful adoption of RISs in future wireless systems is highly dependent on crucial choices for the optimal RIS placement and the efficient transceiver design. In turn, these choices are closely related to certain design parameters that need to be taken into account when designing a RIS-aided link. In this work, the interplay between the RIS size and the properties of the transmitter beam was investigated in terms of the RIS beamforming efficiency. The study revealed that, even for infinitely large RIS, the increase of the transmitter gain does not always guarantee increased power at the receiver, rather there is a maximum, which is further affected by the finite RIS size. The performance of the RIS-aided link was evaluated as a function of the AP gain, revealing two distinct operation regimes, with respect to whether the size of the AP beam footprint is larger or smaller than the RIS area, termed here as *finite RIS regime* and *infinite RIS regime*, respectively. In the infinite RIS regime, an analytical model was presented, predicting the maximum received power and providing insight on the underlying mechanism. The theoretically expected performance was also verified by numerical examples of typical in-door scenarios, elucidating the role of variable RIS size, AP-RIS distance and RIS-UE distance.

5.1.8 Appendix

The footprint of the AP beam on the RIS given by Eq. (5.6), can be written as:

$$S_{RIS}^{inc} = \frac{2P_t}{\pi w_{RIS}^2} U_t, \tag{5.15}$$

with

$$U_t = \exp\left(-2\frac{x^2 + y^2}{w_{RIS}^2}\right) \tag{5.16}$$

expressing the AP radiation pattern at distance $\sim d_{AP}$, where the RIS is located. The power density of Eq. (5.15) can be alternatively expressed in terms of G_t, the AP gain, as:

$$S_{RIS}^{inc} = G_t U_t \frac{P_t}{4\pi d^2}, \tag{5.17}$$

where d is the radius of the sphere centered at the AP, and for pencil beams $d \sim d_{AP}$. From Eq. (5.15) and Eq. (5.17) we find:

$$G_t = 8\left(\frac{d_{AP}}{w_{RIS}}\right)^2. \tag{5.18}$$

The result of Eq. (5.18) is used in Eq. (5.9) of the main text, to express z_R in terms of G_t. Additionally, the radiation pattern given by Eq. (5.16) can be simplified using the result of Eq. (5.18) and the approximation $\sqrt{x^2 + y^2} \sim d_{AP}\sin\theta$, leading to:

$$U_t = \exp\left(-2\frac{d_{AP}^2 \sin^2\theta}{w_{RIS}^2}\right) = \exp\left(-\frac{G_t}{4}\sin^2\theta\right), \tag{5.19}$$

where θ corresponds to the angle $\Theta_{m,n}^t$ shown in Fig. 5.1 and explained in the relevant text.

5.2 An Analytical Framework for Reconfigurable Intelligent Surfaces Placement in a Mobile User Environment

5.2.1 Introduction

Depending on the RIS placement with respect to the AP and UE positions, different power levels can reach the UE, thus affecting the quality of the communication. In most works [110, 115, 117, 121–126], the RIS placement is based on maximizing the cell coverage, the SNR (worst-case or not), the SINR, the minimum user data rate, and the minimization of the blockage probability in both the AP-RIS and RIS-UE links, however with a fixed antenna gain on the transmitter. In this work, the optimal RIS placement in a mobile user environment is studied on the basis of the interplay between three crucial quantities, namely (a) the available AP gain, (b) the available positions for the AP and RIS placement, and (c) the minimum desired power levels at the UE. With the use of the analytical model derived previously in [127] the possibilities for both tunable and non-tunable AP antenna gain are examined. The optimal RIS placement for a fixed UE position is identified, whereas for the mobile UE case a multiplicity of RIS placement solutions is shown to exist. The study is demonstrated through simulations
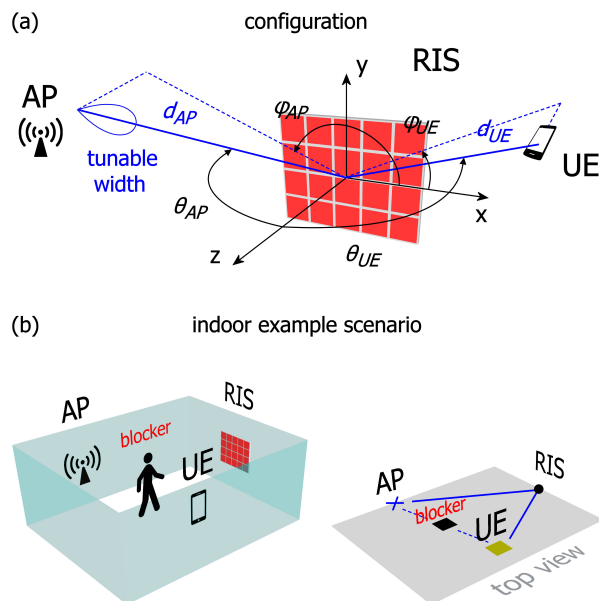
Figure 5.9: System model. (a) Schematic of a RIS-aided link, illustrating the positions and relative angles between the AP, RIS and UE. The red rectangles represent the RIS elements (here shown $4 \times 5 = 20$ elements). (b) Example configuration in a typical indoor scenario. The AP, RIS and UE positions are marked with the blue cross, black dot and yellow box, respectively. A blocker interrupts the line-of-sight connection between the AP and UE (dashed line) and the communication link is restored via the RIS (solid lines).

for an indoor scenario, specifically a rectangular room, where a range of locations for the AP and the RIS are tested for several power thresholds. The results demonstrate that the optimal RIS placement relies on the AP antenna gain tunability and, as with tunable antennas there is always an optimized AP gain that maximizes the received power, the difference in the power delivered to the UE with using the optimized versus non-optimized AP antenna gain is also examined. The results provide insight into the parameters that play a crucial role for the optimal selection of the AP antenna gain and the RIS placement.

5.2.2   System Model

A RIS-aided system model is shown in Fig. 5.9(a), where the RIS is centered at the origin of the local coordinate system (with the RIS as frame of reference) and the red-colored rectangles represent the RIS elements (or unit cells), which are periodically repeated along $x-$ and $y-$ directions. The AP is located at distance d_{AP} from the RIS center, with elevation angle θ_{AP} and azimuth angle ϕ_{AP}, as seen by the RIS. The AP, which is equipped with a highly directional antenna with tunable beam-width, illuminates the RIS, which subsequently redirects the incident beam to any desired direction. Typically, the RIS elements are adjusted appropriately so that the incident beam is redirected towards the UE, which is located at distance d_{UE} from the RIS center, with elevation angle θ_{UE} and azimuth angle ϕ_{UE}, as seen by the RIS. The flat surface of the RIS makes it ideal for placement on a wall, as shown schematically in Fig. 5.9(b).

The power received by the UE can be calculated by Eq. (5.1), using the power

density, S_r, at the UE position and the effective aperture of the UE, A_r. To calculate S_r, the contribution from each unit cell to the total electric field at the UE position is required, leading to an E-field summation over all unit cells, as adopted in several works [26, 106]. Alternatively, due to the sub-wavelength nature of the RIS elements (usually in the order of $\lambda/5$ or less), the RIS can be treated as a continuous surface, i.e. the beam can be considered as a continuous distribution. The advantage of this approach over the E-field summation is that the electric field at the position of the UE can be expressed in a simple closed form, enabling the derivation of analytical expressions that provide insight to the operation of the RIS. For relatively wide beams it is possible that only a fraction of the incident beam is captured by the RIS, thus altering its shape when impinging on the RIS. On the other hand, for pencil beams produced by highly directional antennas, the entire AP beam is captured by the RIS, and the beam from the AP to the UE can be therefore described entirely by the same continuous distribution. Using a Gaussian distribution to describe the main lobe of the AP antenna, in the previous section it was shown that, in the absence of misalignment between the RIS beam and AP antenna, the power density at the UE is given by:

$$S_r = \frac{\frac{2P_t}{\lambda z_R}|R|^2}{\sqrt{\left(1 + \frac{d_{UE}^2}{z_R^2}\right)\left(1 + \frac{d_{UE}^2}{z_R^2 \cos^4 \theta_{UE}}\right)}}. \tag{5.20}$$

In this expression $|R|$ is the common reflection amplitude of all RIS elements and z_R is the Rayleigh length, which can be expressed in terms of $k_0 = \frac{2\pi}{\lambda}$, the free-space wavenumber, and G_t, the AP antenna gain, by Eq. (5.9). Using Eqs.(5.1),(5.20), one can calculate the power received by the UE, e.g. for the typical case where the AP and RIS are at specific positions and the UE is mobile. Alternatively, if there are several available positions for the RIS, one might want to assess the RIS placement that maximizes the received power at the UE, for a certain UE position. In general, this will depend on several parameters, including the available space for the RIS placement.

As an example, in Fig. 5.10, a D-band in-door scenario is considered. The operating frequency is 150 GHz and the AP (marked with the blue cross) is placed at the corner of a 4×5 m^2 rectangular room; the RIS can be placed anywhere at the top wall, which is located at $z = 4$ m in the global coordinate system (with the room as frame of reference) shown in Fig. 5.10(a). The AP transmits a beam of tunable gain, G_t, and constant power $P_t = 1$ W, and the UE gain, G_r, is 20 dB. The position of the UE is marked with the yellow square and all elements (AP, RIS, UE) are on the xz-plane, which coincides with the xz-plane of the local coordinate system of the RIS, i.e. $\sin \phi_{AP} = \sin \phi_{UE} = 0$. The line-of-sight path (dashed line) is interrupted by a blocker and is restored via the RIS. The solid lines denote the propagation path from the AP to the UE for each of the four RIS positions shown here and correspond to a total distance of $d_{AP} + d_{UE}$. The power received by the UE for the four selected RIS positions is shown in Fig. 5.10(b) as a function of the AP gain, G_t. Clearly, with a tunable AP antenna, P_r is maximized when the UE is exactly in front of the RIS, i.e. for $\theta_{UE} = 0$. Indeed, by scanning consecutive RIS positions along the entire available space given by the top wall, it is found that the optimal RIS placement is located at $x = 3$ m (purple dot). If the antenna is non-tunable, i.e. provides a predefined fixed gain, the same conclusion practically holds for relatively high G_t (relatively narrow beams), as can be identified in the high gain region

of Fig. 5.10(b). Interestingly, if the AP antenna beam is relatively wide (low G_t), the calculations reveal that placing the RIS closer to the AP is beneficial over pursuing its placement closer to the UE; this can already be identified in the low gain region of Fig. 5.10(b).
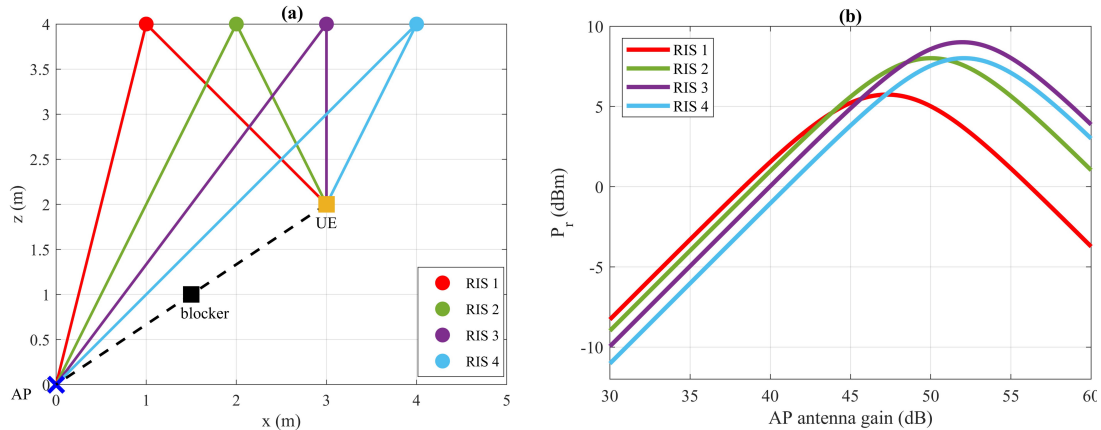

Figure 5.10: Study of RIS placement in an indoor scenario. (a) Configuration, showing a 4×5 m^2 room with the AP and UE at fixed positions, where the top wall is available for the RIS placement. The dashed line shows the line-of-sight path which is interrupted by a blocker. The solid lines denote the restored propagation path from the AP to the UE for variable RIS positions. (b) Power received by the UE as a function of the AP antenna gain (G_t), for the four selected RIS positions shown in (a).

5.2.3 Optimal RIS Placement for Mobile UE

The example of Fig. 5.10 demonstrates how the analytical model presented in this work can be used to locate the optimal RIS placement when the UE is immobile. However, if the user is mobile, different UE positions will lead to the requirement for different RIS placement and, due to the numerous possibilities, it becomes apparent that the complexity of the problem can increase significantly.

A straightforward way to address this challenge is by scanning the available positions for the RIS as a function of the UE location within the expected region of mobility. By imposing a threshold for the minimum acceptable power received by the UE, the parametric scan will yield position ranges instead of isolated points at specific walls. Under the discrete approach adopted in other works [26, 106] this may lead to increasingly long calculation times, because the $E-$field must be calculated at the UE as a sum over all RIS elements; and this must be repeated for every UE and RIS position. Importantly, the calculation complexity can increase severely for a tunable AP antenna, if the problem must be solved repeatedly as a function of the AP gain. However, with the analytical model presented in this work, the calculation complexity is limited to only the extent of the spatial discretization for the possible RIS and UE positions.

Optimized AP Antenna Gain

When the AP antenna is tunable, there is always a maximum in the received power, as shown in Fig. 5.10(b) and explained in detail in [127]. Therefore, the decision for the

optimal RIS placement will be the outcome of the comparison among all such maxima for the various RIS and UE positions. The maximum received power can be calculated directly from Eq.(5.20) by differentiating with respect to z_R and is given by the result of Eq. (5.12), which we repeat here for convenience:

$$P_r^{max} = A_r |R|^2 \frac{2P_t}{\lambda d_{UE}} \frac{\cos^2 \theta_{UE}}{1 + \cos^2 \theta_{UE}}. \tag{5.21}$$

The required gain to achieve this maximum is referred to as *optimized* gain. Using the result of Eq.(5.21), the example of Fig. 5.10 can be revisited with the UE moving freely within the 4×5 m^2 area. For each possible UE and RIS position, the maximum received power is calculated using Eq.(5.21) and is compared with a desired threshold. The RIS positions that can deliver power above the threshold for all UE positions within the designated area (except for a 0.25 m margin from the walls) are marked with a black dot, yielding regions for the RIS placement as shown in Fig. 5.11, the extent of which depends strongly on the imposed threshold. For example, in Fig. 5.11(a), top row, where the AP is placed at the bottom left corner, the threshold is set to -10 dBm. Due to the particular AP placement, the RIS cannot be illuminated by the AP if placed on the left or bottom wall and, therefore, the solutions appear only for the top and right wall. In Fig. 5.11(b), top row, the threshold is reduced to -7.5 dBm, leaving only the right
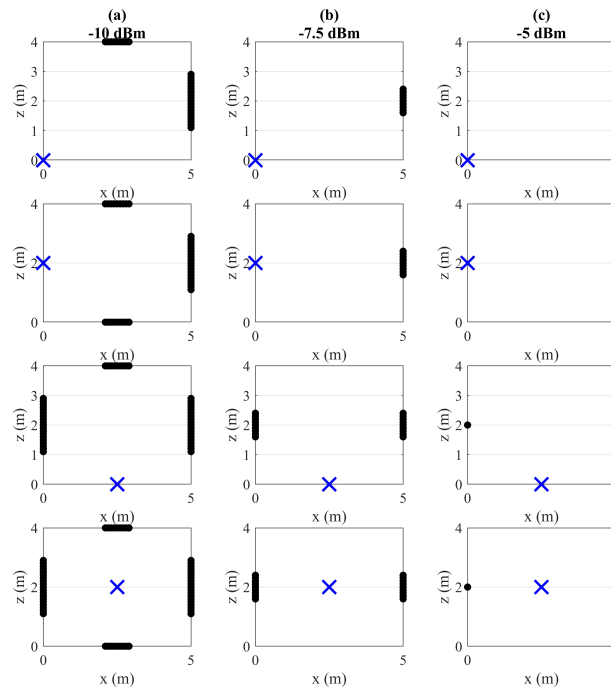

Figure 5.11: Optimal RIS placement for mobile UE, and optimized AP antenna gain. The black dots mark the available positions for placing the RIS, which guarantee that the minimum power received by the UE is at least (a) -10 dBm, (b) -7.5 dBm, and (c) -5 dBm. The UE is free to move within the 4×5 m^2 area designated by each rectangle, and the blue cross marks the position of the AP.

wall for the RIS placement, with a narrower available region. Finally, in Fig. 5.11(c), top row, the threshold is set to -5 dBm, leaving only one position on the right wall. With further increase in the threshold, no available positions are found; on the other hand, with decreasing the threshold below the values shown here, it is found that the

entire top and right walls become available for the RIS when the threshold drops below approximately -18 dBm. Similar conclusions can be drawn from the rest panels in Fig. 5.11, where the AP is placed in the middle of the left wall (second row), in the middle of the bottom wall (third row), and in the middle of the room (last row).

Non-Optimized AP Antenna gain

When the AP antenna gain is not tunable, the received power can be calculated with Eqs.(**??**-**??**), using the specific value of G_t. with the positions shown in Fig. 5.11(c), i.e. those found using a tunable antenna. For the configurations examined in Fig. 5.11, the optimal RIS placement under constant AP gain is found in several cases to agree with the positions shown in Fig. 5.11(c), i.e. those found using a tunable antenna. The thresholds, however, that determine the solutions are in general lower than those of the tunable antenna and depend strongly on the value of G_t. This means that, while the RIS position that maximizes the received power can always be located in the middle of the right or left wall, for example, the maximum received power will depend on the AP antenna gain.

Apparently, the received power cannot exceed the maximum power given by Eq.(5.21), which is achieved with the optimized gain. To estimate the power improvement achieved with optimized versus non-optimized AP antenna gain, the difference in the received power between the two cases is shown in Fig. 5.12 for the AP and RIS positions marked with the blue cross and black dot, respectively. Non-optimized gain refers to the constant (non-tunable) AP gain that consequently cannot satisfy the maximum power given by Eq.(5.21), and in the examples shown here is 40 dB [Fig. 5.12(a)], 50 dB [Fig. 5.12(b)] and 60 dB [Fig. 5.12(c)]. The results show that, overall, the use of tunable AP antenna is advantageous. In particular, compared to relatively wide beams (low G_t) as shown in Fig. 5.12(a), the power difference increases at UE locations close to the RIS and towards small θ_{UE} angles. Compared to beams of higher gain, the UE locations of increased power difference shrinks in the vicinity of the RIS [Fig. 5.12(b)], until for high gain (relatively narrow beams), the picture changes entirely and the received power difference increases as the UE moves away from the RIS [Fig. 5.12(c)]. Hence, from the perspective of the non-tunable AP antenna, these results provide guidance with respect to the choice for the constant AP gain: if the UE is mainly located relatively far from the RIS, the optimized performance can be approached with a relatively low AP gain [Fig. 5.12(a)], while in the opposite case where the UE is mainly located relatively close to the RIS, a relatively high AP gain is more suitable [Fig. 5.12(c)].

5.2.4   Conclusion

In this work, the optimal RIS placement was studied on the basis of the available AP gain, the available positions for the AP and RIS placement, and the minimum desired power levels at the UE. With the use of an analytical model the possibilities for both tunable and non-tunable AP antenna gain were examined in a mobile user environment. The optimal RIS placement for a fixed UE position was identified, whereas for the mobile UE case a multiplicity of RIS placement solutions was shown to exist. It was demonstrated that the optimal RIS placement relies on the AP antenna gain tunability. The results illustrate that increasing the AP antenna gain does not guarantee an
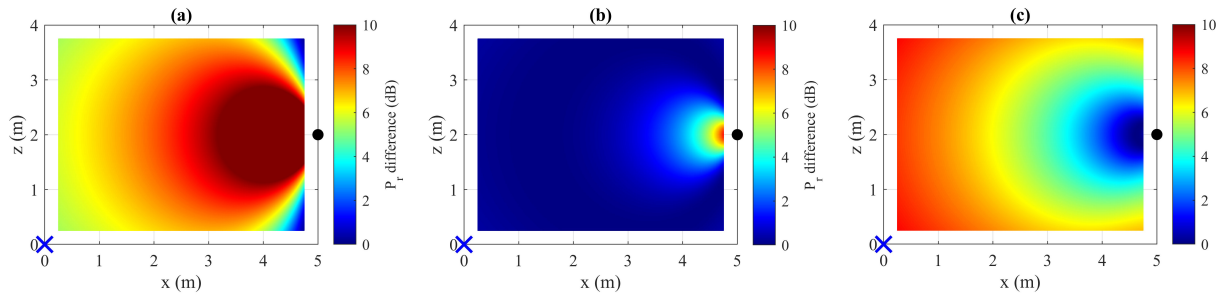
Figure 5.12: Difference in the received power between optimized and fixed AP antenna gain, the latter being (a) 40 dB, (b) 50 dB, and (c) 60 dB. The blue cross marks the position of the AP and the black dot marks the RIS position, which is chosen to implement the optimal placement shown in Fig. 5.11. In (a) the maximum power difference reaches 18 dB, however the color map has been saturated to 10 dB to emphasize the contrast with cases (b) and (c).

increase in received power and provide guidelines for the optimal selection of the AP antenna gain and the RIS placement.

Conclusions of D2.3

This deliverable appends on the results given in D2.1 by giving further analysis and beamforming algorithms for RIS-enabled communication links. Chapter 1 went through the various HW imperfections and their modelling. Chapter 2 analysed the achievable gain of the RIS as a function of the center frequency, size of the RIS, and distance. It was shown that the performance of the RIS in some situations can be very low, but also very high if the distance to and out of the RIS is small and the RIS is large. Chapter 3 considered the RPM scheme for the dual-use of passive beamforming and information transfer of the RIS. It was shown that the proposed RIS-RPM scheme was able to improve the achievable rate performance despite the loss in received signal power as compared to the conventional RIS-assisted system. Chapter 4 proposed a RIS quadrature reflection modulation (RIS-QRM) scheme for simultaneous passive beamforming and information transfer of RIS. The RIS-QRM was shown to be able to improve the error performance of the additional bits delivered by the RIS without deteriorating that of the bits carried on the constellation symbol. Finally, Chapter 5 looked into the optimal RIS placement based on the AP gain, the placement of the AP and RIS, and the minimum desired power levels at the UE. It was shown that the optimal RIS placement relies on the AP antenna gain tunability. The results illustrated that increasing the AP antenna gain does not guarantee an increase in received power and the chapter provided guidelines for the optimal selection of the AP antenna gain and the RIS placement.

Putting together all the results from D2.1 and D2.3, we have a very comprehensive look into directional communication links in Task 2.1. These results produced in T2.1 can be further utilized (and have been utilized) in the other work packages and tasks of ARIADNE. The most notable of those include the machine learning works in WP4 and system performance assessments in WP1. Especially the system performance assessments in T1.3 benefit from the works in these deliverables and gives the link level modelling herein a context within ARIADNE use cases and scenarios. The beamforming models with and without RIS, the HW imperfection models, as well as the network node placement analysis combined with link models from D1.2 give a very wide range of tools not only for the system level evaluations in T1.3, but also for WP4 for the machine learning algorithms. All these aspects are required for successful ARIADNE system level evaluations that determine the performance of the use cases and scenarios of interest for this project. Furthermore, even with the increasing research interests on +100 GHz communications and various aspects therein, the specific research problems at high frequency communications require a lot more research in and outside of the ARIADNE project. As a whole, the results presented herein and in ARIADNE project in general add to the body of knowledge and are very useful for various research groups and companies across Europe and the world.

Bibliography

[1] *Recommendation (18)01: Radio frequency channel/block arrangements for Fixed Service systems operating in the bands 130-134 GHz, 141-148.5 GHz, 151.5-164 GHz and 167-174.8 GHz*, ECC Std.

[2] ARIADNE, "D2.1 Initial results in Dband directional links analysis, system performance assessment, and algorithm designs," Tech. Rep., 2020. [Online]. Available: https://www.ict-ariadne.eu/deliverables/

[3] B. Razavi, *RF Microelectronics, 2nd edition*. Prentice Hall, 2012.

[4] T. Schenk, *RF Imperfections in High-Rate Wireless Systems.* The Netherlands: Springer, 2008.

[5] K. Kundert, *Accurate and Rapid Measurement of IP2 and IP3*. The Designer Guide Community, May 22, 2002.

[6] A. A. Saleh, "Frequency-independent and Frequency-dependent Nonlinear models of TWT Amplifiers," *IEEE Transactions on Communications*, vol. 29, no. 11, p. 1715–20, 1981.

[7] M. Honkanen and S.-G. Haggman, "New aspects on Nonlinear Power Amplifier modeling in Radio communication system simulations," in *Proc. IEEE Int. Symp. On Personal, Indoor, and Mobile Comm, PIMRC*, Helsinki, Finland, Sep.1-4, 1997, pp. 844–848.

[8] C.-S. Choi, Y. Shoji, H. Harada, R. Funada, S. Kat, K. Maruhashi, I. Toyoda, and K. Takahashi, "RF impairment models for 60GHz-band SYS/PHY simulation," *IEEE P802.15 WLAN*, 2006. [Online]. Available: https://mentor.ieee.org/802.15/file/06/15-06-0477-01-003c-rf-impairment-models-60ghz-band-sysphy-simulation.pdf

[9] S. Bicais and J.-B. Dore, "Phase Noise model selection for sub-THz communications," in *IEEE Global Communications Conference (GLOBECOM)*, 2019, pp. 1–6.

[10] M. Wu, D. Wuebben, A. Dekorsy, P. Baracca, V. Braun, and H. Halbauer, "Hardware impairments in millimeter wave communications using OFDM and SC-FDE," in *20th International ITG Workshop on Smart Antennas, WSA*, 2016, pp. 1–8.

[11] Y.-C. Liao and K.-C. Chen, "Estimation of stationary Phase Noise by the autocorrelation of the ICI weighting function in OFDM systems," *IEEE Transactions on Wireless Communications*, vol. 5, no. 12, p. 3370–3374, Dec. 2006.

[12] M. Valkama, M. Renfors, and V. Koivunen, "Advanced methods for I/Q imbalance compensation in communication receivers," *IEEE Transactions on Signal Processing*, vol. 49, no. 10, pp. 2335–2344, 2001.

[13] E. Cetin, I. Kale, and R. C.-S. Morling, "Living and dealing with RF impairments in communication transceivers," in *IEEE International Symposium on Circuits and Systems*, 2007, pp. 21–24.

[14] M. Valkama, M. Renfors, and V. Koivunen, "Compensation of frequency-selective I/Q imbalances in wideband receivers: models and algorithms," in *IEEE Third Workshop on Signal Processing Advances in Wireless Communications in SPAWC*, 2001, pp. 42–45.

[15] A.-A. A. Boulogeorgos, N. Chatzidiamantis, H. G. Sandalidis, A. Alexiou, and M. D. Renzo, "Cascaded composite turbulence and misalignment: Statistical characterization and applications to reconfigurable intelligent surface-empowered wireless systems," *arXiv 2106.15082*, 2021.

[16] M. Di Renzo, A. Zappone, M. Debbah, M. S. Alouini, C. Yuen, J. de Rosny, and S. Tretyakov, "Smart radio environments empowered by reconfigurable intelligent surfaces: How it works, state of research, and the road ahead," *IEEE Journal on Selected Areas in Communications*, vol. 38, no. 11, pp. 2450–2525, 2020.

[17] R. Schroeder, J. He, and M. Juntti, "Channel estimation for hybrid RIS aided MIMO communications via atomic norm minimization," *arXiv:2106.10909*, pp. 1–6, 2021.

[18] N. T. Nguyen, Q.-D. Vu, K. Lee, and M. Juntti, "Hybrid relay-reflecting intelligent surface-assisted wireless communication," *arXiv*, 2021. [Online]. Available: https://arxiv.org/abs/2103.03900

[19] W. Tang, M. Z. Chen, X. Chen, J. Y. Dai, Y. Han, M. Di Renzo, Y. Zeng, S. Jin, Q. Cheng, and T. J. Cui, "Wireless communications with reconfigurable intelligent surface: Path loss modeling and experimental measurement," *arXiv: 1911.05326*, pp. 1–32, 2020.

[20] E. Basar and I. Yildirim, "Indoor and outdoor physical channel modeling and efficient positioning for reconfigurable intelligent surfaces in mmWave bands," *arXiv: 2006.02240*, pp. 1–10, 2020.

[21] J. C. B. Garcia, A. Sibille, and M. Kamoun, "Reconfigurable intelligent surfaces: Bridging the gap between scattering and reflection," *IEEE J. Sel. Areas Commun.*, vol. 38, no. 11, pp. 2538–2547, 2020.

[22] W. Khawaja, O. Ozdemir, Y. Yapici, F. Erden, and I. Guvenc, "Coverage enhancement for NLOS mmWave links using passive reflectors," *IEEE Open J. Commun. Soc.*, vol. 1, pp. 263–281, 2020.

[23] O. Özdogan, E. Björnson, and E. G. Larsson, "Intelligent reflecting surfaces: Physics, propagation, and pathloss modeling," *arXiv:1911.03359*, pp. 1–5, 2019.

[24] F. H. Danufane, M. Di Renzo, J. de Rosny, and S. Tretyakov, "On the path-loss of reconfigurable intelligent surfaces: An approach based on Green's theorem applied to vector fields," *arXiv: 2007.13158*, pp. 1–30, 2020.

[25] M. Di Renzo, F. Habibi Danufane, X. Xi, J. de Rosny, and S. Tretyakov, "Analytical modeling of the path-loss for reconfigurable intelligent surfaces anomalous mirror or scatterer ?" in *IEEE SPAWC*, 2020, pp. 1–5.

[26] A.-A. A. Boulogeorgos and A. Alexiou, "Coverage analysis of reconfigurable intelligent surface assisted THz wireless systems," *IEEE Open J. Vehic. Technol.*, vol. 2, pp. 94–110, 2021.

[27] ——, "Pathloss modeling of reconfigurable intelligent surface assisted THz wireless systems," *arXiv: 2102.08757*, pp. 1–6, 2021.

[28] J. Kokkoniemi and M. Juntti, "Stochastic geometry based interference analysis of multiuser mmWave networks with RIS," in *Proc. IEEE PIMRC*, Sept. 2021, pp. 1–6.

[29] K. Ntontin, A.-A. A. Boulogeorgos, D. Selimis, F. Lazarakis, A. Alexiou, and S. Chatzinotas, "Reconfigurable intelligent surface optimal placement in millimeter-wave networks," *arXiv:2011.09949*, pp. 1–14, 2020. [Online]. Available: https://arxiv.org/abs/2011.09949

[30] J. M. Jornet and I. F. Akyildiz, "Channel modeling and capacity analysis for electromagnetic nanonetworks in the terahertz band," *IEEE Trans. Wireless Commun.*, vol. 10, no. 10, pp. 3211–3221, Oct. 2011.

[31] F. Boccardi, R. W. Heath, A. Lozano, T. L. Marzetta, and P. Popovski, "Five disruptive technology directions for 5G," *IEEE Commun. Mag.*, vol. 52, no. 2, pp. 74–80, Feb. 2014.

[32] F. Sohrabi and W. Yu, "Hybrid digital and analog beamforming design for large-scale antenna arrays," *IEEE J. Sel. Topics Signal Process.*, vol. 10, no. 3, pp. 501–513, Apr. 2016.

[33] H. Q. Ngo, E. Larsson, and T. Marzetta, "Energy and spectral efficiency of very large multiuser MIMO systems," *IEEE Trans. Commun.*, vol. 61, no. 4, pp. 1436–1449, Apr. 2013.

[34] F. Liu, A. Pitilakis, M. S. Mirmoosa, O. Tsilipakos, X. Wang, A. C. Tasolamprou, S. Abadal, A. Cabellos-Aparicio, E. Alarcón, C. Liaskos, N. V. Kantartzis, M. Kafesaki, E. N. Economou, C. M. Soukoulis, and S. Tretyakov, "Programmable metasurfaces: State of the art and prospects," in *Proc. IEEE ISCAS*, Florence, Italy, May 2018, pp. 1–5.

[35] T. J. Cui, M. Q. Qi, X. Wan, J. Zhao, and Q. Cheng, "Coding metamaterials, digital metamaterials and programmable metamaterials," *L. Sci. & Appl.*, vol. 3, no. 10, pp. e218, Oct. 2014.

[36] M. Di Renzo, A. Zappone, M. Debbah, M.-S. Alouini, C. Yuen, J. de Rosny, and S. Tretyakov, "Smart radio environments empowered by reconfigurable intelligent surfaces: How it works, state of research, and road ahead," *arXiv preprint arXiv:2004.09352*, 2020.

[37] X. Tan, Z. Sun, D. Koutsonikolas, and J. M. Jornet, "Enabling indoor mobile millimeter-wave networks based on smart reflect-arrays," in *Proc. IEEE INFOCOM*, Hawaii, USA, Apr. 2018, pp. 270–278.

[38] C. Liaskos, S. Nie, A. Tsioliaridou, A. Pitsillides, S. Ioannidis, and I. Akyildiz, "A new wireless communication paradigm through software-controlled metasurfaces," *IEEE Commun. Mag.*, vol. 56, no. 9, pp. 162–169, Sep. 2018.

[39] E. Basar, M. Di Renzo, J. De Rosny, M. Debbah, M. Alouini, and R. Zhang, "Wireless communications through reconfigurable intelligent surfaces," *IEEE Access*, vol. 7, pp. 116 753–116 773, 2019.

[40] C. Huang, S. Hu, G. C. Alexandropoulos, A. Zappone, C. Yuen, R. Zhang, M. Di Renzo, and M. Debbah, "Holographic MIMO surfaces for 6G wireless networks: Opportunities, challenges, and trends," *IEEE Wireless Commun.*, to appear, 2020.

[41] M. Di Renzo and J. Song, "Reflection probability in wireless networks with metasurface-coated environmental objects: An approach based on random spatial processes," *EURASIP J. Wireless Commun. Netw.*, no. 99, Apr. 2019.

[42] N. S. Perović, M. Di Renzo, and M. F. Flanagan, "Channel capacity optimization using reconfigurable intelligent surfaces in indoor mmWave environments," *arXiv preprint arXiv:1910.14310*, 2019.

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

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

[45] C. Huang, G. C. Alexandropoulos, A. Zappone, M. Debbah, and C. Yuen, "Energy efficient multi-user MISO communication using low resolution large intelligent surfaces," in *Proc. IEEE GLOBECOM*, Abu Dhabi, UAE, Dec. 2018, pp. 1–6.

[46] E. Björnson, Ö. Özdogan, and E. G. Larsson, "Intelligent reflecting surface versus decode-and-forward: How large surfaces are needed to beat relaying?" *IEEE Wireless Commun. Lett.*, vol. 9, no. 2, pp. 244–248, Feb. 2020.

[47] W. Tang, M. Z. Chen, X. Chen, J. Y. Dai, Y. Han, M. Di Renzo, Y. Zeng, S. Jin, Q. Cheng, and T. J. Cui, "Wireless communications with reconfigurable intelligent surface: Path loss modeling and experimental measurement," *arXiv preprint arXiv:1911.05326*, 2019.

[48] Q. Wu and R. Zhang, "Intelligent reflecting surface enhanced wireless network: Joint active and passive beamforming design," in *Proc. IEEE GLOBECOM*, Abu Dhabi, United Arab Emirates, Dec. 2018, pp. 1–6.

[49] A. Taha, M. Alrabeiah, and A. Alkhateeb, "Enabling large intelligent surfaces with compressive sensing and deep learning," *arXiv preprint arXiv:1904.10136*, 2019.

[50] G. C. Alexandropoulos and E. Vlachos, "A hardware architecture for reconfigurable intelligent surfaces with minimal active elements for explicit channel estimation," in *Proc. IEEE ICASSP*, Barcelona, Spain, May 2020, pp. 9175–9179.

[51] C. Huang, G. C. Alexandropoulos, C. Yuen, and M. Debbah, "Indoor signal focusing with deep learning designed reconfigurable intelligent surfaces," in *Proc. IEEE SPAWC*, Cannes, France, Jul. 2019, pp. 1–5.

[52] D. Mishra and H. Johansson, "Channel estimation and low-complexity beamforming design for passive intelligent surface assisted MISO wireless energy transfer," in *Proc. IEEE ICASSP*, Brighton, United Kingdom, May 2019, pp. 4659–4663.

[53] Z. He and X. Yuan, "Cascaded channel estimation for large intelligent metasurface assisted massive MIMO," *IEEE Wireless Commun. Lett.*, vol. 9, no. 2, pp. 210–214, Feb. 2020.

[54] Y. Yang, B. Zheng, S. Zhang, and R. Zhang, "Intelligent reflecting surface meets OFDM: Protocol design and rate maximization," *IEEE Trans. Commun.*, early access, 2020.

[55] B. Zheng and R. Zhang, "Intelligent reflecting surface-enhanced OFDM: Channel estimation and reflection optimization," *IEEE Wireless Commun. Lett.*, vol. 9, no. 4, pp. 518–522, Apr. 2020.

[56] C. You, B. Zheng, and R. Zhang, "Intelligent reflecting surface with discrete phase shifts: Channel estimation and passive beamforming," *arXiv preprint arXiv:1911.03916*, 2019.

[57] T. L. Jensen and E. De Carvalho, "An optimal channel estimation scheme for intelligent reflecting surfaces based on a minimum variance unbiased estimator," in *Proc. IEEE ICASSP*, Barcelona, Spain, May 2020, pp. 5000–5004.

[58] L. Wei, C. Huang, G. C. Alexandropoulos, and C. Yuen, "Parallel factor decomposition channel estimation in RIS-assisted multi-user MISO communication," in *Proc. IEEE SAM*, Hangzhou, China, Jun. 2020.

[59] B. Zheng, C. You, and R. Zhang, "Intelligent reflecting surface assisted multi-user OFDMA: Channel estimation and training design," *arXiv preprint arXiv:2003.00648*, 2020.

[60] C. Huang, R. Mo, and C. Yuen, "Reconfigurable intelligent surface assisted multiuser MISO systems exploiting deep reinforcement learning," *IEEE J. Select. Areas Commun.*, early access, 2020.

[61] E. Basar, "Large intelligent surface-based index modulation: A new beyond MIMO paradigm for 6G," *IEEE Trans. Commun.*, early access, 2020.

[62] R. Karasik, O. Simeone, M. Di Renzo, and S. Shamai, "Beyond max-SNR: Joint encoding for reconfigurable intelligent surfaces," *arXiv preprint arXiv:1911.09443*, 2019.

[63] M. Di Renzo, M. Debbah, D.-T. Phan-Huy, A. Zappone, M.-S. Alouini, C. Yuen, V. Sciancalepore, G. C. Alexandropoulos, J. Hoydis, H. Gacanin, J. de Rosny, A. Bounceur, G. Lerosey, and M. Fink, "Smart radio environments empowered by reconfigurable AI meta-surfaces: An idea whose time has come," *EURASIP J. Wireless Commun. Netw.*, vol. 2019, no. 1, May 2019.

[64] X. Yuan, Y.-J. Zhang, Y. Shi, W. Yan, and H. Liu, "Reconfigurable-intelligent-surface empowered 6G wireless communications: Challenges and opportunities," *arXiv preprint arXiv:2001.00364*, 2020.

[65] W. Yan, X. Yuan, and X. Kuai, "Passive beamforming and information transfer via large intelligent surface," *IEEE Wireless Commun. Lett.*, vol. 9, no. 4, pp. 533–537, Apr. 2020.

[66] W. Yan, X. Yuan, Z.-Q. He, and X. Kuai, "Passive beamforming and information transfer design for reconfigurable intelligent surfaces aided multiuser MIMO systems," *arXiv preprint arXiv:1912.10209*, 2019.

[67] H. Yang, X. Chen, F. Yang, S. Xu, X. Cao, M. Li, and J. Gao, "Design of resistor-loaded reflectarray elements for both amplitude and phase control," *IEEE Antennas Wireless Propag. Lett.*, vol. 16, pp. 1159–1162, Nov. 2017.

[68] A. Man-Cho, J. Zhang, and Y. Ye, "On approximating complex quadratic optimization problems via semidefinite programming relaxations," *Mathematical Programming*, vol. 110, no. 1, pp. 93–110, Jun. 2007.

[69] M. Grant and S. Boyd, "CVX: Matlab software for disciplined convex programming," 2016. [Online]. Available: http://cvxr.com/cvx.

[70] Z.-Q. Luo, W.-K. Ma, A. M.-C. So, Y. Ye, and S. Zhang, "Semidefinite relaxation of quadratic optimization problems," *IEEE Signal Process. Mag.*, vol. 27, no. 3, pp. 20–34, May 2010.

[71] M. Di Renzo, K. Ntontin, J. Song, F. H. Danufane, X. Qian, F. Lazarakis, J. De Rosny, D. Phan-Huy, O. Simeone, R. Zhang, M. Debbah, G. Lerosey, M. Fink, S. Tretyakov, and S. Shamai, "Reconfigurable intelligent surfaces vs. relaying: Differences, similarities, and performance comparison," *IEEE Open J. Commun. Society*, vol. 1, pp. 798–807, Jun. 2020.

[72] X. Qian, M. Di Renzo, J. Liu, A. Kammoun, and M.-S. Alouini, "Beamforming through reconfigurable intelligent surfaces in single-user MIMO systems: SNR distribution and scaling laws in the presence of channel fading and phase noise," *IEEE Wireless Commun. Lett.*, vol. 10, no. 1, pp. 77–81, Jan. 2021.

[73] Q.-U.-A. Nadeem, A. Kammoun, M. D. Anas Chaaban, and M.-S. Alouini, "Asymptotic max-min SINR analysis of reconfigurable intelligent surface assisted MISO systems," *IEEE Trans. Wireless Commun.*, vol. 19, no. 12, pp. 7748–7764, 2020.

[74] B. Zheng, C. You, and R. Zhang, "Double-IRS assisted multi-user MIMO: Cooperative passive beamforming design," *IEEE Trans. Wireless Commun.*, vol. 20, no. 7, pp. 4513–4526, Jul. 2021.

[75] W. Tang, J. Y. Dai, M. Chen, X. Li, Q. Cheng, S. Jin, K.-K. Wong, and T. J. Cui, "Programmable metasurface-based RF chain-free 8PSK wireless transmitter," *Electron. Lett.*, vol. 55, no. 7, pp. 417–420, Apr. 2019.

[76] W. Tang, M. Z. Chen, J. Y. Dai, Y. Zeng, X. Zhao, S. Jin, Q. Cheng, and T. J. Cui, "Wireless communications with programmable metasurface: New paradigms, opportunities, and challenges on transceiver design," *IEEE Wireless Commun.*, vol. 27, no. 2, pp. 180–187, Apr. 2020.

[77] B. Zheng, Q. Wu, and R. Zhang, "Intelligent reflecting surface-assisted multiple access with user pairing: NOMA or OMA?" *IEEE Wireless Commun. Lett.*, vol. 24, no. 4, pp. 753–757, Apr. 2020.

[78] S. Lin, B. Zheng, G. C. Alexandropoulos, M. Wen, F. Chen, and S. Mumtaz, "Adaptive transmission for reconfigurable intelligent surface-assisted OFDM wireless communications," *IEEE J. Select. Areas Commun.*, vol. 38, no. 11, pp. 2653–2665, Nov. 2020.

[79] B. Zheng, C. You, and R. Zhang, "Efficient channel estimation for double-IRS aided multi-user MIMO system," *IEEE Trans. Commun.*, vol. 69, no. 6, pp. 3818–3832, Jun. 2021.

[80] M. Wen, B. Zheng, K. J. Kim, M. Di Renzo, T. A. Tsiftsis, K. Chen, and N. Al-Dhahir, "A survey on spatial modulation in emerging wireless systems: Research progresses and applications," *IEEE J. Select. Areas Commun.*, vol. 37, no. 9, pp. 1949–1972, Sep. 2019.

[81] M. Di Renzo, H. Haas, A. Ghrayeb, S. Sugiura, and L. Hanzo, "Spatial modulation for generalized MIMO: Challenges, opportunities, and implementation," *Proc. of the IEEE*, vol. 102, no. 1, pp. 56–103, Dec. 2014.

[82] R. Mesleh, S. S. Ikki, and H. M. Aggoune, "Quadrature spatial modulation," *IEEE Trans. Veh. Technol.*, vol. 64, no. 6, pp. 2738–2742, Jun. 2015.

[83] B. Zheng, F. Chen, M. Wen, F. Ji, H. Yu, and Y. Liu, "Low-complexity ML detector and performance analysis for OFDM with in-phase/quadrature index modulation," *IEEE Commun. Lett.*, vol. 19, no. 11, pp. 1893–1896, Nov. 2015.

[84] ——, "Single-RF multi-user communication through reconfigurable intelligent surfaces: An information-theoretic analysis," *arXiv preprint arXiv:2101.07556*, 2021.

[85] ——, "Adaptive coding and channel shaping through reconfigurable intelligent surfaces: An information-theoretic analysis," *IEEE Trans. Commun.*, early access, 2021.

[86] W. Yan, X. Yuan, and X. Kuai, "Passive beamforming and information transfer via large intelligent surface," *IEEE Wireless Commun. Lett.*, vol. 9, no. 4, pp. 533–537, Apr. 2020.

[87] S. Lin, B. Zheng, G. C. Alexandropoulos, M. Wen, M. Di Renzo, and F. Chen, "Reconfigurable intelligent surfaces with reflection pattern modulation: Beamforming design and performance analysis," *IEEE Trans. Wireless Commun.*, vol. 20, no. 2, pp. 741–754, Feb. 2021.

[88] S. Guo, S. Lv, H. Zhang, J. Ye, and P. Zhang, "Reflecting modulation," *IEEE J. Select. Areas Commun.*, vol. 38, no. 11, pp. 2548–2561, Nov. 2020.

[89] R. W. Heath, M. Kountouris, and T. Bai, "Modeling heterogeneous network interference using poisson point processes," *IEEE Trans. Signal Process.*, vol. 61, no. 16, pp. 4114–4126, Aug. 2013.

[90] I. S. Gradshteyn and I. M. Ryzhik, *Table of integrals, series, and products (seventh edition)*. Academic press, 2007.

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

[92] *Further advancements for E-UTRA physical layer aspects (Release 9)*, 3GPP TS 36.814, Mar. 2010.

[93] E. N. Papasotiriou, A.-A. A. Boulogeorgos, A. Stratakou, and A. Alexiou, "Performance evaluation of reconfigurable intelligent surface assisted d-band wireless communication," in *IEEE 3rd 5G World Forum (5GWF)*. IEEE, Sep 2020.

[94] A.-A. A. Boulogeorgos, A. Alexiou, T. Merkle, C. Schubert, R. Elschner, A. Katsiotis, P. Stavrianos, D. Kritharidis, P.-K. Chartsias, J. Kokkoniemi, M. Juntti, J. Lehtomaki, A. Teixeira, and F. Rodrigues, "Terahertz technologies to deliver optical network quality of experience in wireless systems beyond 5G," *IEEE Commun. Mag.*, vol. 56, no. 6, pp. 144–151, Jun. 2018.

[95] E. Basar, M. D. Renzo, J. D. Rosny, M. Debbah, M.-S. Alouini, and R. Zhang, "Wireless communications through reconfigurable intelligent surfaces," *IEEE Access*, vol. 7, pp. 116 753–116 773, 2019.

[96] M. Francesco Foglia, A. Clemente, and J. L. Gonzalez-Jimenez, "High-gain d-band transmitarrays in standard PCB technology for beyond-5g communications," *IEEE Transactions on Antennas and Propagation*, vol. 68, no. 1, pp. 587–592, Jan 2020.

[97] G. R. MacCartney, S. Deng, S. Sun, and T. S. Rappaport, "Millimeter-wave human blockage at 73 ghz with a simple double knife-edge diffraction model and extension for directional antennas," in *2016 IEEE 84th Vehicular Technology Conference (VTC-Fall)*, 2016, pp. 1–6.

[98] G. Stratidakis, E. N. Papasotiriou, H. Konstantinis, A.-A. A. Boulogeorgos, and A. Alexiou, "Relay-based blockage and antenna misalignment mitigation in THz wireless communications," in *2020 2nd 6G Wireless Summit (6G SUMMIT)*. IEEE, Mar 2020.

[99] M. Jacob, S. Priebe, R. Dickhoff, T. Kleine-Ostmann, T. Schrader, and T. Kürner, "Diffraction in mm and sub-mm wave indoor propagation channels," *IEEE Trans. Microw. Theory Techn.*, vol. 60, no. 3, pp. 833–844, Mar. 2012.

[100] S. Abeywickrama, R. Zhang, Q. Wu, and C. Yuen, "Intelligent reflecting surface: Practical phase shift model and beamforming optimization," *IEEE Transactions on Communications*, vol. 68, no. 9, pp. 5849–5863, 2020.

[101] I. Yildirim, A. Uyrus, and E. Basar, "Modeling and analysis of reconfigurable intelligent surfaces for indoor and outdoor applications in future wireless networks," *IEEE Transactions on Communications*, vol. 69, no. 2, pp. 1290–1301, 2021.

[102] N. S. Perović, M. D. Renzo, and M. F. Flanagan, "Channel capacity optimization using reconfigurable intelligent surfaces in indoor mmwave environments," in *ICC 2020 - 2020 IEEE International Conference on Communications (ICC)*, 2020, pp. 1–7.

[103] C. Huang, G. C. Alexandropoulos, C. Yuen, and M. Debbah, "Indoor signal focusing with deep learning designed reconfigurable intelligent surfaces," in *2019 IEEE 20th International Workshop on Signal Processing Advances in Wireless Communications (SPAWC)*, 2019, pp. 1–5.

[104] Q. Wu, S. Zhang, B. Zheng, C. You, and R. Zhang, "Intelligent reflecting surface-aided wireless communications: A tutorial," *IEEE Transactions on Communications*, vol. 69, no. 5, pp. 3313–3351, 2021.

[105] O. Ozdogan, E. Bjornson, and E. G. Larsson, "Intelligent reflecting surfaces: Physics, propagation, and pathloss modeling," *IEEE Wireless Communications Letters*, vol. 9, no. 5, pp. 581–585, May 2020.

[106] W. Tang, M. Z. Chen, X. Chen, J. Y. Dai, Y. Han, M. D. Renzo, Y. Zeng, S. Jin, Q. Cheng, and T. J. Cui, "Wireless communications with reconfigurable intelligent surface: Path loss modeling and experimental measurement," *IEEE Transactions on Wireless Communications*, vol. 20, no. 1, pp. 421–439, Jan 2021.

[107] M. D. Renzo, K. Ntontin, J. Song, F. H. Danufane, X. Qian, F. Lazarakis, J. D. Rosny, D.-T. Phan-Huy, O. Simeone, R. Zhang, M. Debbah, G. Lerosey, M. Fink, S. Tretyakov, and S. Shamai, "Reconfigurable intelligent surfaces vs. relaying: Differences, similarities, and performance comparison," *IEEE Open Journal of the Communications Society*, vol. 1, pp. 798–807, 2020.

[108] Y. Han, W. Tang, S. Jin, C.-K. Wen, and X. Ma, "Large intelligent surface-assisted wireless communication exploiting statistical CSI," *IEEE Transactions on Vehicular Technology*, vol. 68, no. 8, pp. 8238–8242, Aug 2019.

[109] 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.

[110] K. Ntontin, A.-A. A. Boulogeorgos, D. G. Selimis, F. I. Lazarakis, A. Alexiou, and S. Chatzinotas, "Reconfigurable intelligent surface optimal placement in millimeter-wave networks," *IEEE Open Journal of the Communications Society*, vol. 2, pp. 704–718, 2021.

[111] K. Ntontin, D. Selimis, A.-A. A. Boulogeorgos, A. Alexandridis, A. Tsolis, V. Vlachodimitropoulos, and F. Lazarakis, "Optimal reconfigurable intelligent surface placement in millimeter-wave communications," in *2021 15th European Conference on Antennas and Propagation (EuCAP)*. IEEE, Mar 2021.

[112] Y. Pan, K. Wang, C. Pan, H. Zhu, and J. Wang, "Sum rate maximization for intelligent reflecting surface assisted terahertz communications," *CoRR*, vol. abs/2008.12246, 2020. [Online]. Available: https://arxiv.org/abs/2008.12246

[113] A. Zappone, M. D. Renzo, X. Xi, and M. Debbah, "On the optimal number of reflecting elements for reconfigurable intelligent surfaces," *IEEE Wireless Communications Letters*, pp. 1–1, 2020.

[114] H. Yu, H. D. Tuan, A. A. Nasir, T. Q. Duong, and H. V. Poor, "Joint design of reconfigurable intelligent surfaces and transmit beamforming under proper and improper gaussian signaling," *IEEE Journal on Selected Areas in Communications*, vol. 38, no. 11, pp. 2589–2603, Nov 2020.

[115] S. Zeng, H. Zhang, B. Di, Z. Han, and L. Song, "Reconfigurable intelligent surface (RIS) assisted wireless coverage extension: RIS orientation and location optimization," *IEEE Communications Letters*, pp. 1–1, 2020.

[116] X. Qian, M. Di Renzo, J. Liu, A. Kammoun, and M.-S. Alouini, "Beamforming through reconfigurable intelligent surfaces in single-user mimo systems: Snr distribution and scaling laws in the presence of channel fading and phase noise," *IEEE Wireless Communications Letters*, vol. 10, no. 1, pp. 77–81, 2021.

[117] L. Subrt and P. Pechac, "Controlling propagation environments using intelligent walls," in *2012 6th European Conference on Antennas and Propagation (EuCAP)*, 2012, pp. 1–5.

[118] Q. Wu and R. Zhang, "Beamforming optimization for wireless network aided by intelligent reflecting surface with discrete phase shifts," *IEEE Transactions on Communications*, vol. 68, pp. 1838–1851, 2020.

[119] A. Kazemipour, M. Hudlička, R. Dickhoff, M. Salhi, T. Kleine-Ostmann, and T. Schrader, "The horn antenna as gaussian source in the mm-wave domain," *Journal of Infrared, Millimeter, and Terahertz Waves*, vol. 35, no. 9, pp. 720–731, Jun 2014.

[120] Y. Hadad and T. Melamed, "Non-orthogonal domain parabolic equation and its tilted gaussian beam solutions," *IEEE Transactions on Antennas and Propagation*, vol. 58, no. 4, pp. 1164–1172, 2010.

[121] H. Lu, Y. Zeng, S. Jin, and R. Zhang, "Enabling panoramic full-angle reflection via aerial intelligent reflecting surface," in *2020 IEEE International Conference on Communications Workshops (ICC Workshops)*, 2020, pp. 1–6.

[122] G. Ghatak, "On the placement of intelligent surfaces for rssi-based ranging in mm-wave networks," *IEEE Communications Letters*, vol. 25, no. 6, pp. 2043–2047, 2021.

[123] H. Hashida, Y. Kawamoto, and N. Kato, "Intelligent reflecting surface placement optimization in air-ground communication networks toward 6g," *IEEE Wireless Communications*, vol. 27, no. 6, pp. 146–151, 2020.

[124] C. You, B. Zheng, and R. Zhang, "How to deploy intelligent reflecting surfaces in wireless network: Bs-side, user-side, or both sides?" 2021.

[125] S. Zhang and R. Zhang, "Intelligent reflecting surface aided multiple access: Capacity region and deployment strategy," in *2020 IEEE 21st International Workshop on Signal Processing Advances in Wireless Communications (SPAWC)*, 2020, pp. 1–5.

[126] X. Mu, Y. Liu, L. Guo, J. Lin, and R. Schober, "Joint deployment and multiple access design for intelligent reflecting surface assisted networks," *IEEE Transactions on Wireless Communications*, pp. 1–1, 2021.

[127] G. Stratidakis, S. Droulias, and A. Alexiou, "Analytical performance assessment of beamforming efficiency in reconfigurable intelligent surface-aided links," *IEEE Access*, vol. 9, pp. 115 922–115 931, 2021.