Channel Estimation and User Association for Distributed Intelligent Reflecting Surfaces Assisted MISO Communications

by

Hibatallah Asem M.N Alwazani

B.Sc. Hons., Effat University, 2019

A THESIS SUBMITTED IN PARTIAL FULFILLMENT OF
THE REQUIREMENTS FOR THE DEGREE OF

MASTER OF APPLIED SCIENCE

in

THE COLLEGE OF GRADUATE STUDIES

(Electrical Engineering)

THE UNIVERSITY OF BRITISH COLUMBIA

(Okanagan)

March 2021

The following individuals certify that they have read, and recommend to the College of Graduate Studies for acceptance, a thesis/dissertation entitled:

Channel Estimation and User Association for Distributed Intelligent Reflecting Surfaces Assisted MISO Communications

submitted by Hibatallah Asem M.N Alwazani in partial fulfilment of the requirements of the degree of Master of Applied Science

Prof. Anas Chaaban, School of Engineering
Supervisor

Prof. Hossain Jahangir, School of Engineering
Supervisory Committee Member

Prof. Loic Markley, School of Engineering
Supervisory Committee Member

Prof. Abbas Milani, School of Engineering
University Examiner

Abstract

Distributed intelligent reflecting surfaces (IRS)s deployed in wireless systems promise improved system performance, while spurring on diverse challenges such as channel estimation (CE), average signal-to-interference and noise (SINR) analysis, and IRS-user association. This is due to the passive nature of the reflecting elements and large multi-user system dimensions. In light of these challenges, we undertake the CE problem for the distributed IRSs-assisted multi-user MISO system. An optimal CE protocol requiring relatively low training overhead is developed using Bayesian techniques under the practical assumption that the base-station (BS)-IRSs channels are dominated by the line-of-sight (LoS) components. Simulation results corroborate the normalized MSE (NMSE) analysis and establish the advantage of the proposed protocol as compared to a benchmark scheme in terms of training overhead. In addition, we derive the average SINR for the distributed IRSs system with perfect channel state information (CSI) using a sub-optimal IRS reflecting configuration. After that, a successive refinement method is developed to find IRS-user association for the formulated max-min SINR problem which motivates user-fairness. Simulations validate the average SINR analysis while confirming the superiority of deploying a distributed IRSs scheme as well as an optimized IRS-user association as opposed to a centralized IRS deployment and random assignment.

Lay Summary

For a traditional wireless communication system, the transmitter sends information carried by electromagnetic (EM) waves across the radio environment, in free-space, to a particular receiver. The transmitter and receiver are controlled and optimized for this transmission, while the radio environment is largely random and uncontrolled. To meet the ever-increasing demands of communications, engineers converged on the idea of deploying distributed intelligent reflecting surfaces (IRS)s in the radio environment as a way to optimize the transmission over it. What the distributed IRSs do essentially is add more degrees of freedom into the wireless system presented as re-configurable entities that enable the radio environment to be smart. We focus on the updated wireless system which includes the distributed IRSs, then we provide the theoretical foundations for this distributed IRSs assisted wireless system. Moreover, we operate the distributed IRSs by pairing each IRS with a receiver then re-configuring that IRS to steer the information signals to the intended receiver. Finally, we formulate a conclusion on the usage of distributed IRSs in the wireless system based on the simulation results.

Preface

Prof. Anas and Dr. Qurrat Ul Ain (Annie) supervised this research. Moreover, both have reviewed the thesis and validated the theoretical derivations and numerical results. They have also authored with me the following publications. This thesis extends the system model defined in Chapter 2 from the journal paper titled (Intelligent Reflecting Surface-Assisted Multi-User MISO Communication: Channel Estimation and Beamforming Design) and published in 2020 IEEE Open Journal of the Communication Society (OJCOM) [1]. Chapter 3 of this thesis is in the conference paper titled (Channel Estimation for Distributed Intelligent Reflecting Surfaces Assisted Multi-User MISO Systems) and published during 2020 in IEEE GLOBECOM [2]. A journal paper version of Chapter 4 along with some of the other chapters in the thesis is prepared for submission at a prestigious IEEE journal such as IEEE Transactions in Wireless Communication (TWC).

Table of Contents

List of Tables

List of Figures

List of Notation

$(\cdot)^H$	the Hermetian of a complex vector or matrix.
$(\cdot)^\dagger$	the pseudo inverse of a matrix.
$\angle\cdot$	phase of a complex number for vectors it is taken element-wise.
$\langle\cdot,\cdot\rangle$	the inner product of two vectors.
$\mathbb{E}[\cdot]$	expected value of random variables.
$\mathbf{X}$	a matrix is written in bold upper-case letters or symbols.
$\mathbf{x}$	a vector is written in bold lower-case letters or symbols.
$\odot$	Hadamard product.
$\otimes$	Kronecker product.
$tr(\cdot)$	trace of a matrix.
$x \sim \mathcal{CN}(\mu, \sigma^2)$	x is a circularly symmetric complex Gaussian with mean μ and variance σ^2.
x^*	the conjugate of a complex number x.
$\lvert\cdot\rvert$	the absolute value of a number, for vectors it is taken element-wise.
$\lvert\lvert\cdot\rvert\rvert$	the Euclidean norm of a vector.
$\mathbf{x} \in \mathbb{C}^{M\times 1}$	$\mathbf{x}$ is a complex M-dimensional vector.
$\mathrm{diag}(\cdot)$	creates a diagonal matrix out of a vector.
$\mathrm{Var}(\cdot)$	variance of a random variable.
$\mathrm{vec}(\cdot)$	creates a vector out of a diagonal matrix.

List of Abbreviations

AoA	Angle-of-Arrival.
AoD	Angle-of-Departure.
BS	Base-station.
CE	Channel estimation.
CSI	Channel state information.
DFT	Discrete Fourier transform.
EM	Electromagnetic.
IRS	Intelligent reflecting surfaces.
LoS	Line-of-sight.
LS	Least squares.
MIMO	Multiple-input multiple-output.
MISO	Multiple-input single-output.
MMSE	Minimum mean square error.
MRT	Maximum Ratio Transmission.
NLoS	Non line-of-sight.
NMSE	Normalized mean square error.

SINR Signal-to-interference and noise ratio.
SNR Signal-to-noise ratio.

TDD Time-division duplexing.

ULA Uniform linear array.
UPA Uniform planar array.

Acknowledgements

To Prof. Anas and Dr. Annie, each of your presence has made this research roller-coaster full-filling, meaningful, and forever sketched into memory. Not only your incomparable guidance and your incredible insight, but also your understanding and patient natures in the transfer and teaching of this knowledge proved invaluable to a student stepping foot into the once esoteric research community. With such a prime view into what it is to be excellent researchers, dedicated teachers, and patient supervisors, my list of role models has definitely been updated.

To Prof. Jahangir, Prof. Markley, and Prof. Milani, I am doubly grateful to you as committee members and course instructors, where key concepts taught to me during the past year were crucial for building my research foundations. I am also thankful for your time and constructive feedback.

I am thankful to the wonderful friends I gained here at UBCO who have helped me during my research, special thanks to Bayan Alnahhas for first introducing this exciting subject to me. Also huge thanks to Sarah Bahanshal, my dearest friend and roommate who patiently listened to my research problems and always cheered me up. To Kelowna ducks everywhere, whenever I felt down I went to Okanagan lake and your presence made me feel lighter and put things into perspective, thank you.

I am extremely thankful to my mom whose prayers are the reason I approached this capacity in knowledge. All such strides former and future, I channel to you. I am also grateful towards uncle Hussain, whose support cannot be measured. Finally, I am thankful to my sisters and brothers who were always there for me and have given me motivation to pursue higher studies.

Dedication

To my little niece and nephews, I dedicate this thesis to you.

Chapter 1

Introduction

There are two uncannily accurate and related trends that trace out the meteoric pace of technological evolution in the past decades: Moore's law and Cooper's law. The former, extrapolated by inventor and engineer Gordon Moore in 1965, is related to the doubling of transistors per chip every two years, while the latter, discovered by cell-phone inventor Martin Cooper is about the doubling of data rates per link every two years. Recently, Moore's law, an exponentially increasing trend, has hit a plateau sending shock waves across the technological world [3]. Since these laws are dependent to some extent, communication engineers must discover cost-effective ways to prolong the inevitable plateau of Cooper's law, each generation standard in cellular communication has managed to stay on track with the exponentially increasing trend and increase data rates by introducing new key technologies or by densifying old technologies. The next sections introduce a novel idea tossed around in beyond 5G and towards 6G discussions to tackle the increasing data rate demand by smartly re-configuring radio environments.

1.1 Beyond 5G and towards 6G

With 5G underway, significant technologies such as massive multiple-input multiple-output (MIMO) and milimeter (mm)Wave have been steadily commercialized. Inherent limitations within these technologies are being brought to the fore [4], one of which is the inability to control the radio environment while being energy-cognizant [5]. These limitations are spurring on the development of the next generation 6G wireless system which is expected to be a convergence of several technology trends including that of smart environments and smart surfaces examined below.

1.2 Smart Radio Environments

The goal of future wireless systems is not only speed and connectivity as it was before. In fact, future wireless systems achieve several objectives to serve several information and communication technology applications like computing, sensing, and distributed intelligent communications. Two fundamental limitations exist in current communication standards [6]:

- High power consumption of wireless systems. With the expected addition of millions of devices, the resultant power consumption will be prohibitive.
- Interruptions in connectivity due to random wireless channels, limited mmWave propagation, and susceptibility to blockage.

Usually, to work around these limitations and enable supporting high wireless traffic, wireless networks operators densify the network by deploying more base stations (BS)s , more antennas at the BS, or use more spectrum. Every generation of wireless networks has incurred more power consumption and more radio wave emissions to fulfill its specification (rate, latency, number of connected devices, etc.).

Fleshing out the concept of smart-radio environment might just be the needed alternative of providing better data rates with lower power consumption. Note that current radio environments are not completely controlled, due to random fluctuating channels being the cause for interruptions in connectivity limitation, leading to excessive attenuation, fading, blockage, etc. in some scenarios. To mitigate these effects, smart-radio environments create a tunable space playing an active role in customizing and transferring information signals.

The fundamental limitations stated above need to be resolved with several emerging solutions being considered under the umbrella of smart-radio environments. Considering the high power consumption limitation, one of the more cost-effective enablers of smart-radio environments is called intelligent reflecting surfaces (IRS)s discussed next.

1.3 What are Intelligent Reflecting Surfaces?

To enable smart radio environments, communication engineers converged on the idea of deploying software-controlled surfaces, i.e. IRSs, on objects' surfaces in the environment such as buildings. Intelligent reflecting surfaces (IRS)s are a seemingly recent concept, built on decades of work [7]. Each

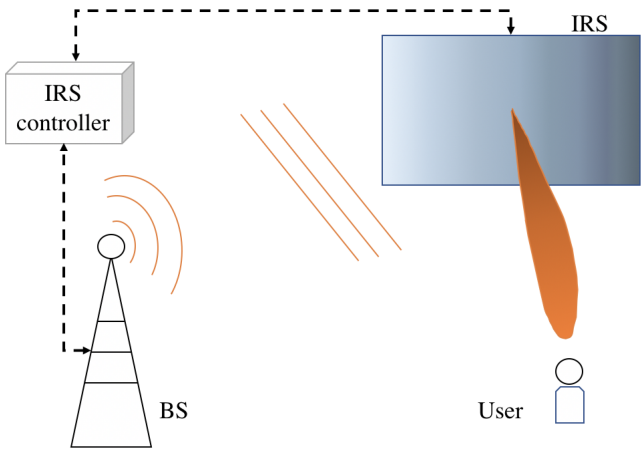

Figure 1.1: IRS passive beamforming for a certain user with an incident EM plane-wave.

IRS contains a large number of economical, passive, and scattering elements, where each element can change the amplitude and/or a phase shift of the impinging electromagnetic (EM) waves to achieve reflect beamforming gains. The distributed smart surfaces or more accurately distributed IRSs would reflect and/or scatter the incoming incident wave to achieve a desired objective. This objective can be in the form of increasing the received power of the desired signal at a receiver, or decreasing the power of interference, or increasing the ratio between the two. An IRS can help to achieve these objectives via beam-focusing, which is tuning the magnitude of the radiation pattern in a certain direction, or beam-steering which is modifying the direction of the beam (see Fig. 1.1). There are many use cases where IRSs may find their niche [5], including: interference mitigation, enhanced coverage, benefiting Tetrahertz bands communication, etc.

1.3.1 Architecture

IRSs are imagined to be reflect antenna arrays or meta-surfaces coating walls in the environment, electrically reconfigured to affect propagating electromagnetic (EM) waves. Currently, IRS research is veering towards meta-surfaces, which are man-made surfaces made up of sub-wavelength sized meta-atoms [8]. These meta-atoms reconfigure and modify the overall impedance of the surface thus manipulating the impinging EM wave and achieving polarization control, beam-steering, or beam-focusing [9]. The

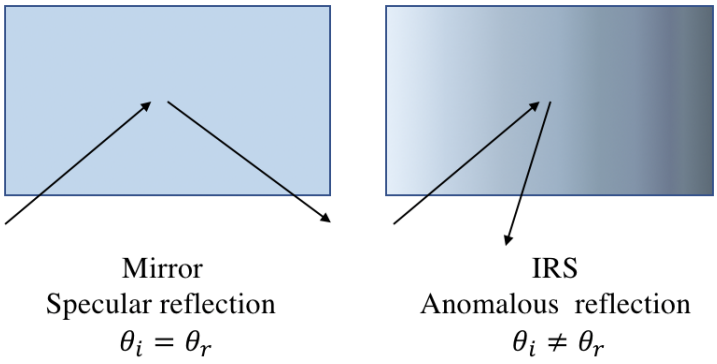

Figure 1.2: Anomalous reflection occurring through continuous changing surface impedance vs specular reflection with constant surface impedance

meta-surface can be implemented with switching diodes which produces a binary approach to changing the amplitude or phase which would have two states, and with varactors (voltage-controlled variable capacitors) which give continuous variations for these properties.

1.3.2 Anomalous vs Specular Reflection

Fig. 1.2 illustrates the difference between anomalous and specular reflection. The former is defined as reflecting to the user by shaping the scattered field so that the main beam is in a desired direction, the color gradient indicates continuously changing surface impedance. While the latter encompasses Snell's law with angle of incidence equaling angle of reflection and has constant surface impedance.

1.4 State of the Art

In this section, the state of the art, focusing on comparisons with relaying technology, passive beamforming, channel estimation, and distributed IRSs deployment is reviewed and the limitations are presented.

1.4.1 Relays vs. Intelligent Reflecting Surfaces

The IRS fills a gap in the area of relay-assisted wireless systems, which is namely the question of achieving full-duplex relaying with low power consumption. An IRS can be considered as a full-duplex transparent relay with transparent relays re-transmitting after amplifying signals such as Amplify-and-Forward (AF) relays, while non-transparent relays re-transmitting after regenerating signals such as Decode-and-Forward (DF) relays. Another distinguishing feature is the mode of operation for relays, classical active relays usually operate in half-duplex, i.e. re-transmission and reception of signals is separated temporally, whereas full-duplex has no such separation. Here, IRSs passively reflect signals without generating new radio signals and are therefore considered energy-efficient.

[10] compares energy efficiency of AF relays with IRSs where IRSs are found to perform better in that regard since they enhance received power strength without using power amplifiers but by appropriate phase shifts design.

On the other hand, [5] makes a valid argument about the uncertain energy consumption of IRSs since their control interface would likely consume lots of power, and until efficient CE protocols and real-time re-configurability of the IRS are solidified, one cannot make such concrete statements about their energy efficiency or the lack thereof. Another meaningful comparison between a DF relay which is non-transparent vs an IRS concludes that IRSs need a large number of reflecting elements to remain competitive to Decode-and-Forward relaying technology [11].

1.4.2 Passive Beamforming

Designing IRS phase shifts as to shape impinging electromagnetic (EM) waves has been studied before extensively, [10, 12–14]. These works assume that the IRS is an array of passive elements able to independently add phase shifts onto the incident EM wave to reflect passively in desired ways, illustrated in Fig. 1.1. Adjusting these phase shifts at the IRSs is referred to as passive beamforming or reflect beamforming. Additionally, some of these works [10, 12, 13] focus on jointly optimizing transmit beamforming at the BS and passive beamforming at the IRS to meet a certain goal such as maximize energy efficiency [10], maximize minimum user SINR subject to a transmit power constraint [12], or minimize transmit power subject to quality of service constraints [13].

Prior works deal with optimizing phase shifts at the IRS while amplitude

coefficients are assumed to have unity magnitude, on the other hand, [13] tackles this assumption and introduces a phase shift model that includes practical phase-dependent amplitude responses of the IRS elements. Furthermore, [14] paves the way for coordinated passive beamforming for the distributed IRS-assisted wireless systems, and jointly optimizes the power at the BS and the phase shifts at each IRS to maximize sum-rate.

1.4.3 Explicit Channel Estimation

Most existing work on designing different IRS schemes assume perfect CSI available at the BS and the IRS. However, this assumption is impractical. The passivity of the IRS translates to an inability to estimate the channels or assist the BS in estimating the channels by transmitting pilot symbols, which poses the biggest challenge in developing CE protocols for these systems [12, 15, 16]. Given the IRS phase shifts have to be updated in each coherence interval, CSI needs to be acquired for all IRS-assisted links at the pace of channel variation, which overburdens the system.

The first estimation protocol that appeared for a single-user IRS-assisted MISO system, known as the ON/OFF protocol [17], estimates all IRS-assisted links one-by-one by serially turning one element ON while keeping the others OFF. This protocol was significantly improved in terms of the normalized mean squared error (NMSE) performance in [18] for a single-user system and in [1] for a multi-user system, where an optimal solution for IRS phase shifts vectors in the CE phase was developed.

However, the CE time scales linearly with the number of IRS elements in all these works, imposing prohibitively high training overhead. To combat this, [19] and [20] introduce the idea of grouping adjacent IRS elements into sub-surfaces, which will decrease the training overhead but also reduce the reflect beamforming gains promised by using a large number of IRS elements. Other solutions exploit channel sparsity that exists in mmWave and massive MIMO channels to develop low-overhead algorithms for CE [21, 22].

A two-timescale CE framework is proposed in [23] that exploits the property that the BS-IRS channel is quasi-static, while the IRS-user channel is dynamic. The assumption that BS-IRS channel is LoS and fixed is also made in many other works such as [1, 12, 15, 21, 24] and is quite practical given BS and IRS have fixed positions with few obstacles around. The authors in [25] envision a novel IRS architecture where a few IRS elements are active and can sense the channels to aid in CE. They then propose compressive sensing and deep learning based methods to design the system.

1.4.4 Distributed Intelligent Reflecting Surfaces: An Extension

The few existing works on CE and beamforming all deal with single IRS-assisted MISO systems and still impose large training overheads. In a multi-user MISO system, severe blockages between the base station (BS) and users can downgrade the overall system performance. Adding a single IRS with a sufficient number of elements provides coverage for a certain area, yet the number of degrees of freedom the IRS adds to the MISO system depends heavily on the IRS placement which can reduce to only one under certain channel conditions. Introduced first in [26], the rank deficiency problem in the BS-IRS link states that the degrees of freedom offered by the channel through the IRS when the BS-IRS channel is LoS will equal to one yielding gains only when there is one user in the system.

Recently, the idea of deploying multiple IRSs in a distributed manner has been considered to overcome several signal blockages by coating the blocking structures with IRSs. Moreover, distributed IRSs are a promising solution to the rank-deficiency problem stated earlier, since the overall BS-IRS link in the distributed case would be the sum of multiple rank one channels which guarantees higher rank channels [27], thereby benefiting the multi-user setting. So far, the couple of works that have appeared on distributed IRSs assume perfect CSI and design beamforming to increase coverage and performance [27, 28].

Only one paper so far handles the IRS-user association problem [29] which is about assigning IRSs to users to optimally balance the passive beamforming gains among different BS-user links. They derive the signal-to-interference-and-noise (SINR) expression at each user in closed form, and create an SINR balancing optimization problem to find the IRS-user association parameters. However, their system model assumes single-antenna BSs to simplify analysis and thus is impractical in a world where massive MIMO is being deployed quickly.

1.5 Problem Statement

This research extends current single-IRS MISO systems to the distributed IRSs MISO systems with a focus on channel estimation, theoretical analysis for average signal-to-interference-and-noise (SINR) at each user, and optimization of IRS-user association. The key aspect of this work is the development of statistical channel estimation protocol for the distributed IRSs assisted MISO system where each BS-IRS link is dominated by the LoS link

while all other links undergo fast fading. A novel closed-form average SINR expression for the system is derived and validated through Monte Carlo simulations. Utilizing the derived average SINR expression, the IRS-user association problem is solved to maximize the minimum user SINR using a successive refinement algorithm. Simulation results illustrate the positive effect of deploying IRSs in a distributed fashion with optimized association parameters.

1.6 Objectives

The main objectives are stated in the following:

- To provide a holistic, unprecedented discourse of distributed IRSs as an enabling technology for beyond 5G (B5G) and 6G, and pinpoint the advantages and disadvantages of deploying distributed IRSs.
- To provide a practical system model and a channel estimation protocol with relatively low channel training overhead for distributed IRSs assisted MISO system under LoS BS-IRS channels and Rayleigh fading IRS-user and direct channels.
- To analyze the average SINR as a performance metric of the distributed IRSs assisted MISO system under LoS BS-IRS channels and Rayleigh fading IRS-user and direct channels.
- To optimize over IRS-user association parameters for a formulated max-min average SINR problem.

1.7 Contributions

The main contributions can be summed up as:

- Tackling the unprecedented CE problem for a distributed IRSs-assisted MISO system. To the best of our knowledge, this research marks the first contribution in developing a CE scheme for the distributed IRSs case (published in [2]).
- Deriving average SINR expression for the distributed IRSs assisted MISO system under MRT precoding and with IRS-user association parameters.

– Designing an effective yet subptimal successive refinement algorithm for the IRS-user association with lower complexity than existing methods. This algorithm finds optimized IRS-user association parameters and performs closely to full enumeration.

– Drawing insights from resulting simulations on the CE protocol as compared to a benchmark protocol as well as on the effect of optimized IRS-user association on the system performance.

1.8 Outline

The rest of this thesis is partitioned as follows: Chapter 2 gives a glimpse onto some wireless communication concepts used in later chapters and introduces the distributed IRSs assisted MISO system model with Rayleigh fading channels and LoS BS-IRS channel. Chapter 3 provides a novel channel estimation protocol to estimate the downlink channel using uplink training. Chapter 4 derives the ergodic analysis for the SINR and then solves an association problem with max-min average SINR as an objective. Chapter 5 includes the conclusion, limitations, and future directions.

Chapter 2

System Model and Preliminaries

In this chapter, we cover the system model utilized in this work and review some key concepts regarding wireless channels. The channels are modelled as Rayleigh fading for the direct and IRS-user links, and as a rank one matrix containing array responses of the BS and each IRS l for the BS-IRS l line-of-sight (LoS) link. The below figure depicts a distributed IRS-assisted MISO system, with IRSs installed on high-rise buildings, a multi-antenna BS, and single-antenna users. Furthermore, there is a centralized control interface between the BS and all the other IRSs.

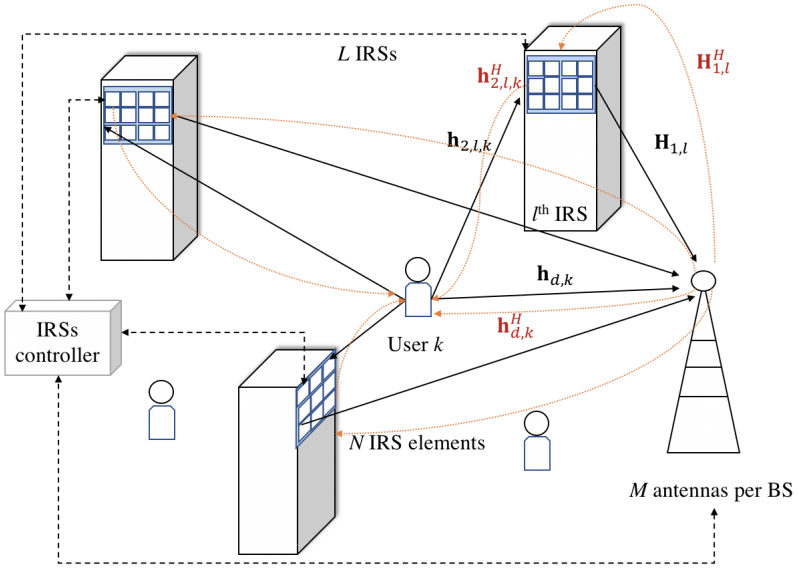

Figure 2.1: Distributed IRSs MISO system model in uplink and downlink.

Consider M antennas at the BS serving K single-antenna users. The communication between the BS and users is assisted by L IRSs, each equipped

with N elements (see Fig. 2.1). The IRSs are deployed in the environment in a distributed manner with fixed positions and their operation is controlled by IRS controllers that communicate with the BS over a backhaul link [12].

The collective channel between the BS and user k in the uplink is given by

$$\mathbf{h}_k = \mathbf{h}_{d,k} + \sum_{l=1}^{L} \mathbf{H}_{1,l} \boldsymbol{\Theta}_l \mathbf{h}_{2,l,k}, \tag{2.1}$$

where $\mathbf{H}_{1,l} \in \mathbb{C}^{M \times N}$ is the channel between BS and IRS l, $\mathbf{h}_{2,l,k} \in \mathbb{C}^{N \times 1}$ is the channel between IRS l and user k, and $\boldsymbol{\Theta}_l = \text{diag}(\alpha_{l,1} \exp(j\theta_{l,1}), \ldots, \alpha_{l,N} \exp(j\theta_{l,N})) \in \mathbb{C}^{N \times N}$ is the reflection matrix for IRS l, where $\theta_{l,n} \in [0, 2\pi]$ is the phase-shift applied by element n of IRS l and $\alpha_{n,l} \in [0, 1]$ is the amplitude reflection coefficient. Also $\mathbf{h}_{d,k} \in \mathbb{C}^{M \times 1}$ is the direct channel between BS and user k.

The distributed IRSs assisted MISO are a generalization of the single IRS scenario, therefore, we build on [26] and extend the system model discussed to distributed surfaces while retaining the similar transmission conditions explained in details below.

2.1 Transmission Mode

Uplink and downlink transmissions can be separated in time or frequency. The uplink is described as the transmission from users to the BS, while the downlink is the transmission of signals from BS to users. Time separation, i.e. time-division duplexing (TDD), occurs when the downlink transmission operates in a different times slot than the uplink transmission but both occupy the same frequency band, whereas frequency division-duplexing (FDD) happens when the downlink operates in a different frequency band than the uplink at the same time as depicted in Fig. 2.2. While FDD is deployed in practice, TDD has become part of the canonical model for massive multiple-input multiple-output (MIMO) technology mainly because the channel estimation for TDD has less training time [30].

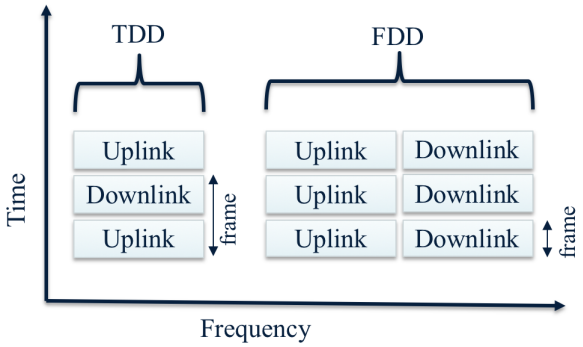

Figure 2.2: Time Division Duplexing (TDD) vs Frequency Division Duplexing (FDD) transmission

In this IRS-assisted MISO system model, we consider a TDD transmission mode due to the similarities between IRS-assisted wireless systems and massive MIMO. One of which is the prohibitively large system dimensions ergo large numbers of channels to estimate (cf. sec. 2.2.2). Here, TDD exploits channel reciprocity and thus eliminates the need to estimate channels both in the uplink and downlink (for example $\mathbf{h}_k^H$ in the downlink is the Hermetian of $\mathbf{h}_k$ defined in (2.1) in the uplink). In the next section, we lay out the fading model for the direct channel $\mathbf{h}_{d,k}$ and IRS l-user k channel $\mathbf{h}_{2,l,k}$, in addition, we present some important definitions regarding fading channels.

2.2 Fading Direct and IRS-User Channels

Electromagnetic waves propagating through wireless channels are sufficiently yet incompletely characterized by multipath fading (depicted in Fig. 2.3) and shadowing among other effects. This is due to the complicated propagation through dynamic moving terminals in the environments or to rich scattering that cannot be mathematically fully described. However, considerable work has been done to statistically describe and model these random fading channels in certain settings. One of the simpler models is Rayleigh fading which captures the multipath component found in wireless channels [31].

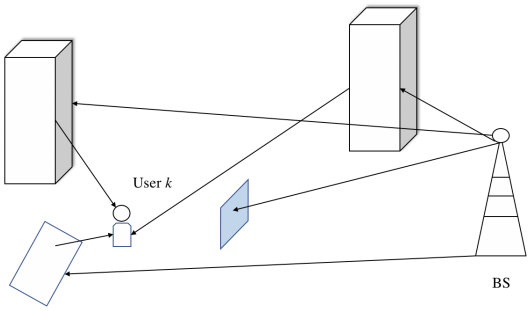

Figure 2.3: Rayleigh fading with no LoS path

Definition 2.1. Multipath fading is characteristic of wireless channels where the propagating wave is randomly scattered, delayed, reflected, and diffracted which creates multiple paths in the environment due to scatterers that create these additional paths, thus leading to destructive and/or constructive interference at the terminal.

Fading is modeled as random channels with some distribution, one of which is the Rayleigh distribution for a fading channel $\mathbf{h}$, so that

$$\mathbf{h} \in \mathcal{CN}(0, \mathbf{C}) \tag{2.2}$$

where $\mathbf{C} \in C^{M \times M}$ is the covariance matrix, and M is the number of antennas. One practical scenario which is often modelled using Rayleigh fading is mobile systems, where no line-of-sight (LoS) path exists between the transmitter and receiver. Considering the system model illustrated in Fig. 2.1, the ground users will likely be blocked by surrounding objects. Thus, the modelling of Rayleigh fading for the direct and IRS-user channels is justified. Moreover, we assume the fading these channels undergo is classified as block-fading and flat-fading.

Definition 2.2. Block fading occurs when the response of the channel is constant (static) for a certain period of time. This period is known as coherence interval.

The length of a coherence interval depends on the fading channel characteristics and the mobility of the user, assuming that the transmitter is fixed. Higher mobility of the user quickens the pace of channel variations, and lessens the period of static behavior. For a TDD transmission mode previously defined in sec. 2.1, portions for uplink and downlink transmission are reserved in the coherence period, where a fraction of the uplink is

also reserved for pilot training to achieve channel estimation (CE), demonstrated in Fig. 2.4. This poses a time constraint on the CE overhead, and effectively on the number of pilots (preamble signals used for CE) a user can send.

Figure 2.4: Channel coherence interval divided into uplink training phase, uplink phase, and downlink phase.

As for the flat-fading assumption for the $\mathbf{h}_{2,l,k}$ and $\mathbf{h}_{d,k}$ channels, we note that these channels remain frequency non-selective under certain conditions, which implies no channel induced inter-symbol interference (ISI). More formally, we present the following definition.

Definition 2.3. Flat-fading channels have the property that all received multipath components (see **Definition 2.1**) of a symbol arriving within the duration of that symbol [32], in other words, when the bandwidth of the channel is larger than the bandwidth of the signal.

Now, we model statistically $\mathbf{h}_{2,l,k}$ and $\mathbf{h}_{d,k}$ channels by correlated Rayleigh fading represented as

$$\mathbf{h}_{2,l,k} = \sqrt{\beta_{2,l,k}}\mathbf{R}^{1/2}_{IRS_{l,k}}\mathbf{z}_{l,k}, \tag{2.3}$$

$$\mathbf{h}_{d,k} = \sqrt{\beta_{d,k}}\mathbf{R}^{1/2}_{BS_k}\mathbf{z}_{d,k}, \tag{2.4}$$

where $\mathbf{R}_{IRS_{l,k}}$ is the spatial correlation matrix of the channel from IRS l to user k, $\mathbf{R}_{BS_k}$ is the spatial correlation matrix from the BS to user k, and $\mathbf{z}_{l,k} \sim \mathcal{CN}(0, \mathbf{I}_N)$ and $\mathbf{z}_{d,k} \sim \mathcal{CN}(0, \mathbf{I}_M)$ describe the fast fading vectors of the IRS-user link and the BS-user link, respectively. The path loss factor is denoted as $\beta_{d,k}$ for the direct channel and $\beta_{2,l,k}$ for the IRS-user k link. The notion of spatial correlation is explained below in sec. 2.2.1, and it arises in these vector channel models because the antennas at the BS as well as the reflecting elements in the IRSs are arranged compactly.

2.2.1 Spatial Correlation

Spatial correlation is a property related to multiple antennas at the transmitter and/or receiver. In layman words, it arises in practice when received power is stronger in certain directions than others. More formally, spatial correlation happens due to two factors, one of which is propagation environment, where the channel vector is more probable to point in some directions more than others. The other is because of insufficient spacing between antennas [33]. Take $\mathbf{R}_{IRS_{l,k}}$ in (2.3) for instance, the off-diagonal elements will represent the correlation between the antenna elements, while the diagonal elements are all set to one.

Definition 2.4. Spatially Correlated Channels. A fading channel vector $\mathbf{h}$ is Rayleigh distributed if it follows $\mathbf{h} \sim \mathcal{CN}(0, \mathbf{C})$, where $\mathbf{C}$ is the covariance matrix of the distribution (also its correlation matrix). When the channel's direction $\mathbf{h}/\|\mathbf{h}\|$ and its magnitude $\|\mathbf{h}\|$ are independent random variables then $\mathbf{h}$ is said to be spatially uncorrelated. Otherwise, it is spatially correlated [30].

2.2.2 Channel Estimation

Channel estimation (CE) is a process of sending pilots (known sequence of symbols between transmitter and receiver) through an unknown channel. The transmission coefficients which are complex-valued channel gains can then be estimated by correlating the received signal with the known pilots at the receiver's end.

Since the IRSs are completely passive and do not have any signal processing capabilities, the BS has to estimate all channels and compute the desired configurations for all the IRSs then share this information with the IRSs' micro-controllers over back-haul connections. Each IRS then applies the required reflection configuration $\mathbf{\Theta}_l$, $l = 1, \ldots, L$.

The reason for CE is to negate the detrimental effects of channels or at least suppress those effects, to do that one needs to sense these channels first. If we model the effects of the channels as linear, we may use a linear estimator, one of which is the least squares (LS) estimator which is an optimal linear estimator given no prior information of the channels. On the other hand, if the BS ascertains the large-scale statistics of the Rayleigh fading channels given in (2.4) and (2.3), we can shift to a Bayesian channel estimation method called minimum mean squared error (MMSE) estimation which is optimal for Rayleigh-distributed fading channels.

In the following section, we present the channel model for $\mathbf{H}_{1,l}$, which we show next to be of a LoS nature and known a priori at the BS. Note that since we make the assumption of knowing $\mathbf{H}_{1,l}$, the number of channels to estimate in the system model, illustrated in Fig. 2.1, is reduced to $(M+NL)K$ instead $(M++NL)K + MNL$. Although the number of unknowns is reduced we may incur high training time dedicated for channel estimation especially in the multi-user setting and large M, N, L system dimensions.

2.3 BS-IRS l Channel

Previously, we established practical models for the direct and IRS-user k channels. Recall there are two channels which make up each IRS cascaded channel, one of which is the BS-IRS l channel $\mathbf{H}_{1,l}$ which is defined here. Many works on IRS-assisted systems, for example: [1, 12, 15, 21, 24], assume the BS-IRS channel to be LoS dominated. This assumption is quite practical and is supported in literature with the following remarks:

- First, there will always exist a LoS path between the BS and IRS. As the BS tower is generally elevated high and the IRS is also envisioned to be integrated onto the walls of (high-rise) buildings, so both will have few obstacles around. Given the positions of BS and IRSs are fixed, a stable LoS channel between the BS and each IRS will exist and can be constructed at the BS using directional (LoS angle of departure (AoD) and angle of arrival (AoA)) information.

- Second, the path loss in NLoS paths is much larger than that in LoS path loss in the next generation systems, resulting in any NLoS paths constituting the BS-IRS channel to be much weaker. In fact it is noted that in mmWave systems, the typical value of Rician factor (ratio of energy in LoS to that in NLoS) is 20dB and can be as large as 40dB in some cases [24], which is sufficiently large to neglect any NLoS channel components. Under these remarks, we assume that each BS-IRS l channel is LoS.

We make the next assumption that the BS would have prior knowledge of $\mathbf{H}_{1,l}$ because of the LoS dominant nature of each IRS l and the BS link explained earlier. The BS can have knowledge of the locations, orientations, and array parameters of the IRS to construct the BS-to-IRS channel.

Assuming a uniform linear array (ULA) at the BS and uniform planar array (UPA) at the IRSs depicted in Fig. 2.5, the LoS BS-IRS l channel can

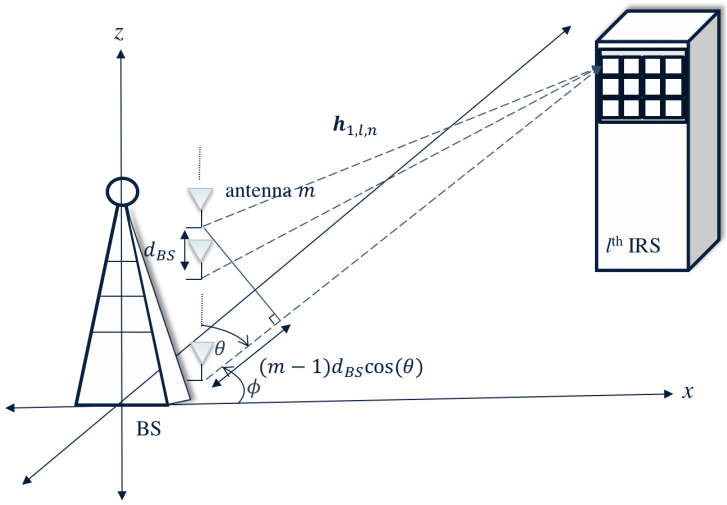

Figure 2.5: Model of $\mathbf{h}_{1,l,n} \in \mathbb{C}^{M\times 1}$ which is the n^{th} column vector of of LoS BS-IRS l channel matrix $\mathbf{H}_{1,l}$ with multiple transmit antennas at the BS and multiple elements at IRS l. The signals from the transmit antennas arrive almost in parallel at the n^{th} receive element of the IRS.

be written as

$$\mathbf{H_{1,l}} = \sqrt{\beta_{1,l}}\mathbf{a}_l\mathbf{b}_l^H. \tag{2.5}$$

Here, $\beta_{1,l}$ the path loss factor for BS-IRS l channel and $\mathbf{a}_l$ is the array response vector for the BS, while $\mathbf{b}_l$ is the array response vector for IRS l. Since we assume a ULA at the BS, the steering vector $\mathbf{a}_l \in \mathbb{C}^{M\times 1}$ can be written as

$$\mathbf{a}_l = [1, e^{-jkd_{BS}\cos(\theta)}, \ldots, e^{-jkd_{BS}(M-1)\cos(\theta)}]^T, \tag{2.6}$$

where d_{BS} is the inter-antenna spacing, $k = 2\pi/\lambda_c$ is the wave number, λ_c is carrier wavelength, θ is the elevation angle-of-departure (AoD) from the BS to IRS l. [34].

The IRS is envisioned to be a planar array of $N = N_xN_z$ elements, where N_x denotes the number of horizontally placed elements and N_z is the number of vertically placed elements. The array response vector $\mathbf{b}_l \in \mathbb{C}^{N\times 1}$ for a UPA at IRS l is expressed as

$$\mathbf{b}_l = \mathbf{b}_{l,x}^T \otimes \mathbf{b}_{l,z}^T \tag{2.7}$$

Here, $\otimes$ denotes the Kronecker product, and

$$\mathbf{b}_{l,x} = [1, e^{-jkd_{IRS}\sin(\varphi)\cos(\vartheta)}, \ldots, e^{-jkd_{IRS}(N_x-1)\sin(\varphi)\cos(\vartheta)}], \tag{2.8}$$

$$\mathbf{b}_{l,z} = [1, e^{-jkd_{IRS}\cos(\varphi)}, \ldots, e^{-jkd_{IRS}(N_z-1)\cos(\varphi)}], \tag{2.9}$$

where d_{IRS} is the inter-antenna spacing, and $\vartheta(\varphi)$ denote the azimuth (elevation) angle-of-arrival AoA of the path from BS to IRS l [35], [36]. It is apparent that each BS-IRS l channel matrix has a rank one structure (because it is made up of the outer product of two vectors), thus $\mathbf{H}_{1,l}$ is a rank-one matrix with a unique non-zero singular value $\lambda_{\mathbf{H}_{1,l}} = \sqrt{\beta_{1,l}}$ [34]. The rank one assumption is accurate since the elements at both the BS and IRSs are co-located, as both the BS and IRSs are elevated such that no obstacles lie in-between, and the IRSs are assumed to be in the far-field, such that the propagation distance d_{BS-IRS_l} between the BS and IRS l is much bigger than the largest dimension of that IRS. To ensure far-field assumption holds true, we have that

$$d_{BS-IRS_l} \geq \frac{2D^2}{\lambda_c} \tag{2.10}$$

and D is the largest dimension of the antenna array. Note that D in the ULA case is the total length of the array. While D for UPA is the diagonal length of the UPA array. Although a channel with rank one means only one degree of freedom for the system, with distributed surfaces the overall channel between BS and each user will at least have rank L, which allows multiple users to be served efficiently.

2.4 Maximum Ratio Transmission Precoding

In sec. 2.2, we establish the definition of multipath fading, here we attempt to mitigate it. Multipath fading is thought to be the most adverse propagation effect for wireless communications, but a way to combat it is to use antenna diversity techniques, one of which is known as maximum ratio precoding (MRT) [37].

Note that to achieve precoding, one needs at least two transmit dimensions in general, since one can precode in time or frequency in addition to space (antennas). Define an MRT digital precoding vector $\mathbf{f}_k \in \mathbb{C}^{M\times 1}$, with M being the number of antennas at the transmitter, written as

$$\mathbf{f}_k = \frac{\mathbf{h}_k}{\sqrt{\mathbb{E}[\|\mathbf{h}_k\|^2]}} \tag{2.11}$$

with $\mathbf{h}_k$ stated in (2.1) and similar to [38].

With precoding such as (2.11) used at the BS, the transmit signal $\mathbf{x} \in \mathbb{C}^{M\times 1}$ is formulated as

$$\mathbf{x} = \sum_{k=1}^{K} \sqrt{p_k}\mathbf{f}_k s_k, \tag{2.12}$$

where p_k and $s_k \in \mathcal{CN}(0,1)$ are the allocated power and data symbol of user k, respectively.

Given s_k's are independently and identically distributed (i.i.d.) $\mathcal{CN}(0,1)$ variables, $\mathbf{x}$ has to satisfy the average Tx power constraint as

$$\mathbb{E}[||\mathbf{x}||^2] = \text{tr } (\mathbf{P}\mathbf{F}^H\mathbf{F}) \leq P_{max}, \tag{2.13}$$

where $P_{max} > 0$ is the power constraint at the BS, $\mathbf{P} = \text{diag}(p_1, \ldots, p_K) \in \mathbb{C}^{K\times K}$ is the power allocation matrix and $\mathbf{F} = [\mathbf{f}_1, \ldots, \mathbf{f}_K] \in \mathbb{C}^{M\times K}$ is the precoding matrix.

MRT precoding maximizes signal-to-noise ratio (SNR) which is optimal for single-user wireless systems, yet for the multi-user setting this is not the optimal precoder as this precoding technique does not factor in interference from other users. However, it is simpler than other precoding techniques and thus utilized in the theoretical analysis in Chapter 4.

2.5 Downlink Transmission

Assuming flat-fading channels (cf. **Definition 2.3**), the received baseband signal in the downlink at user k is defined as

$$y_k = \mathbf{h}_k^H \mathbf{x} + n_k, \tag{2.14}$$

and $n_k \sim \mathcal{CN}(0, \sigma^2)$ is the noise at the receiver with noise variance σ^2, $\mathbf{x}$ is the transmit signal determined in (2.12), and $\mathbf{h}_k$ defined as in (2.1). The expression for (2.14) is used in Chapter 4 to compute the average signal-to-interference and noise (SINR) at each user.

2.6 Summary

In this chapter, we described the system model, the channels' models, and the transmission scheme. To realize an optimized transmission, estimated channels are obtained at the BS which is covered in Chapter 3

building on the system model and concepts covered here to find a novel CE protocol for the distributed IRSs assisted MISO system. After this, the BS can use a precoding scheme such as MRT to send precoded signals to the users in the downlink. Since this is a TDD system, the BS can compute the average SINRs for all the users. The SINR is then used as an important performance metric at the BS to optimize transmission through associating IRSs to users and configuring IRSs optimally discussed in Chapter 4. The reason SINR is chosen to represent the system holistically is that other performance metrics such as transmission rate, outage probability, and bit error rate are all functions of SINR.

Chapter 3

Channel Estimation

This chapter presents the proposed CE protocol and derives analytical expressions of the MMSE estimates and NMSE. We consider correlated Rayleigh fading channels between the BS and users, and the IRSs and users, while each BS-IRS channel is LoS. The LoS BS-IRSs channel matrices can be computed a priori at the BS using the locations of the IRS, which are fixed and can be used to compute the LoS angle of departure (AoD) and angle of arrival (AoA). Leveraging this, we develop an optimal CE protocol using the Bayesian technique of minimum mean squared error (MMSE) estimation, that promises low CE error and imposes a significantly lower training overhead than what we need when we extend the conventional protocols from [1, 17, 18] to this setup. The normalized mean squared error (NMSE) in the MMSE estimates is analytically derived and studied using simulations.

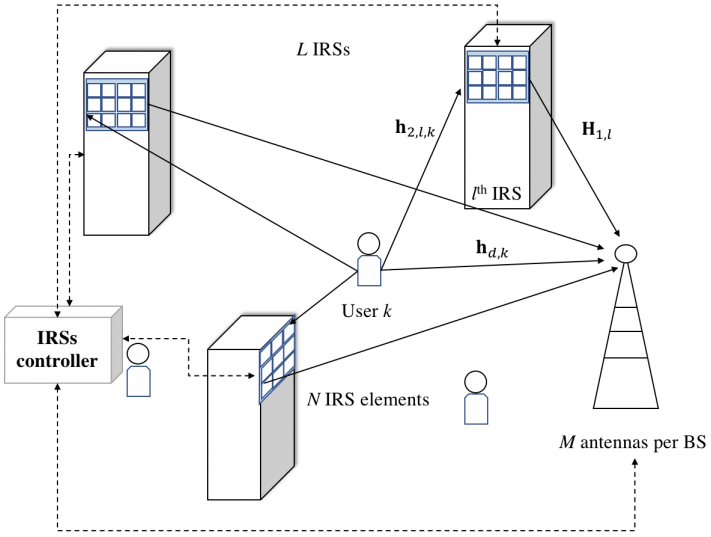

Figure 3.1: Distributed Intelligent Reflecting Surfaces multi-user MISO system uplink transmission.

3.1 Proposed Protocol

In this research, we assume a time division duplexing (TDD) system where the downlink channels are estimated using the received uplink pilot symbols by exploiting channel reciprocity. Since the IRSs are passive, the BS has to estimate all channels and determine optimal IRS configurations.

We assign an uplink training phase of τ_C seconds (sec) in the coherence interval (cf. Fig. 2.4) which is further broken down into S sub-phases of length $\tau_S = \frac{\tau_C}{S}$ sec. Throughout the CE phase, users transmit mutually orthogonal pilot sequences of length T_S symbols, with $T_S = \frac{\tau_S}{\tilde{\tau}}$ where $\tilde{\tau}$ is the duration of each symbol. The pilot sequence of user k is denoted as $\mathbf{x}_{p,k} = [x_{p,k,1}, \ldots, x_{p,k,T_S}]^T \in \mathbb{C}^{T_S \times 1}$, such that $\mathbf{x}_{p,k}^H \mathbf{x}_{p,l} = 0$, for $k \neq l$, $k, l = 1, \ldots, K$ and $\mathbf{x}_{p,k}^H \mathbf{x}_{p,k} = P_C T_S \tilde{\tau} = P_C \tau_S$ Joules, where P_C is the transmit power of each user. The IRS l applies the reflect beamforming matrix $\mathbf{\Theta}_{l,s} = \text{diag}[\phi_{l,s,1}, \ldots, \phi_{l,s,N}]^T \in \mathbb{C}^{N \times N}$ in sub-phase s, where $\phi_{l,s,n} = \alpha_{l,s,n} \exp(j\theta_{l,s,n})$. The received training signal at the BS, $\mathbf{Y}_s^{tr} \in \mathbb{C}^{M \times T_S}$ in sub-phase s is then given as

$$\mathbf{Y}_s^{tr} = \sum_{k=1}^{K} (\mathbf{h}_{d,k} + \sum_{l=1}^{L} \mathbf{H}_{1,l} \mathbf{\Theta}_{l,s} \mathbf{h}_{2,l,k}) \mathbf{x}_{p,k}^H + \mathbf{N}_s^{tr} \tag{3.1}$$

where $\mathbf{N}_s^{tr} \in \mathbb{C}^{M \times T_S}$ is the matrix of noise vectors at the BS, with each column distributed independently as $\mathcal{CN}(\mathbf{0}, \sigma^2 \mathbf{I}_M)$. After correlating the received training signal with the training sequence of user k, the BS obtains the observation vector, $\mathbf{r}_{s,k}^{tr} \in \mathbb{C}^{M \times 1}$, for user k in sub-phase s as

$$\mathbf{r}_{s,k}^{tr} = (\mathbf{h}_{d,k} + \sum_{l=1}^{L} \mathbf{H}_{1,l} \mathbf{\Theta}_{l,s} \mathbf{h}_{2,l,k}) + \mathbf{n}_{s,k}^{tr}, \tag{3.2}$$

which can be compactly written as

$$\mathbf{r}_{s,k}^{tr} = \mathbf{h}_{d,k} + \mathbf{H}_1 \mathbf{\Theta}_s \mathbf{h}_{2,k} + \mathbf{n}_{s,k}^{tr},\ k = 1, \ldots, K, \tag{3.3}$$

where $\mathbf{h}_{2,k} = [\mathbf{h}_{2,1,k}^T, \ldots, \mathbf{h}_{2,L,k}^T]^T \in \mathbb{C}^{NL \times 1}$ is the concatenated vector of all L IRS-user k channels, $\mathbf{\Theta}_s = \text{diag}(\mathbf{\Theta}_{1,s}, \ldots, \mathbf{\Theta}_{L,s}) \in \mathbb{C}^{NL \times NL}$ is the reflection matrix of all the IRSs in sub-phase s, $\mathbf{H}_1 = [\mathbf{H}_{1,l}, \ldots, \mathbf{H}_{1,l}] \in \mathbb{C}^{M \times NL}$ is the concatenation of all L BS-IRS channel matrices, and $\mathbf{n}_{s,k}^{tr} = \frac{\mathbf{N}_s^{tr} \mathbf{x}_{p,k}}{P_C \tau_S} \in \mathbb{C}^{M \times 1}$ is the noise.

Since we remarked earlier that $\mathbf{H}_{1,l} \in \mathbb{C}^{M \times N}$ can be perfectly known at the BS using locations of the IRSs, we will focus only on the estimation of

$\mathbf{h}_{d,k}$s and $\mathbf{h}_{2,k}$s. To this end, we define

$$\bar{\mathbf{H}}_1 = \text{diag}(\sqrt{M}\mathbf{I}_M, \mathbf{h}_{1,1,1}, \ldots, \mathbf{h}_{1,1,N}, \mathbf{h}_{1,2,1} \ldots, \mathbf{h}_{1,2,N}, \ldots, \mathbf{h}_{1,L,1}, \ldots, \mathbf{h}_{1,L,N}) \in \mathbb{C}^{M(NL+1)\times NL+M}, \tag{3.4}$$

as the block diagonal matrix of $\mathbf{h}_{1,l,n}$s, where $\mathbf{h}_{1,l,n}$ is the nth column of $\mathbf{H}_{1,l}$. Aggregating the observation vectors across the S sub-phases as $\mathbf{r}_k^{tr} = [\mathbf{r}_{1,k}^{tr^T}, \ldots, \mathbf{r}_{S,k}^{tr^T}]^T \in \mathbb{C}^{SM\times 1}$, we obtain

$$\mathbf{r}_k^{tr} = (\mathbf{V}_{tr} \otimes \mathbf{I}_M)\bar{\mathbf{H}}_1 \mathbf{h}_{d,2,k} + \mathbf{n}_k^{tr},\ k = 1, \ldots, K, \tag{3.5}$$

where,

$$\mathbf{V}_{tr} = \begin{bmatrix} 1 & \mathbf{v}_{1,1}^T & \cdots & \mathbf{v}_{L,1}^T \\ \vdots & \vdots & & \vdots \\ 1 & \mathbf{v}_{1,S}^T & \cdots & \mathbf{v}_{L,S}^T \end{bmatrix} \in \mathbb{C}^{S\times(NL+1)} \tag{3.6}$$

$\mathbf{h}_{d,2,k} = [\frac{1}{\sqrt{M}}\mathbf{h}_{d,k}^T, \mathbf{h}_{2,k}^T]^T \in \mathbb{C}^{(NL+M)\times 1}$ cascades the direct and IRS-user channels and $\mathbf{n}_k^{tr^T} = [\mathbf{n}_{1,k}^{tr^T}, \ldots, \mathbf{n}_{S,k}^{tr^T}]^T \in \mathbb{C}^{SM\times 1}$ is the concatenated noise across all sub-phases. Note that $\mathbf{v}_{l,s} = \text{vec}(\mathbf{\Theta}_{l,s})$ is the phase shifts vector for IRS l in sub-phase s.

Define $\tilde{\mathbf{V}}_{tr} = (\mathbf{V}_{tr} \otimes \mathbf{I}_M)\bar{\mathbf{H}}_1 \in \mathbb{C}^{SM\times NL+M}$. To guarantee the existence of left pseudo-inverse of $\tilde{\mathbf{V}}_{tr}$ (given as $\tilde{\mathbf{V}}_{tr}^{\dagger} = (\tilde{\mathbf{V}}_{tr}^H\tilde{\mathbf{V}}_{tr})^{-1}\tilde{\mathbf{V}}_{tr}^H$), we must have

$$SM \geq NL + M. \tag{3.7}$$

Rewriting the condition to accentuate only the parameter S, we obtain

$$S \geq \frac{NL}{M} + 1. \tag{3.8}$$

Because of prior knowledge of $\mathbf{H}_{1,l}$, the proposed CE protocol estimates channels sequentially while training time is decreased by a factor of M due to having M copies of the $\mathbf{h}_{2,k}$ link, 1 sub-phase remains for the direct channel, while for cascaded IRS channels it becomes NL/M since partial knowledge of the cascaded IRS links is present at the BS along with large-scale statistics of the unknown part of the link. Thus, the minimum number of CE sub-phases is actually reduced by a factor of M, as compared to the conventional protocols. For example: as compared to the protocol in [18] for a single IRS ($L = 1$) system, which requires $N + 1$ sub-phases to estimate the channels, we will require $N/M + 1$ sub-phases, which is a substantial reduction given next generation systems use large numbers of BS antennas.

Multiplying (3.5) by $\tilde{\mathbf{V}}_{tr}^{\dagger}$ results in

$$\tilde{\mathbf{r}}_k^{tr} = \mathbf{h}_{d,2,k} + (\tilde{\mathbf{V}}_{tr}^H \tilde{\mathbf{V}}_{tr})^{-1} \tilde{\mathbf{V}}_{tr}^H \mathbf{n}_k^{tr}. \tag{3.9}$$

We denote by $\tilde{\mathbf{n}}_k^{tr} = (\tilde{\mathbf{V}}_{tr}^H \tilde{\mathbf{V}}_{tr})^{-1} \tilde{\mathbf{V}}_{tr}^H \mathbf{n}_k^{tr} \in \mathbb{C}^{(NL+M)\times 1}$ the measurement noise and by $\mathbf{C}_{\tilde{\mathbf{n}}_k^{tr}} = \mathbb{E}[\tilde{\mathbf{n}}_k^{tr} \tilde{\mathbf{n}}_k^{tr^H}] \in \mathbb{C}^{(NL+M)\times(NL+M)}$ its covariance matrix. Next, we minimize the noise variance in (3.9) and derive the MMSE estimates.

3.2 Noise Variance Minimization

Next step is to design $\mathbf{V}_{tr}$ such that variance of the noise is minimized and the noise across the estimated channels is uncorrelated.

Recalling that $\mathbf{n}_{s,k}^{tr} = \frac{\mathbf{N}_s^{tr} \mathbf{x}_{p,k}}{P_C \tau_S}$ we obtain

$$\mathbf{C}_{\tilde{\mathbf{n}}_k^{tr}} = (\tilde{\mathbf{V}}_{tr}^H \tilde{\mathbf{V}}_{tr})^{-1} \tilde{\mathbf{V}}_{tr}^H \mathbb{E}\left[\mathbf{n}_k^{tr} \mathbf{n}_k^{tr^H}\right] \tilde{\mathbf{V}}_{tr} (\tilde{\mathbf{V}}_{tr}^H \tilde{\mathbf{V}}_{tr})^{-1}, \tag{3.10}$$

$$= \frac{\sigma^2 P_C \tau_S}{(P_C \tau_S)^2} (\tilde{\mathbf{V}}_{tr}^H \tilde{\mathbf{V}}_{tr})^{-1} = \frac{\sigma^2}{P_C \tau_S} (\tilde{\mathbf{V}}_{tr}^H \tilde{\mathbf{V}}_{tr})^{-1}, \tag{3.11}$$

where,

$$\tilde{\mathbf{V}}_{tr}^H \tilde{\mathbf{V}}_{tr} = \bar{\mathbf{H}}_1^H (\mathbf{V}_{tr}^H \mathbf{V}_{tr} \otimes \mathbf{I}_M) \bar{\mathbf{H}}_1. \tag{3.12}$$

We note that by construction $\bar{\mathbf{H}}_1^H \bar{\mathbf{H}}_1 = M\boldsymbol{\Sigma}$, where $\boldsymbol{\Sigma} = \text{diag}(\mathbf{I}_M, \beta_{1,1}\mathbf{I}_N, ..., \beta_{1,L}\mathbf{I}_N) \in \mathbb{C}^{NL+M \times NL+M}$. Therefore, to ensure uncorrelated channel estimates and equal noise variance in each estimate, $\mathbf{V}_{tr}$ must have equally scaled orthogonal columns such that $(\mathbf{V}_{tr}^H \mathbf{V}_{tr})^{-1} = \zeta \mathbf{I}_{NL+1}$. Minimizing the noise variance is equivalent now to minimizing ζ, under the constraints:

1. $\mathbf{V}_{tr}$ has the structure in (3.6).
2. $\alpha_{l,s,n} \in [0, 1]$.
3. $\theta_{l,s,n} \in [0, 2\pi]$.
4. $(\mathbf{V}_{tr}^H \mathbf{V}_{tr})^{-1} = \zeta \mathbf{I}_{NL+1}$.

The last constraint can be written as

$$\zeta = \frac{NL+1}{tr(\mathbf{V}_{tr}^H \mathbf{V}_{tr})} = \frac{NL+1}{\sum_{s=1}^{S} \sum_{n=1}^{NL+1} |[\mathbf{V}_{tr}]_{s,n}|^2}. \tag{3.13}$$

Using the second constraint in (3.13), we see that

$$\zeta \geq \frac{1}{S}. \tag{3.14}$$

A solution that meets the lower bound in (3.14) with equality is the $NL+1$ leading columns of an $S \times S$ DFT matrix as expressed below [1, 18].

$$[\mathbf{V}_{tr}]_{s,n} = w^{(n-1)(s-1)},\ n = 1, \ldots, NL+1, s = 1, \ldots, S, \tag{3.15}$$

where $w = e^{-j2\pi/S}$ is the primitive Sth root of unity. This design of $\mathbf{V}^{tr}$ results in $(\mathbf{V}_{tr}^H \mathbf{V}_{tr})^{-1} = \frac{1}{S}\mathbf{I}_{NL+1}$ and therefore achieves the minimum noise variance in (3.14) since $\zeta = \frac{1}{S}$. Using this we can simplify (3.12) as

$$\tilde{\mathbf{V}}_{tr}^H \tilde{\mathbf{V}}_{tr} = S\bar{\mathbf{H}}_1^H \bar{\mathbf{H}}_1 = SM\boldsymbol{\Sigma}, \tag{3.16}$$

and $(\tilde{\mathbf{V}}_{tr}^H \tilde{\mathbf{V}}_{tr})^{-1} = \frac{\boldsymbol{\Sigma}^{-1}}{SM}$.

To derive the MMSE estimates, we refer back to (3.9) and simplify it with the optimal DFT design in (3.15) to obtain

$$\tilde{\mathbf{r}}_k^{tr} = \mathbf{h}_{d,2,k} + \frac{1}{SM}\boldsymbol{\Sigma}^{-1}\tilde{\mathbf{V}}_{tr}^H \mathbf{n}_k^{tr},\ k = 1, \ldots, K, \tag{3.17}$$

Recalling that $\mathbf{h}_{d,2,k} = [\frac{1}{\sqrt{M}}\mathbf{h}_{d,k}^T, \mathbf{h}_{2,k}^T]^T$, we can write $\tilde{\mathbf{r}}_k^{tr} = [\tilde{\mathbf{r}}_{0,k}^{tr^T}, \tilde{\mathbf{r}}_{1,k}^{tr^T}, \ldots, \tilde{\mathbf{r}}_{L,k}^{tr^T}]^T \in \mathbb{C}^{(M+NL)\times 1}$, where $\tilde{\mathbf{r}}_{0,k}^{tr} \in \mathbb{C}^{M\times 1}$ is used to estimate the direct channel and $\tilde{\mathbf{r}}_{l,k}^{tr} \in \mathbb{C}^{N\times 1}$ is used to estimate the channel from IRS l to user k.

The relationship between $\tilde{\mathbf{r}}_{0,k}$ and $\mathbf{h}_{d,k}$ for user k can be written using (3.17) and definition of $\tilde{\mathbf{V}}_{tr}$ as

$$\begin{aligned}\tilde{\mathbf{r}}_{0,k}^{tr} &= \frac{1}{\sqrt{M}}\mathbf{h}_{d,k} + \frac{1}{SM}\boldsymbol{\Sigma}_0^{-1}\sqrt{M}\mathbf{I}_M(\mathbf{v}_1^{tr} \otimes \mathbf{I}_M)^H \mathbf{n}_k^{tr} \\ &= \frac{1}{\sqrt{M}}\mathbf{h}_{d,k} + \frac{1}{S\sqrt{M}}(\mathbf{v}_1^{tr} \otimes \mathbf{I}_M)^H \mathbf{n}_k^{tr}.\end{aligned} \tag{3.18}$$

where $\sqrt{M}\mathbf{I}_M$ corresponds to the first $M \times M$ diagonal matrix of $\bar{\mathbf{H}}_1$, $\boldsymbol{\Sigma}_0^{-1} = \mathbf{I}_M$ is the first $M \times M$ diagonal matrix of $\boldsymbol{\Sigma}^{-1}$ and $\mathbf{v}_1^{tr}$ is the first $S \times 1$ column of $\mathbf{V}_{tr}$. This observation vector can be scaled by a factor of $\sqrt{M}$ at the BS to obtain

$$\tilde{\mathbf{r}}_{0,k}^{tr} = \mathbf{h}_{d,k} + \frac{1}{S}(\mathbf{v}_1^{tr} \otimes \mathbf{I}_M)^H \mathbf{n}_k^{tr}. \tag{3.19}$$

Note that (3.19) is actually the LS estimate of $\mathbf{h}_{d,k}$ which we denote as $\hat{\mathbf{h}}_{d,k}^{LS}$.

3.3 Minimum Mean Squared Error Estimation

The BS can compute the MMSE estimate of $\mathbf{h}_{d,k}$ as stated in the following lemma.

Theorem 3.1. *The MMSE estimate* $\hat{\mathbf{h}}_{d,k}$ *of* $\mathbf{h}_{d,k}$ *is given as*

$$\hat{\mathbf{h}}_{d,k} = \beta_{d,k}\mathbf{R}_{BS_k}\left(\beta_{d,k}\mathbf{R}_{BS_k} + \frac{\sigma^2\mathbf{I}_M}{SP_C\tau_S}\right)^{-1}\tilde{\mathbf{r}}_{0,k}^{tr}, \tag{3.20}$$

which is distributed as $\hat{\mathbf{h}}_{d,k} \sim \mathcal{CN}(\mathbf{0}, \mathbf{\Psi}_{d,k})$ *where*

$$\mathbf{\Psi}_{d,k} = \beta_{d,k}^2\mathbf{R}_{BS_k}\left(\beta_{d,k}\mathbf{R}_{BS_k} + \frac{\sigma^2\mathbf{I}_M}{SP_C\tau_S}\right)^{-1}\mathbf{R}_{BS_k}^H. \tag{3.21}$$

Proof. The proof follows from noting that the MMSE estimate is given as as $\hat{\mathbf{h}}_{d,k} = \mathbf{W}\tilde{\mathbf{r}}_{0,k}^{tr}$, where $\mathbf{W} = \mathbb{E}[\mathbf{h}_{d,k}\tilde{\mathbf{r}}_{0,k}^{tr^H}](\mathbb{E}[\tilde{\mathbf{r}}_{0,k}^{tr}\tilde{\mathbf{r}}_{0,k}^{tr^H}])^{-1}$. The expressions for $\mathbb{E}[\mathbf{h}_{d,k}\tilde{\mathbf{r}}_{0,k}^{tr^H}]$ and $\mathbb{E}[\tilde{\mathbf{r}}_{0,k}^{tr}\tilde{\mathbf{r}}_{0,k}^{tr^H}]$ can be derived by noting that $\mathbf{n}_k^{tr}$ and $\mathbf{h}_{d,k}$ are independent Gaussian vectors. More details can be found in Appendix A. □

The uncorrelated estimation error $\tilde{\mathbf{h}}_{d,k} = \hat{\mathbf{h}}_{d,k} - \mathbf{h}_{d,k}$ is also statistically independent of $\hat{\mathbf{h}}_{d,k}$ since both vectors are jointly Gaussian. Moreover, it is distributed as $\tilde{\mathbf{h}}_{d,k} \sim \mathcal{CN}(\mathbf{0}, \tilde{\mathbf{\Psi}}_{d,k})$, where $\tilde{\mathbf{\Psi}}_{d,k} = \beta_{d,k}\mathbf{R}_{BS_k} - \mathbf{\Psi}_{d,k}$.

Next we estimate $\mathbf{h}_{2,l,k} \in \mathbb{C}^{N\times 1}$ using the observation vector $\tilde{\mathbf{r}}_{l,k}^{tr}$. We extract $\bar{\mathbf{H}}_{1,l} = \text{diag}(\mathbf{h}_{1,l,1}, \ldots, \mathbf{h}_{1,l,N}) \in \mathbb{C}^{MN\times N}$ from $\bar{\mathbf{H}}_1$ and note that $\bar{\mathbf{H}}_{1,l}^H\bar{\mathbf{H}}_{1,l} = M\mathbf{\Sigma}_l = \beta_{1,l}M\mathbf{I}_N$ for $l = 1, \ldots, L$. The relationship between $\tilde{\mathbf{r}}_{l,k}^{tr} \in \mathbb{C}^{N\times 1}$ and $\mathbf{h}_{2,l,k}$s can now be written using (3.17) as

$$\begin{aligned}\tilde{\mathbf{r}}_{l,k}^{tr} &= \mathbf{h}_{2,l,k} + \frac{1}{SM}\mathbf{\Sigma}_l^{-1}\bar{\mathbf{H}}_{1,l}^H(\mathbf{V}_l^{tr}\otimes\mathbf{I}_M)^H\mathbf{n}_k^{tr} \\ &= \mathbf{h}_{2,l,k} + \frac{1}{SM\beta_{1,l}}\bar{\mathbf{H}}_{1,l}^H(\mathbf{V}_l^{tr}\otimes\mathbf{I}_M)^H\mathbf{n}_k^{tr},\end{aligned} \tag{3.22}$$

where $\mathbf{V}_l^{tr} = \mathbf{V}^{tr}(:, [N(l-1)+2 : Nl+1]) \in \mathbb{C}^{S\times N}$ for $l = 1, \ldots, L$ and $(:, \cdot)$ denotes all the rows for a particular column or set of columns of a matrix. Note that (3.22) is actually the LS estimate of $\mathbf{h}_{2,l,k}$ which we denote as $\hat{\mathbf{h}}_{2,l,k}^{LS}$.

Theorem 3.2. *The MMSE estimate* $\hat{\mathbf{h}}_{2,l,k}$ *of* $\mathbf{h}_{2,l,k}$ *is given as*

$$\hat{\mathbf{h}}_{2,l,k} = \beta_{2,l,k}\mathbf{R}_{IRS_{l,k}}\left(\beta_{2,l,k}\mathbf{R}_{IRS_{l,k}} + \frac{\sigma^2\mathbf{I}_N}{\beta_{1,l}SMP_C\tau_S}\right)^{-1}\tilde{\mathbf{r}}_{l,k}, \tag{3.23}$$

for $l = 1, \ldots, L$, $k = 1, \ldots, K$. *It is distributed as* $\hat{\mathbf{h}}_{2,l,k} \sim \mathcal{CN}(\mathbf{0}, \boldsymbol{\Psi}_{2,l,k})$, *where*

$$\boldsymbol{\Psi}_{2,l,k} = \beta_{2,l,k}^2 \mathbf{R}_{IRS_k} \left(\beta_{2,l,k} \mathbf{R}_{IRS_k} + \frac{\sigma^2 \mathbf{I}_N}{\beta_{1,l} SMP_C \tau_S} \right)^{-1} \mathbf{R}_{IRS_k}^H. \tag{3.24}$$

Proof. The proof is similar to that of $\hat{\mathbf{h}}_{d,k}$ and found in Appendix B. □

The uncorrelated estimation error $\tilde{\mathbf{h}}_{2,l,k} = \mathbf{h}_{2,l,k} - \hat{\mathbf{h}}_{2,l,k}$ is also statistically independent of $\hat{\mathbf{h}}_{2,l,k}$ since both vectors are jointly Gaussian. Moreover, it is distributed as $\tilde{\mathbf{h}}_{2,l,k} \sim \mathcal{CN}(\mathbf{0}, \tilde{\boldsymbol{\Psi}}_{2,l,k})$, where $\tilde{\boldsymbol{\Psi}}_{2,l,k} = \beta_{2,l,k} \mathbf{R}_{IRS_k} - \boldsymbol{\Psi}_{2,l,k}$.

In order to compute the MMSE estimates, the BS must have previous knowledge of the correlation matrices at the IRSs and the BS. However, it is well-known that the correlation matrices vary slowly and remain static over many coherence intervals. The BS can easily obtain the prior knowledge of these matrices since possessing second-order channel statistics at the BS is pervasive in massive MIMO literature [1].

3.4 Normalized Mean Squared Error Analysis

Here we analyze the NMSE in the LS and MMSE estimates under the developed CE protocol. Note that normalization occurs with respect to the channel power which is defined for the direct channel as $tr(\mathbb{E}[\mathbf{h}_{d,k}\mathbf{h}_{d,k}^H])$. The reason we compute the normalization is to enable comparison between the direct channel and the IRS-user channels and to provide a better visualization. We define the NMSE (e.g. for the direct channel) as

$$\text{NMSE}(\hat{\mathbf{h}}_{d,k}) = \frac{tr(\mathbb{E}[\tilde{\mathbf{h}}_{d,k}\tilde{\mathbf{h}}_{d,k}^H])}{tr(\mathbb{E}[\mathbf{h}_{d,k}\mathbf{h}_{d,k}^H])} = \frac{tr(\tilde{\boldsymbol{\Psi}}_{d,k})}{tr(\mathbb{E}[\mathbf{h}_{d,k}\mathbf{h}_{d,k}^H])}. \tag{3.25}$$

The NMSEs for the LS estimates in (3.19) and (3.22) are independent of the structure of $\mathbf{R}_{IRS_{l,k}}$ and $\mathbf{R}_{BS_k}$. The NMSE in the MMSE estimates does depend on these matrices, but we assume them to be identity matrices for the purpose of this section to obtain closed-form expressions and yield insights. The effect of correlation on NMSE is shown in the simulations later.

First, the NMSE in the MMSE estimate of $\mathbf{h}_{d,k}$ is derived using the fact

that $tr(\mathbb{E}[\mathbf{h}_{d,k}\mathbf{h}_{d,k}^H]) = tr(\beta_{d,k}\mathbf{I}_M) = M\beta_{d,k}$ as

$$\begin{aligned}\text{NMSE}(\hat{\mathbf{h}}_{d,k}) &= \frac{\beta_{d,k}}{M\beta_{d,k}}\Big(M - \beta_{d,k}tr\Big(\beta_{d,k}\mathbf{I}_M + \frac{\sigma^2\mathbf{I}_M}{SP_C\tau_S}\Big)^{-1}\Big) \\ &= \frac{1}{M\beta_{d,k}}\frac{M\beta_{d,k}\frac{\sigma^2}{SP_C\tau_S}}{\Big(\beta_{d,k} + \frac{\sigma^2}{SP_C\tau_S}\Big)} = \frac{\frac{\sigma^2}{SP_C\tau_S}}{\beta_{d,k} + \frac{\sigma^2}{SP_C\tau_S}}. \end{aligned} \tag{3.26}$$

The expression reveals that the NMSE increases to 1 as σ^2 grows large or S, P_C, τ_S grow small (using L'Hopital's rule). Moreover by increasing S, the quality of estimation can be improved but the training time will also increase. Similarly, the NMSE increases as $\beta_{d,k}$ decreases.

The NMSE in the LS estimate of $\mathbf{h}_{d,k}$ in (3.19) is obtained using $\mathbb{E}[\mathbf{n}_k^{tr}\mathbf{n}_k^{tr^H}] = \frac{\sigma^2 P_C\tau_S}{(P_C\tau_S)^2}\mathbf{I}_{MS}$ and $\mathbf{v}_1^{tr^H}\mathbf{v}_1^{tr} = S$ as

$$\text{NMSE}(\hat{\mathbf{h}}_{d,k}^{LS}) = \frac{\sigma^2 P_C\tau_S tr\left((\mathbf{v}_1^{tr}\otimes\mathbf{I}_M)^H(\mathbf{v}_1^{tr}\otimes\mathbf{I}_M)\right)}{M\beta_{d,k}S^2(P_C\tau_S)^2} \tag{3.27}$$

$$= \frac{\sigma^2}{\beta_{d,k}SP_C\tau_S}. \tag{3.28}$$

We see that the NMSE in LS estimate will grow unboundedly if either σ^2 grows large, or the argument in the denominator grows too small. The difference in the NMSE of MMSE and LS estimates can be straightforwardly calculated to be
$\text{NMSE}(\hat{\mathbf{h}}_{d,k}^{LS}) - \text{NMSE}(\hat{\mathbf{h}}_{d,k})=$

$$\frac{\sigma^4}{(\beta_{d,k}SP_C\tau_S)^2 + \beta_{d,k}SP_C\tau_S\sigma^2} \geq 0, \tag{3.29}$$

for all $\beta_{d,k}$, P_C, τ_S, and S. Therefore the MMSE estimate of the direct channel will outperform the LS estimate. We can also observe the growth behavior of both NMSEs with respect to noise and see that the NMSE in MMSE estimate grows slower than that in the LS estimate, making MMSE robust even at low values of signal-to-noise ratio (SNR).

The NMSE in the estimates of $\mathbf{h}_{2,l,k}$s can be derived similarly by noting that from (2.3) that under $\mathbf{R}_{IRS_{l,k}} = \mathbf{I}_N$, we have $tr(\mathbb{E}[\mathbf{h}_{2,l,k}\mathbf{h}_{2,l,k}^H]) = tr(\beta_{2,l,k}\mathbf{I}_N) = N\beta_{2,l,k}$. Therefore $\text{NMSE}(\hat{\mathbf{h}}_{2,l,k}) = \frac{1}{N\beta_{2,l,k}}tr(\tilde{\mathbf{\Psi}}_{2,l,k})$ and we

get $\text{NMSE}(\hat{\mathbf{h}}_{2,l,k}) =$

$$\frac{\beta_{2,l,k}}{N\beta_{2,l,k}}\Big(N - \beta_{2,l,k}tr\Big(\beta_{2,l,k}\mathbf{I}_N + \frac{\sigma^2\mathbf{I}_N}{\beta_{1,l}SMP_C\tau_S}\Big)^{-1}\Big),$$
$$= \frac{\frac{\sigma^2}{SMP_C\tau_S}}{\beta_{2,l,k}\beta_{1,l} + \frac{\sigma^2}{SMP_C\tau_S}}. \tag{3.30}$$

Also, the NMSE in LS estimate in (3.22) can be shown to be

$$\text{NMSE}(\hat{\mathbf{h}}^{LS}_{2,l,k}) = \frac{1}{\beta_{2,l,k}\beta_{1,l}}\frac{\sigma^2}{SMP_C\tau_S}. \tag{3.31}$$

The performance gap between the MMSE and LS estimate of $\mathbf{h}_{2,l,k}$ is $\text{NMSE}(\hat{\mathbf{h}}^{LS}_{2,l,k}) - \text{NMSE}(\hat{\mathbf{h}}_{2,l,k}) =$

$$= \frac{\sigma^4}{(\beta_{2,l,k}\beta_{1,l}SMP_C\tau_S)^2 + \beta_{2,l,k}\beta_{1,l}SMP_C\tau_S\sigma^2} \geq 0, \tag{3.32}$$

because $\beta_{2,l,k}$, $\beta_{1,l}$ and M are all non-negative. Therefore the MMSE estimate of $\mathbf{h}_{2,l,k}$ outperforms the LS estimate.

3.5 Benchmark protocol

As a benchmark (BM), we extend the protocol in [1] for a single IRS to the distributed IRSs setup. The derivation is similar to [1], and is repeated here for completeness. Under this protocol, we derive the estimates of the cascaded IRS-assisted channels, i.e. the columns of $\mathbf{H}_{0,l,k} = \mathbf{H}_{1,l}\text{diag}(\mathbf{h}_{2,l,k}) \in \mathbb{C}^{M\times N}$, without exploiting full knowledge of $\mathbf{H}_{1,l}$s. As a result, (3.3) is written as

$$\mathbf{r}^{BM}_{s,k} = (\mathbf{h}_{d,k} + \mathbf{H}_{0,k}\mathbf{v}_s) + \mathbf{n}^{tr}_{s,k},\ k = 1,\ldots,K, \tag{3.33}$$

where $\mathbf{H}_{0,k} = \mathbf{H}_1\text{diag}(\mathbf{h}_{2,k}) \in \mathbb{C}^{M\times NL}$ is the cascaded matrix of all IRS-assisted channel vectors and $\mathbf{v_s} = \text{diag}(\mathbf{\Theta}_s) \in \mathbb{C}^{NL\times 1}$ is the concatenated beamforming vector for all the IRSs in subphase s. Note that $\mathbf{H}_1$, $\mathbf{h}_{2,k}$ and $\mathbf{\Theta}_s$ are defined after (3.3). We let $\bar{\mathbf{h}}_{d,2,k} = [\mathbf{h}^T_{d,k}, \mathbf{h}^T_{0,1,1,k}, \ldots, \mathbf{h}^T_{0,1,N,k}, \mathbf{h}^T_{0,2,1,k}, \ldots, \mathbf{h}^T_{0,L,N,k}]^T \in \mathbb{C}^{M(NL+1)\times 1}$, where $\mathbf{h}_{0,l,n,k}$ is the n^{th} column of $\mathbf{H}_{0,l,k}$, and collect the observation vectors across S subphases to obtain

$$\mathbf{r}^{BM}_k = (\mathbf{V}_{tr}\otimes\mathbf{I}_M)\bar{\mathbf{h}}_{d,2,k} + \mathbf{n}^{tr}_k,\ k = 1,\ldots,K, \tag{3.34}$$

where $\mathbf{V}_{tr}$ is given as (3.6) and $\mathbf{n}_k^{tr}$ is the same as in (3.5). To ensure that the left pseudo-inverse exists for $\mathbf{V}_{tr}$, we must have $S \geq NL+1$. Compared with the proposed protocol's condition of $S \geq \frac{NL}{M}+1$, the benchmark would impose almost an M times larger training overhead which will compromise the downlink performance of the system.

With optimal design for $\mathbf{V}_{tr}$ found to be the DFT matrix, we can perform the pseudo-inverse operation on (3.34) and get

$$\tilde{\mathbf{r}}_k^{BM} = \bar{\mathbf{h}}_{d,2,k} + \frac{1}{S}(\mathbf{V}_{tr} \otimes \mathbf{I}_M)^H \mathbf{n}_k^{tr}. \tag{3.35}$$

Write $\tilde{\mathbf{r}}_k^{BM} = [\tilde{\mathbf{r}}_{0,k}^{BM^T}, \tilde{\mathbf{r}}_{1,1,k}^{BM^T}, \ldots, \tilde{\mathbf{r}}_{1,N,k}^{BM^T}, \ldots, \tilde{\mathbf{r}}_{L,1,k}^{BM^T}, \ldots, \tilde{\mathbf{r}}_{L,N,k}^{BM^T}]^T \in \mathbb{C}^{M(NL+1)\times 1}$, we can obtain the MMSE estimates of $\mathbf{h}_{d,k}$ and $\mathbf{h}_{0,l,n,k}$ as follows

$$\hat{\mathbf{h}}_{d,k}^{BM} = \beta_{d,k}\mathbf{R}_{BS_k}\mathbf{Q}_{d,k}\tilde{\mathbf{r}}_{0,k}^{BM}, \tag{3.36}$$

where $\mathbf{Q}_{d,k} = \left(\beta_{d,k}\mathbf{R}_{BS_k} + \frac{\sigma^2}{SP_C\tau_S}\mathbf{I}_M\right)^{-1}$, and

$$\hat{\mathbf{h}}_{0,l,n,k}^{BM} = r_{l,n,k}\beta_{l,k}\mathbf{h}_{1,l,n}\mathbf{h}_{1,l,n}^H\mathbf{Q}_{l,n,k}\tilde{\mathbf{r}}_{l,n,k}^{BM}, \tag{3.37}$$

for $n = 1, \ldots, N$, $l = 1, \ldots, L$, $k = 1, \ldots, K$, where

$$\mathbf{Q}_{l,n,k} = \left(r_{l,n,k}\beta_{l,k}\mathbf{h}_{1,l,n}\mathbf{h}_{1,l,n}^H + \frac{\sigma^2}{SP_C\tau_S}\mathbf{I}_M\right)^{-1}, \tag{3.38}$$

and $\beta_{l,k} = \beta_{1,l}\beta_{2,l,k}$, $r_{l,n,k}$ is the $(n,n)^{th}$ entry of the matrix $\mathbf{R}_{IRS_{l,k}}$ and $\mathbf{h}_{1,l,n}$ is the n^{th} column of $\beta_{1,l}^{-1/2}\mathbf{H}_{1,l}$.

Note that with $\hat{\mathbf{H}}_{0,l,k}^{BM} = [\hat{\mathbf{h}}_{0,l,1,k}^{BM}, \ldots, \hat{\mathbf{h}}_{0,l,N,k}^{BM}] \in \mathbb{C}^{M\times N}$, the expression for $\hat{\mathbf{h}}_{2,l,k}^{BM}$ can be recovered under the benchmark protocol through the relationship $\hat{\mathbf{H}}_{0,l,k}^{BM} = \mathbf{H}_{1,l}\text{diag}(\hat{\mathbf{h}}_{2,l,k}^{BM})$, when $\mathbf{H}_{1,l}$ is known apriori.

This is done for comparison with the proposed protocol where we directly estimate $\mathbf{h}_{2,l,k}$s.

3.6 Simulations and Discussion

In the simulation, we set $N = 32$, $L = 4$, $M = 8$, $P_C = 1W$, and $\tau_S = 50\mu s$. Denoting (x, y) metres (m) as the coordinates, the BS is placed at (0,0)m and the four IRSs are placed at (0,100), (100,0), (-100,0) and (0,-100)m. The path loss factors are computed as $\beta_{d,k} = \frac{10^{-2.8}}{d_{BS-user}^{3.67}}$, $\beta_{2,l,k} =$

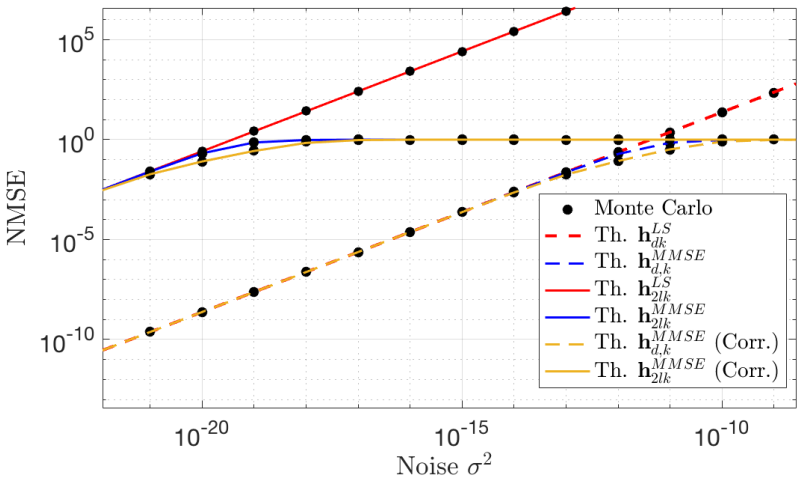

Figure 3.2: Normalized Mean Square Error vs noise. Th. refers to the theoretical expressions given. Corr. refers to the scenario of correlated Rayleigh fading channels.

$\frac{10^{-2.8}}{d_{IRSl-user}^{3.67}}$ which are non-LoS (NLoS) path loss factors computed at 2.5 GHz carrier frequency for the 3GPP Urban Micro (UMi) scenario from TR36.814 (also found in Section V of [1]), and $\beta_{1,l} = \frac{10^{-2.6}}{d_{BS-IRSl}^{2.2}}$ is the LoS path loss factor, where $d_{(.)}$ stands for Euclidean distance for a particular link. The number of sub-phases is set as $S = \frac{NL}{M} + 1 = 17$ for the proposed protocol. Fig. 3.2 shows the NMSE curves for both MMSE and LS estimates of $\mathbf{h}_{d,k}$ and $\mathbf{h}_{2,l,k}$ (for the latter we plot $\frac{1}{L}\sum_{l=1}^{L} \text{NMSE}(\hat{\mathbf{h}}_{2,l,k})$). The Monte-Carlo simulated values match the theoretical analysis in Sec. III-C. As expected the MMSE estimates performs better than LS estimates especially in the high noise regime. Also, the NMSE in $\mathbf{h}_{2,l,k}$ is higher than that in $\mathbf{h}_{d,k}$ due to the multiplicative path loss factor $\beta_{1,l}\beta_{2,l,k}$ in (3.30) and (3.31).

The orange curves in the figure show the NMSE for the correlated scenario where $[\mathbf{R}_{IRS_{l,k}}]_{i,j} = \eta^{|i-j|}$ (similar definition for $\mathbf{R}_{BS_k}$) and $\eta = 0.95$. The NMSE in the LS estimates is unaffected (not plotted for clarity of figures) while that in the MMSE estimates is actually less for the correlated case [1].

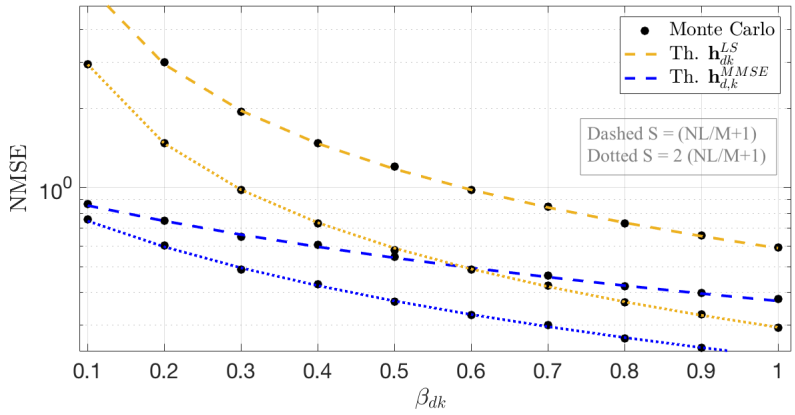

Figure 3.3: NMSE vs path loss factor for direct channel. Dashed lines represent a smaller number of sub-phases S than the dotted lines.

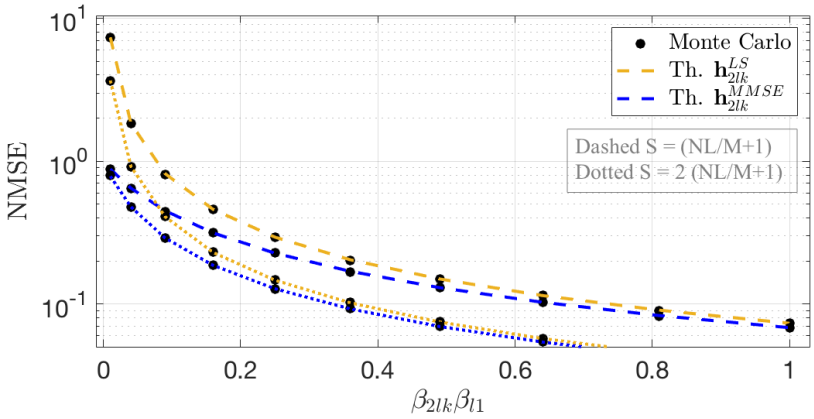

Figure 3.4: NMSE vs path loss factor for IRS assisted channels.

The NMSE vs the path loss factor is also plotted for the direct channel and IRS assisted channel in Fig. 3.3 and Fig. 3.4 respectively, where it can be seen that the LS estimate performs worse than the MMSE estimates for all values of path loss factor. Increasing S reduces the NMSE in both LS and MMSE estimates. However, increasing S means increasing training overhead, which reduces the time left for downlink transmission.

Table 3.1: Estimation accuracy per unit time $\frac{1}{NMSE \times \tau_C}(s^{-1})$

L	**2**	**6**	**8**	**10**
Proposed ($\times 10^5$)	1.809	1.784	1.776	1.782
Benchmark ($\times 10^5$)	1.767	1.777	1.766	1.769

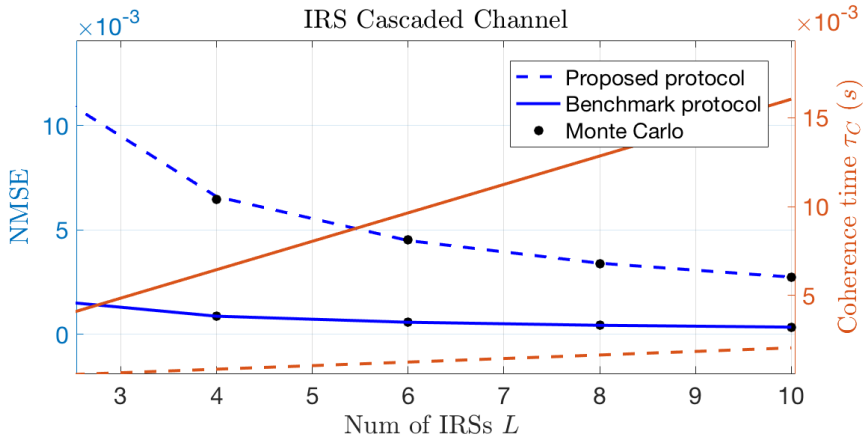

Figure 3.5: Comparison of proposed and benchmark protocols. The black dots are the Monte Carlo simulations of NMSE. Pink lines correspond to channel estimation time for the proposed protocol (dashed) and benchmark protocol (solid). Blue lines correspond to NMSE for both protocols.

In Fig. 3.5, we compare the proposed protocol which needs $S = \frac{NL}{M} + 1$ sub-phases to the benchmark protocol which needs $S = NL + 1$ sub-phases, in terms of NMSE and training time versus L. To ensure a fair comparison, both protocols are tested for the NMSE for the cascaded IRS-assisted channel. The NMSE in the benchmark protocol is lower due to it using a larger S for estimation, with the gap between the two NMSE curves becoming smaller with L. On the other hand, the time required for CE grows linearly with the number of IRSs in the benchmark protocol, which is detrimental for the overall downlink performance, while that for the proposed protocol grows with a much smaller slope. In fact, the additional gap in required training time for the BS relative to the proposed scheme becomes significantly large as L (or N) increase. To better capture this trade-off, we tabulate in Tab. 3.1 a figure of merit defined as $\frac{1}{NMSE \times \tau_C}(s^{-1})$, which can be interpreted as the measure of estimation accuracy per unit training time. Ideally, its value

should be high. We observe that the proposed protocol actually performs better in terms of estimation accuracy per unit training time for all L. This makes the proposed protocol scalable to large-scale distributed deployment of IRSs.

In the next chapter, we derive the average SINR and formulate an optimization problem dedicated to solving for optimized IRS-user association parameters.

Chapter 4

Average SINR and IRS-User Association

With distributed IRSs deployed in a multi-user system, a natural next step is to associate certain IRSs to different users in an optimal manner to achieve a performance objective such as max-min average signal-to-interference and noise ratio (SINR). The association implies each IRS will be *tuned* to a particular user. The motivation behind single-user association with an IRS is that it is user-centric and simplifies the optimal phase shifts design since the design can be based on the local channel state information (CSI) of the associated user, thus reducing complexity.

This chapter focuses on the average (ergodic) analysis of the SINR at each user in the distributed IRSs-assisted multi-user MISO system, where each IRS is associated with a user in the system. For a particular IRS-user association pair, we choose the design for IRS phase shifts that would maximize the received signal strength at that user. Under that design, we utilize statistical tools to obtain a closed form expression for the average SINR at each user. Then, we develop average SINR expression under IRS-user association parameters, optimized IRS passive beamforming, and maximum ratio transmission (MRT) precoding at the BS. Next, we formulate and solve a max-min average SINR optimization problem to find optimal IRS-user association.

While the CE scheme provided in the previous chapter can be applied in theory to any channel model, in this chapter, we focus on uncorrelated fading at the BS and IRS to obtain tractable closed-form expression for the average SINR under the umbrella of perfect CSI. However, this analysis is extendible (with some tedious algebraic manipulations) to correlated Rayleigh fading and imperfect CSI. We formally state the assumptions made on the system model below.

Assumption 1. *The correlation matrices of the direct and IRS-user channel are set to identity i.e. ($\mathbf{R}_{IRS_{l,k}} = \mathbf{I}_N$ and $\mathbf{R}_{BS_k} = \mathbf{I}_M$) which shifts the model to uncorrelated Rayleigh fading.*

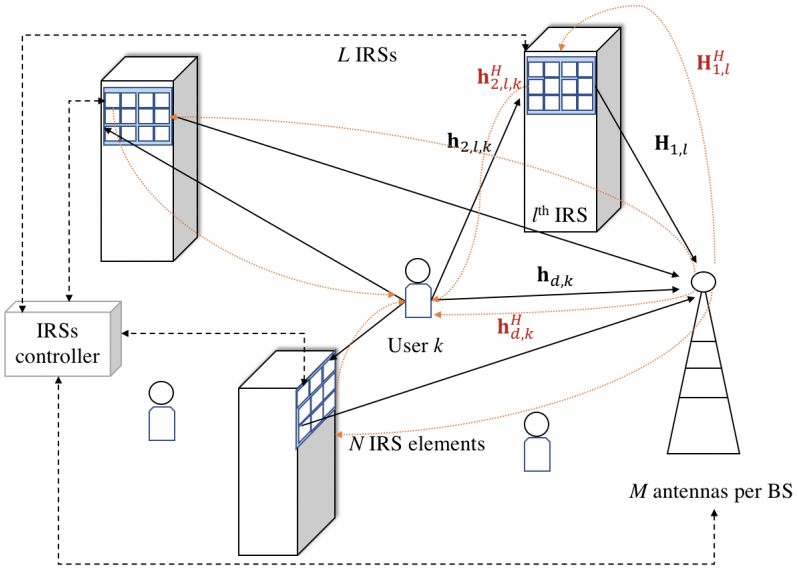

Figure 4.1: Distributed IRSs MISO system model. Red lines are for downlink transmission.

Assumption 2. *The analysis is done under the assumption that the BS has perfect CSI to design the precoder and IRS reflect beamforming vectors.*

Assumption 2 is made to ensure a tractable analysis since the distributed IRSs MISO system is complex where many links are involved. The results we obtain are relevant since they pave the way for imperfect CSI analysis and can serve as meaningful upper bounds for imperfect CSI scenarios.

Assumption 3. *The reflection coefficients* $\alpha_{l,n} = 1, \forall n, \forall l$.

Under idealized conditions, metasurfaces can achieve unity magnitude amplitude in the scattered field even with arbitrary phase [39]. Also, most current work on IRSs make this assumption.

4.1 Downlink Transmission

Recall from Chapter 2 that in the downlink, the k^{th} user receives

$$y_k = \mathbf{h}_{d,k}^H \mathbf{x} + \sum_{l=1}^{L} \mathbf{h}_{2,l,k}^H \boldsymbol{\Theta}_l^H \mathbf{H}_{1,l}^H \mathbf{x} + n_k, \tag{4.1}$$

where $n_k \sim \mathcal{CN}(0, \sigma^2)$ is the channel noise and the channels are given in (2.3), (2.4), and (2.5), as well as the transmit signal vector $\mathbf{x}$ has the structure shown in (2.12). The downlink transmission scheme is described elaborately in Chapter 2 and illustrated in Fig. 4.1. To better visualize the instantaneous SINR expression γ_k for user k, we can separate the received signal y_k into a desired term y_{kD} and interference and noise term y_{kIN} written as

$$y_{kD} = \sqrt{p}_k \mathbf{h}_{d,k}^H \mathbf{f}_k s_k + \sqrt{p}_k \sum_{l=1}^{L} \mathbf{h}_{2,l,k}^H \mathbf{\Theta}_l^H \mathbf{H}_{1,l}^H \mathbf{f}_k s_k, \tag{4.2}$$

$$y_{kIN} = \sum_{t \neq k, t=1}^{K} \left(\sqrt{p}_t \mathbf{h}_{d,k}^H \mathbf{f}_t s_t + \sqrt{p}_t \sum_{l=1}^{L} \mathbf{h}_{2,l,k}^H \mathbf{\Theta}_l^H \mathbf{H}_{1,l}^H \mathbf{f}_t s_t \right) + n_k, \tag{4.3}$$

The direct and IRS cascaded channels in the downlink can be rewritten collectively as

$$\mathbf{h}_k^H = \mathbf{h}_{d,k}^H + \sum_{l=1}^{L} \mathbf{h}_{2,l,k}^H \mathbf{\Theta}_l^H \mathbf{H}_{1,l}^H \tag{4.4}$$

$$= \mathbf{h}_{d,k}^H + \sum_{l=1}^{L} \mathbf{v}_l^H \mathbf{H}_{0,l,k}^H, \tag{4.5}$$

where

$$\mathbf{H}_{0,l,k} = \mathbf{H}_{1,l} \text{diag}(\mathbf{h}_{2,l,k}) = \sqrt{\beta_{1,l}} \mathbf{a}_l \mathbf{b}_l^H \text{diag}(\mathbf{h}_{2,l,k}) \tag{4.6}$$

Note that $\mathbf{H}_{0,l,k} \in \mathbb{C}^{M \times N}$ and $\mathbf{v}_l = \text{vec}(\mathbf{\Theta}_l) \in \mathbb{C}^{N \times 1}$, where $\text{vec}(\cdot)$ is used to create a vector from a diagonal matrix, and $\text{diag}(\cdot)$ creates a diagonal matrix out of a vector.

At this stage, we can introduce association variables $\lambda_{l,k} \in \{0, 1\}$ which are binary variables denoting association between the l^{th} IRS and k^{th} user. The collected channel (4.4) with association variables can be reformulated as

$$\mathbf{h}_k^H(\boldsymbol{\lambda}_k) = \mathbf{h}_{d,k}^H + \sum_{l=1}^{L} \lambda_{l,k} \mathbf{h}_{2,l,k}^H \mathbf{\Theta}_l^{k_l H} \mathbf{H}_{1,l}^H + \sum_{l=1}^{L} (1 - \lambda_{l,k}) \mathbf{h}_{2,l,k}^H \mathbf{\Theta}_l{}^H \mathbf{H}_{1,l}^H, \tag{4.7}$$

$$= \mathbf{h}_{d,k}^H + \sum_{l=1}^{L} \lambda_{l,k} \mathbf{v}_l^{k_l H} \mathbf{H}_{0,l,k}^H + \sum_{l=1}^{L} (1 - \lambda_{l,k}) \mathbf{v}_l^H \mathbf{H}_{0,l,k}^H, \tag{4.8}$$

where $\boldsymbol{\lambda}_k \in \mathbb{B}^{1\times L}$ is the row vector of associated or non-associated IRSs with respect to user k. To make the association between the optimized (tuned) beamforming vector $\mathbf{v}_l$ to a particular user k, we replace $\mathbf{v}_l$ by $\mathbf{v}_l^{k_l}$ henceforth, where k_l is the value of k for which $\lambda_{l,k} = 1$. In addition, we can rewrite (4.1) with association parameters, simply as

$$y_k = \mathbf{h}_k^H(\boldsymbol{\lambda}_k)\mathbf{x} + n_k. \tag{4.9}$$

We find optimal $\mathbf{v}_l^{k_l H}$ for IRS l associated with user k such that $\lambda_{l,k} = 1$ as to maximize the channel gain via that IRS to that user. Note that for the multi-user case (which we have), there is no known closed-form optimal solution for $\mathbf{v}_l^{k_l H}$ that maximizes the SINR. Furthermore, finding $\mathbf{v}_l^{k_l H}$ for each IRS l such that SINR at its associated user is maximized will result in a highly intractable joint optimization problem involving the beamforming vectors of all IRSs and their association parameters with the users. Therefore, we assume that for an IRS l associated with user k the channel gain is maximized via that IRS to the user, similar to [29]. Hence, the passive beamforming optimization problem for an IRS l associated with user k can be defined as

$$(P0) \quad \max_{\mathbf{v}_l} \quad \|\mathbf{h}_{d,k}^H + \mathbf{v}_l^H \mathbf{H}_{0,l,k}^H\|^2 \tag{4.10}$$

$$\text{s.t.} \quad |v_{l,n}| = 1, \forall n. \tag{4.11}$$

Note that $P0$ finds the optimized beamforming vector $\mathbf{v}_l^{k_l}$ for user k, when $\lambda_{l,k} = 1$. Expanding the objective function yields $\|\mathbf{h}_{d,k}^H\|^2 + 2\langle \mathbf{v}_l^H \mathbf{H}_{0,l,k}^H, \mathbf{h}_{d,k}^H\rangle + \|\mathbf{v}_l^H \mathbf{H}_{0,l,k}^H\|^2$, where we can drop the first term since it does not depend on the optimization variable to get

$$(P0') \quad \max_{\mathbf{v}_l} \quad 2\langle \mathbf{v}_l^H \mathbf{H}_{0,l,k}^H, \mathbf{h}_{d,k}^H\rangle + \|\mathbf{v}_l^H \mathbf{H}_{0,l,k}^H\|^2 \tag{4.12}$$

$$\text{s.t.} \quad |v_{l,n}| = 1, \forall n. \tag{4.13}$$

To make this more general, we drop the subscripts and define

$$(P0") \quad \max_{\mathbf{v}} \quad 2\langle \mathbf{v}^H \mathbf{H}^H, \mathbf{h}^H\rangle + \|\mathbf{v}^H \mathbf{H}^H\|^2 \tag{4.14}$$

$$\text{s.t.} \quad |v| = 1, \forall n. \tag{4.15}$$

Proposition 4.1. *Under Assumptions 2 and 3, and by copying the structure of* $\mathbf{H} = \mathbf{a}\mathbf{b}^H diag(\mathbf{c})$ *as in* (4.6) *and* $\mathbf{c} = \sqrt{\beta_{1,l}} diag(\mathbf{h}_{2,l,k})$*, we can formulate the optimal solution of* $(P0")$ *as*

$$\mathbf{v}^{k_l} = \exp(j\angle diag(\mathbf{c}^H)\mathbf{b}) \exp(j\angle \mathbf{a}^H \mathbf{h}) \tag{4.16}$$

Proof. To find the angle of a complex (Hermitian) inner product, it suffices to find $\Re(\langle \mathbf{v}^H\mathbf{H}^H, \mathbf{h}^H\rangle) = \sum_{m=1}^{M} |g_m||h_m| \cos(\angle g_m - \angle h_m)$. Where $\Re(\langle\cdot,\cdot\rangle)$ denotes the real part of the inner product of two complex vectors, and g_m (h_m) is the m^{th} component of $\mathbf{g} = \mathbf{H}\mathbf{v}$ ($\mathbf{h}$). This expression achieves its maximum value when $\angle g_m = \angle h_m$, equivalently ($\angle \mathbf{v}^H \text{diag}(\mathbf{c}^H)\mathbf{b}\mathbf{a}^H = \angle \mathbf{h}^H$) and $\angle\cdot$ for element-wise phase of the vector. Noting that when we choose $\mathbf{v}^H$ as in (4.16) we satisfy the relation, since we achieve ($\angle \alpha_v \mathbf{h} = \angle \mathbf{h}$), where $\alpha_v = \|\text{diag}(\mathbf{c}^H)\mathbf{b}\|^2\|\mathbf{a}\|^2$ is a scalar.

Similar argument can be done for the second term in the objective function (4.14), where $\max_{\mathbf{v}} \quad \|\mathbf{v}^H\mathbf{H}^H\|^2$

$$= \max_{\mathbf{v}} \quad \mathbf{v}^H \text{diag}(\mathbf{c}^H)\mathbf{b}\mathbf{a}^H\mathbf{a}\mathbf{b}^H\text{diag}(\mathbf{c})\mathbf{v} \tag{4.17}$$

$$= \max_{\mathbf{v}} \quad \|\mathbf{a}\|^2\mathbf{v}^H \text{diag}(\mathbf{c}^H)\mathbf{b}\mathbf{b}_l^H\text{diag}(\mathbf{c})\mathbf{v}, \tag{4.18}$$

$$= \max_{\mathbf{v}} \quad \|\mathbf{a}\|^2|\mathbf{v}^H \text{diag}(\mathbf{c}^H)\mathbf{b}|^2, \tag{4.19}$$

$$= \max_{v_n} \quad \|\mathbf{a}\|^2\left|\sum_{n=1}^{N} |v_n||c_n||b_n| \exp(j(\angle v_n^* + \angle c_n^* + \angle b_n))\right|^2, \tag{4.20}$$

$$= \|\mathbf{a}\|^2(\sum_{n=1}^{N} |c_n||b_n|)^2, \tag{4.21}$$

which achieves it maximum value with optimal reflection beamforming vector given in (4.22). Note that c_n^* is the complex conjugate of the complex number c_n. Moreover, b_n, v_n^*, and c_n denote the n^{th} elements of $\mathbf{b}$, $\mathbf{v}^H$, and $\mathbf{c}$, respectively.

□

We can now write the general result in (4.16) with proper notation for the beamforming vector of IRS l dedicated to user k, such that $\lambda_{l,k} = 1$, as

$$\mathbf{v}_l^{k_l} = \exp(j\angle \text{diag}(\mathbf{h}_{2,l,k}^H)\mathbf{b}_l) \exp(j\angle \mathbf{a}_l^H \mathbf{h}_{d,k}). \tag{4.22}$$

To shed some light on how to maximize a squared norm given unit-modulus constraint, a the second term in the objective function (4.12) is plotted in Fig. 4.2 with optimal reflection beamforming vector and (4.20) is also plotted against $-1 \leq \exp(j(\angle {v^*}_n + \angle c_n^* + \angle b_n)) \leq 1$ encompassing the sum of phases of all the terms inside the squared norm. Furthermore, (4.21) is plotted as a validation step.

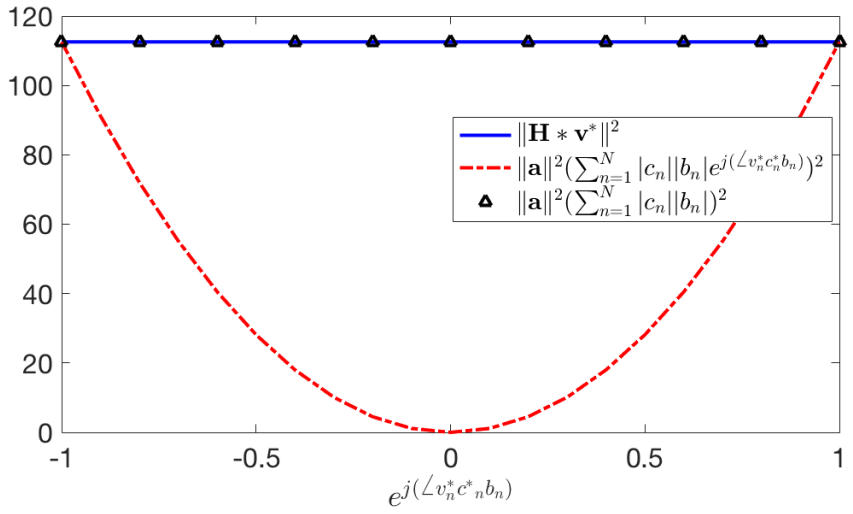

Figure 4.2: Maximizing a squared norm with unit-modulus constraints. This shows that the solution for reflection beamforming vector is the optimal solution.

Once we substitute the optimized beamforming vector given in (4.22) in (4.8) we obtain

$$\begin{aligned}\mathbf{h}_k(\boldsymbol{\lambda}_k) = |\mathbf{h}_{d,k}^H| \cdot e^{j\angle \mathbf{h}_{d,k}^H} + \\ \sum_{l=1}^{L} \lambda_{l,k} e^{j\angle \mathbf{h}_{d,k}^H \mathbf{a}_l} \sum_{n=1}^{N} |h_{2,l,k,n}||b_{l,n}|\mathbf{a}_l^H \\ + \sum_{l=1}^{L} (1-\lambda_{l,k}) \mathbf{v}_l^H \mathbf{H}_{0,l,k}^H\end{aligned} \tag{4.23}$$

where $\angle$ and $|.|$ refer to element wise phase and magnitude of the vectors. An intuitive observation to make is that this solution for $\mathbf{v}_l^{k_l}$ negates the phases from the the cascaded IRS channel $\mathbf{H}_{0,l,k}^H$ and lies in-phase with the direct channel $\mathbf{h}_{d,k}^H$ which indicates that the signal coming from the direct channel and from the cascaded IRS channels at the associated user will combine coherently and constructively.

4.2 Average SINR

The SINR γ_k at user k under the transmission model just described with IRS-user association parameters is given as

$$\gamma_k = \frac{p_k \|\mathbf{h}_k^H(\boldsymbol{\lambda}_k)\mathbf{f}_k(\boldsymbol{\lambda}_k)\|^2}{\sum_{t\neq k, t=1}^{K} p_t \|\mathbf{h}_k^H(\boldsymbol{\lambda}_k)\mathbf{f}_t(\boldsymbol{\lambda}_t)\|^2 + \sigma^2}. \tag{4.24}$$

In general, this SINR can not be perfectly computed at the user except when perfect CSI is available at the receiver which is a moot possibility. The downlink SINR is an important metric of performance which can be fed back to the BS, or the BS can discover it given sufficient information of the channels, and aids in the decision-making process for certain hyper-parameters of the distributed IRS assisted MISO system: BS active beamforming, power control at the BS, and IRSs passive beamforming. In this work, we focus on the latter which can split into IRSs re-configurability and IRS-user association, and assume the previous two are given or fixed.

Since practical wireless channels are often modelled having random fading (c.f. sec. 2.2), we direct our attention to average SINR $\bar{\gamma}_k$ which can be tracked at the user and depends only on large-scale statistics of the channels. Large-scale statistics such as path loss factor and correlation matrices change slowly as compared to the channels themselves. Therefore, the user will not need to compute the instantaneous SINR in each coherence interval (more on coherence interval is found in sec. 2.2) and instead use $\bar{\gamma}_k$ as metric for performance. By using $\bar{\gamma}_k$ as a performance metric, the association parameters will not need to be updated on the coherence time scale but after multiple coherence intervals depending on the varying large-scale statistics of the channels. Each IRS will then use local CSI of the associated user to determine its optimal configuration as in (4.22).

Next, we formulate Lemma 4.2 which conveys the expectation of a ratio of terms defining the structure of SINR.

Lemma 4.2. *The average SINR $\bar{\gamma}_k$ can be approximated as*

$$\mathbb{E}[\gamma_k] \approx \frac{p_k \mathbb{E}[\|\mathbf{h}_k^H(\boldsymbol{\lambda}_k)\mathbf{f}_k(\boldsymbol{\lambda}_k)\|^2]}{\sum_{t\neq k, t=1}^{K} p_t \mathbb{E}[\|\mathbf{h}_k^H(\boldsymbol{\lambda}_k)\mathbf{f}_t(\boldsymbol{\lambda}_t)\|^2] + \sigma^2} \tag{4.25}$$

Proof. For average SINR, we obtain the expectation of a ratio, where $\mathbb{E}[\gamma_k] = \mathbb{E}[\frac{X}{Y}]$. Now assume $\mathbb{E}[X] = \mu_X$ and $\mathbb{E}[Y] = \mu_Y$, we can expand the ratio (using the bivariate first-order Taylor series expansion) around the point

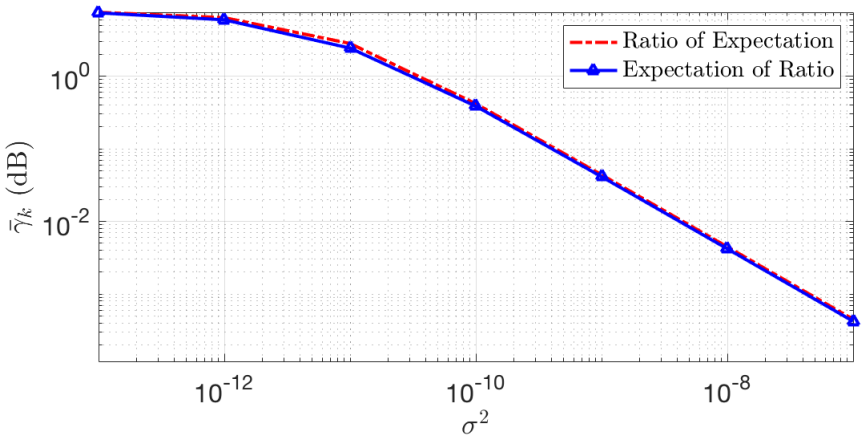

Figure 4.3: Tightness of approximation for $\bar{\gamma}_k$

(μ_X, μ_Y) so that

$$\frac{X}{Y} = \frac{\mu_X}{\mu_Y} + \frac{1}{\mu_Y}(X - \mu_X) - \frac{\mu_X}{\mu_Y^2}(Y - \mu_Y) + R \tag{4.26}$$

where R is the remainder of the the other order terms in the expansion. Taking the first moment of this ratio provides

$$\mathbb{E}\left[\frac{X}{Y}\right] = \mathbb{E}\left[\frac{\mu_X}{\mu_Y} + \frac{1}{\mu_Y}(X - \mu_X) - \frac{\mu_X}{\mu_Y^2}(Y - \mu_Y) + R\right] \tag{4.27}$$

$$\underset{(i)}{\approx} \mathbb{E}\left[\frac{\mu_X}{\mu_Y}\right] + \mathbb{E}\left[\frac{1}{\mu_Y}(X - \mu_X)\right] - \mathbb{E}\left[\frac{\mu_X}{\mu_Y^2}(Y - \mu_Y)\right] \tag{4.28}$$

$$\underset{(ii)}{=} \frac{\mathbb{E}[X]}{\mathbb{E}[Y]} \tag{4.29}$$

where the last line is a ratio of expectations, (i) follows from dropping the lesser order terms R, and (ii) follows since the expectation of the second and third terms is zero. □

Remark 1 This approximation for $\mathbb{E}[\gamma_k]$ is justified as follows: since X and Y contain quadratic forms, the variance relative to the mean of these random variables is small [40], [41], [42]. Hence, this approximation is tight (as also verified numerically using the same set-up in Tab. 4.1 and illustrated in Fig. 4.3).

Table 4.1: Simulation parameters.

Parameter	**Value**
Array parameters:	
BS configuration	Uniform linear array
IRS configuration	Uniform planar array
Antenna gain	5dBi
d_{BS}, d_{IRS}	0.5λ
Noise level	-60dBm
Path Loss:	
Model	$\frac{10^{-C/10}}{d^{\alpha}}$
C (Fixed loss at $d = 1$m)	25dB (β_1), 30dB ($\beta_{2,k}, \beta_{d,k}$)
α (Path loss exponent)	2.2 (β_1), 3.67 ($\beta_{2,k}, \beta_{d,k}$)
Penetration Loss:	
$\mathbf{h}_{d,k}$	20dB
$\mathbf{h}_{2,l,k}$	5dB
System Dimensions:	
N	$8, 16$
L	8
K	4
M	16

4.2.1 Performance Under Maximum Ratio Transmission

With maximum ratio transmission (MRT) and using Assumption 2, the digital precoder is found to be

$$\mathbf{f}_k(\boldsymbol{\lambda}_k) = \frac{\mathbf{h}_k(\boldsymbol{\lambda}_k)}{\sqrt{\mathbb{E}[\|\mathbf{h}_k(\boldsymbol{\lambda}_k)\|^2]}}, \tag{4.30}$$

Note that MRT is optimal for single-user systems (refer to sec. 2.4). Nevertheless, it is shown in [43] that for large system dimensions such as when when M becomes large MRT reaches the performance of optimal linear precoding due to channel hardening and favorable propagation. Another reason MRT is chosen as the BS precoding technique is because its simpler than its counterparts (for example. ZF precoding). Since this work focuses on IRSs (and not on BS beamforming techniques), we resort to MRT to find

tractable expressions. Substituting the MRT precoder in (4.25) yields

$$\bar{\gamma}_k = \frac{p_k \frac{\mathbb{E}[\|\mathbf{h}_k^H(\boldsymbol{\lambda}_k)\|^4]}{\mathbb{E}[\|\mathbf{h}_k^H(\boldsymbol{\lambda}_k)\|^2]}}{\sum_{t\neq k, t=1}^{K} p_t \frac{\mathbb{E}[\|\mathbf{h}_k^H(\boldsymbol{\lambda}_k)\mathbf{h}_t(\boldsymbol{\lambda}_t)\|^2]}{\mathbb{E}[\|\mathbf{h}_t^H(\boldsymbol{\lambda}_t)\|^2]} + \sigma^2} \tag{4.31}$$

4.2.2 Main Results

In this section, we derive all the terms in (4.31). This would require us to find the first and second moments of a complex Gaussian quadratic form abbreviated as (CGQF). Furthermore, we combine these results to obtain the average SINR expression under MRT with association parameters. We can see that $\mathbf{h}_k \in \mathbb{C}^{M\times 1} \sim \mathcal{CN}(0, \mathbf{R}_k)$ since it contains the addition of complex Gaussian vectors. As a result, $\mathbf{h}_k$ is a complex Gaussian vector with zero mean and correlation matrix $\mathbf{R}_k$, which will be worked out later. Expanding the gain of the collective channel in (4.8) gives

$$\begin{aligned}\mathbb{E}[\|\mathbf{h}_k(\boldsymbol{\lambda}_k)\|^2] = \mathbb{E}[\mathbf{h}_{d,k}^H\mathbf{h}_{d,k} + 2\sum_{l=1}^{L}\lambda_{l,k}{\mathbf{v}_l^{k_l}}^H\mathbf{H}_{0,l,k}^H\mathbf{h}_{d,k} + \\ \sum_{l=1}^{L}\sum_{\bar{l}=1}^{L}\lambda_{l,k}\lambda_{\bar{l},k}{\mathbf{v}_l^{k_l}}^H\mathbf{H}_{0,l,k}^H\mathbf{H}_{0,\bar{l},k}\mathbf{v}_{\bar{l}}^{k_l} + \sum_{l=1}^{L}\sum_{\bar{l}=1}^{L}(1-\lambda_{l,k})(1-\lambda_{\bar{l},k}){\mathbf{v}_l}^H\mathbf{H}_{0,l,k}^H\mathbf{H}_{0,\bar{l},k}\mathbf{v}_{\bar{l}} \\ + 2\sum_{l=1}^{L}(1-\lambda_{l,k}){\mathbf{v}_l}^H\mathbf{H}_{0,l,k}^H\mathbf{h}_{d,k} + 2\sum_{l=1}^{L}\sum_{\bar{l}=1}^{L}(1-\lambda_{l,k})\lambda_{\bar{l},k}{\mathbf{v}_l}^H\mathbf{H}_{0,l,k}^H\mathbf{H}_{0,\bar{l},k}\mathbf{v}_{\bar{l}}^{k_l}].\end{aligned} \tag{4.32}$$

Notice that some of these terms cancel due to independence between channels $\mathbf{h}_{2,l,k}$ and $\mathbf{h}_{d,k}$. Moreover, $\mathbf{v}_l^{k_l}$ is of the form given in (4.22). Now, we shift our attention to computing this expectation which is demonstrated in the following lemma.

Lemma 4.3. *After computing the expectation,* (4.32) *is found to be*

$$\begin{aligned}\mathbb{E}[\|\mathbf{h}_k(\boldsymbol{\lambda}_k)\|^2] = M\beta_{d,k} + \sum_{l=1}^{L}\lambda_{l,k}\Bigg(\sqrt{M}N\sqrt{\beta_{1,l}\beta_{2,l,k}\beta_{d,k}}\frac{\pi}{2} \\ + tr(\mathbf{H}_{1,l}^H\mathbf{H}_{1,l}\boldsymbol{\Sigma}_{\tilde{\mathbf{v}}_l^*}) - \beta_{2,l,k}tr(\mathbf{H}_{1,l}^H\mathbf{H}_{1,l})\Bigg) + \\ \sum_{l=1}^{L}\beta_{2,l,k}tr(\mathbf{H}_{1,l}^H\mathbf{H}_{1,l})\end{aligned} \tag{4.33}$$

Note that $\mathbb{E}[\|\mathbf{h}_k(\lambda_{l,k})\|^2] = tr(\mathbf{R}_k)$.

Proof. Details found in Appendix C. □

It can be seen that average channel gain increases with respect to M. As for L and N, it depends on the user association. For instance, if user k does not have any IRSs associated so that $\lambda_{l,k} = 0, \forall l$, then we still gain something from all the non-associated IRSs. Lemma 4.4 showcases a significant result in CGQF and thus completes the analysis of the numerator for $\bar{\gamma}_k$.

Lemma 4.4. *The fourth moment* $\mathbb{E}\{\|\mathbf{h}_k(\lambda_{l,k})\|^4\}$ *in the numerator of* $\bar{\gamma}_k$ (4.31) *is derived using a result of a CGQF's second moment [40] as*

$$\mathbb{E}[\|\mathbf{h}_k(\boldsymbol{\lambda}_k)\|^4] = tr(\mathbf{R}_k^2) + tr(\mathbf{R}_k)^2. \tag{4.34}$$

Proof.

$$\mathbb{E}[\|\mathbf{h}_k(\boldsymbol{\lambda}_k)\|^4] = Var\{\|\mathbf{h}_k(\boldsymbol{\lambda}_k)\|^2\} + (\mathbb{E}\{\|\mathbf{h}_k(\boldsymbol{\lambda}_k)\|^2\})^2 \tag{4.35}$$

$$= Var\{\|\mathbf{h}_k(\boldsymbol{\lambda}_k)\|^2\} + tr(\mathbf{R}_k)^2 \tag{4.36}$$

$$= tr(\mathbf{R}_k^2) + tr(\mathbf{R}_k)^2, \tag{4.37}$$

where $Var\{\|\mathbf{h}_k(\boldsymbol{\lambda}_k)\|^2\}$ is taken from the following result of the second moment of the CGQF [44]

$$Var\{\|\mathbf{h}_k(\lambda_{l,k})\|^2\} = tr(\mathbf{R}_k^2). \tag{4.38}$$

□

One can see from Lemma 4.4 that this result relies heavily on the correlation matrix $\mathbf{R}_k$, as the size of $\mathbf{R}_k$ increases so does the channel gain. Increasing M increases quadratically the fourth-order moment in (4.36). However, increasing the size of $\mathbf{R}_k$ also affects the interference seen below.

Lemma 4.5. *In the interference term* $\sum_{t\neq k, t=1}^{K} p_t \frac{\mathbb{E}[\|\mathbf{h}_k^H(\boldsymbol{\lambda}_k)\mathbf{h}_t(\boldsymbol{\lambda}_t)\|^2]}{\mathbb{E}[\|\mathbf{h}_t^H(\boldsymbol{\lambda}_t)\|^2]}$, *the numerator is found to be*

$$\mathbb{E}[\|\mathbf{h}_k^H(\boldsymbol{\lambda}_k)\mathbf{h}_t(\boldsymbol{\lambda}_t)\|^2] = tr(\mathbf{R}_t\mathbf{R}_k). \tag{4.39}$$

Proof.

$$\mathbb{E}[\|\mathbf{h}_k^H(\boldsymbol{\lambda}_k)\mathbf{h}_t(\boldsymbol{\lambda}_t)\|^2] = \mathbb{E}[\mathbf{h}_k^H(\boldsymbol{\lambda}_k)\mathbf{h}_t(\boldsymbol{\lambda}_t)\mathbf{h}_t^H(\boldsymbol{\lambda}_t)\mathbf{h}_k(\boldsymbol{\lambda}_k)]. \tag{4.40}$$

Since $\mathbf{h}_t(\boldsymbol{\lambda}_t)$ and $\mathbf{h}_k(\boldsymbol{\lambda}_k)$ are independent, we can use conditional expectation as

$$\mathbb{E}_{\mathbf{h}_k}[\mathbf{h}_k^H(\lambda_{l,k})\mathbb{E}_{\mathbf{h}_t}[\mathbf{h}_t(\lambda_{l,k})\mathbf{h}_t^H(\lambda_{l,k})|\mathbf{h}_k(\lambda_{l,k})]\mathbf{h}_k(\lambda_{l,k})], \tag{4.41}$$
$$= \mathbb{E}_{\mathbf{h}_k}[\mathbf{h}_k^H(\lambda_{l,k})\mathbf{R}_t\mathbf{h}_k(\lambda_{l,k})], \tag{4.42}$$
$$= \mathbb{E}_{\mathbf{h}_k}[tr(\mathbf{h}_k^H(\lambda_{l,k})\mathbf{R}_t\mathbf{h}_k(\lambda_{l,k}))], \tag{4.43}$$
$$= \mathbb{E}_{\mathbf{h}_k}[tr(\mathbf{R}_t\mathbf{h}_k(\lambda_{l,k})\mathbf{h}_k^H(\lambda_{l,k}))], \tag{4.44}$$
$$= tr(\mathbf{R}_t\mathbb{E}_{\mathbf{h}_k}[\mathbf{h}_k(\lambda_{l,k})\mathbf{h}_k^H(\lambda_{l,k})]), \tag{4.45}$$
$$= tr(\mathbf{R}_t\mathbf{R}_k). \tag{4.46}$$

□

It is straightforward to note that when $\mathbf{R}_k$ increases in size, so does the overall interference. Intuitively, increasing the channel gain for one user, increases the interference sensed at another. Previously, Corollaries 4.8 and 4.9 inspected the impact of variations of the correlation matrices and the level of diversity between correlation matrices for different users. The conclusion is that having diverse channels lends to a higher overall SINR, while the presence of similar channels, or similar correlation matrices leads to a degradation in SINR performance.

To obtain expressions in (4.46) and (4.37) we need to find the expression for $\mathbf{R}_k$ which is next computed. Define the correlation matrix for $\mathbf{h}_k$ to be

$$\begin{aligned}\mathbf{R}_k =&\mathbb{E}[\mathbf{h}_{d,k}\mathbf{h}_{d,k}^H + 2\sum_{l=1}^{L}\lambda_{l,k}\mathbf{h}_{d,k}{\mathbf{v}_l^{k_l}}^H\mathbf{H}_{0,l,k}^H \\ &+ \sum_{l=1}^{L}\sum_{\bar{l}=1}^{L}\lambda_{l,k}\lambda_{\bar{l},k}\mathbf{H}_{0,l,k}\mathbf{v}_l^{k_l}{\mathbf{v}_{\bar{l}}^{k_l}}^H\mathbf{H}_{0,\bar{l},k}^H \\ &+ \sum_{l=1}^{L}\sum_{\bar{l}=1}^{L}(1-\lambda_{l,k})(1-\lambda_{\bar{l},k})\mathbf{H}_{0,l,k}\mathbf{v}_l\mathbf{v}_{\bar{l}}^H\mathbf{H}_{0,\bar{l},k}^H],\end{aligned} \tag{4.47}$$

where other terms are zeros as shown in Appendix C. The expression given in the next lemma for $\mathbf{R}_k$ unifies the effect of the system dimensions M, N, L, path loss factors $\beta_{1,l}, \beta_{2,l,k}, \beta_{d,k}$, and association parameters $\lambda_{l,k}$ on $\bar{\gamma}_k$ which depends on $\mathbf{R}_k$. Sec. 4.3 optimizes the association parameters thus tuning $\mathbf{R}_k, k = 1, \ldots, K$ as to maximize the minimum $\bar{\gamma}_k$.

Lemma 4.6. *Using Assumption 1, the correlation matrix for the collective channel in* (4.8) *is shown to be*

$$\mathbf{R}_k = \beta_{d,k}\mathbf{I}_M + \sum_{l=1}^{L} \lambda_{l,k}\Bigg(2\sqrt{\beta_{1,l}\beta_{2,l,k}\beta_{d,k}}\frac{N\pi}{4\sqrt{M}}\mathbf{e}^{j\angle \mathbf{H}_{1,l}\mathbf{H}_{1,l}^H} + \mathbf{H}_{1,l}\mathbf{\Sigma}_{\tilde{\mathbf{v}}_l^{kl}}\mathbf{H}_{1,l}^H - \beta_{2,l,k}\mathbf{H}_{1,l}\mathbf{H}_{1,l}^H\Bigg) + \sum_{l=1}^{L}\beta_{2,l,k}\mathbf{H}_{1,l}\mathbf{H}_{1,l}^H. \tag{4.48}$$

Proof. The proof is postponed to Appendix D. □

Now, we can combine these results in the following Theorem 4.7 to find a closed-form expression for the SINR at user k, note that this closed-form expression manifests due to the Rayleigh fading nature of the direct and IRS l-user k channels, and the MRT precoding used at the BS. Moreover, it would be more accurate to call this an approximation of the average SINR following with Lemma 4.2.

Theorem 4.7. *Using the results from Lemmas 4.3, 4.4, 4.5, the ergodic SINR under MRT with user association parameters is given as*

$$\bar{\gamma}_k = \frac{\frac{p_k}{tr(\mathbf{R}_k)}\left(tr(\mathbf{R}_k^2) + tr(\mathbf{R}_k)^2\right)}{\sum_{t\neq k, t=1}^{K}\frac{p_t}{tr(\mathbf{R}_t)}tr(\mathbf{R}_t\mathbf{R}_k) + \sigma^2}. \tag{4.49}$$

Without loss of generality (WLOG), set $p_k = c_k tr(\mathbf{R}_k)\forall k$, where c_k is a coefficient that will keep this expression from violating the average power constraint in (2.13) to get

$$\bar{\gamma}_k = \frac{c_k(tr(\mathbf{R}_k^2) + tr(\mathbf{R}_k)^2)}{\sum_{t\neq k, t=1}^{K} c_t tr(\mathbf{R}_t\mathbf{R}_k) + \sigma^2}. \tag{4.50}$$

Corollary 4.8. *When the correlation matrices of different users are orthogonal, i.e.* $\mathbf{R}_t\mathbf{R}_k = \mathbf{0}_M$ *. This average SINR in* (4.50) *is simplified to*

$$\bar{\gamma}_{k,Up} = \frac{c_k(tr(\mathbf{R}_k^2) + tr(\mathbf{R}_k)^2)}{\sigma^2} \tag{4.51}$$

This hardly occurs in practice but implies that the spatial correlation matrices are vastly different between users, which also means that the number of M effective antennas for each user reduces to M/K in order to support the induced orthogonality [30]. This special case essentially decomposes the multi-user channel into K orthogonal single-user channels without interference.

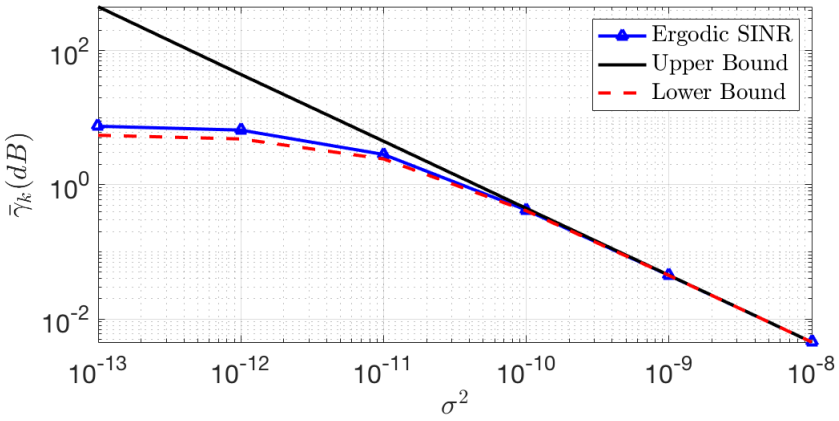

Figure 4.4: Upper and lower bounds of $\bar{\gamma}_k$. The black solid line is linear in σ^2 since the upper bound sets the interference to zero which changes SINR to signal-to-noise ratio (SNR). This figure is simulated using Tab. 4.1

Corollary 4.9. *For the worst case when the correlation matrices of the users are all identical, i.e.,* $\mathbf{R}_t = \mathbf{R}_k$ *[30], we obtain*

$$\bar{\gamma}_{k,Low} = \frac{1 + tr(\mathbf{R}_k)^2/tr(\mathbf{R}_k^2)}{(K-1) + \sigma^2/c_k tr(\mathbf{R}_k^2)}. \tag{4.52}$$

This is also an extreme case since it implies that the users are not sufficiently physically separated. Essentially, Corollaries 4.8 and 4.9 can be thought of as meaningful bounds for $\bar{\gamma}_k$, where we can check that $\bar{\gamma}_{k,Low} \leq \bar{\gamma}_k^s \leq \bar{\gamma}_{k,Up}$.

Proof. To check that $\bar{\gamma}_k^s \leq \bar{\gamma}_{k,Up}$, it suffices to check that $tr(\mathbf{R}_t\mathbf{R}_k)$ is non-negative, one can see that the definition of $tr(\mathbf{R}_t\mathbf{R}_k)$ in (4.42), is the expectation of a non-negative value, since $\mathbf{h}_k^H(\lambda_{l,k})\mathbf{R}_t\mathbf{h}_k(\lambda_{l,k}) \geq 0$, for $\mathbf{R}_t$ is a covariance matrix which is positive semi-definite matrix. Hence, $tr(\mathbf{R}_t\mathbf{R}_k) \geq 0$. We can also check that interference is a power term which can only be non-negative.

As for $\bar{\gamma}_{k,Low} \leq \bar{\gamma}_k^s$, we need to make a fair comparison so we assume that the channel gains $\|\mathbf{h}_k\|^2$ for all users, $k = 1, \ldots, K$ are equal.

Thus, it suffices to show that

$$\sum_{t\neq k,t=1}^{K} tr(\mathbf{R}_t\mathbf{R}_k) \leq (K-1)tr(\mathbf{R}_k^2), \tag{4.53}$$

$$\implies \sum_{t\neq k,t=1}^{K} \mathbf{h}_k^H(\boldsymbol{\lambda}_k)\mathbf{h}_t(\boldsymbol{\lambda}_t) \leq (K-1)\mathbf{h}_k^H(\boldsymbol{\lambda}_k)\mathbf{h}_k(\boldsymbol{\lambda}_k), \tag{4.54}$$

$$\implies \sum_{t\neq k,t=1}^{K} \hat{\mathbf{h}}_k^H(\boldsymbol{\lambda}_k)\hat{\mathbf{h}}_t(\boldsymbol{\lambda}_t) \leq (K-1)\hat{\mathbf{h}}_k^H(\boldsymbol{\lambda}_k)\hat{\mathbf{h}}_k(\boldsymbol{\lambda}_k), \tag{4.55}$$

which is true, since the inner product will rely on the angle between the channels, where $\hat{\mathbf{h}}_k$ and $\hat{\mathbf{h}}_t$ are unit vectors denoting the directions of $\mathbf{h}_k$ and $\mathbf{h}_t$, respectively. □

Lastly, Fig. 4.4 illustrates the tightness of the lower bound and upper bound as noise variance σ^2 is relatively large.

In the next section, we formulate the max-min SINR problem using the average SINR derived in here to find the optimal IRS-user association pairs.

4.3 IRS-User Association Optimization Problem

For this multi-user system, we consider max-min SINR as the metric to improve the performance and fairness of the system. This objective for an IRS-assisted system was previously tackled in [12]. We can formulate the max-min average SINR problem which promotes user fairness as

$$(P1) \quad \max_{\boldsymbol{\Lambda}} \quad \min_{k} \quad \bar{\gamma}_k \tag{4.56}$$

$$s.t. \quad \|\boldsymbol{\lambda}_l\|_1 = 1, 1 \leq l \leq L, \tag{4.57}$$

$$\|\boldsymbol{\lambda}_k\|_1 \leq L, 1 \leq k \leq K, \tag{4.58}$$

$$\lambda_{l,k} \in \{0,1\}, 1 \leq l \leq L, 1 \leq k \leq K, \tag{4.59}$$

Here, $\boldsymbol{\Lambda} \in \mathbb{B}^{K\times L}$ denotes the binary association matrix between all users and IRSs, $\boldsymbol{\lambda}_l \in \mathbb{B}^{K\times 1}$ is a column vector which one element equalling 1 and the rest zeros, the index of 1 in the vector points to the associated user k for the l^{th} IRS, and $\boldsymbol{\lambda}_k \in \mathbb{B}^{1\times L}$ is a row vector of associated and non-associated L IRSs for a particular user k. Constraint (4.57) indicates that each IRS must be assigned to only one user, where $\|\mathbf{x}\|_1 = \sum_i |x_i|$ is the l_1-norm, and x_is are elements of a vector $\mathbf{x}$. On the other hand, constraint (4.58) considers

the user's perspective, as any user has at most L IRSs associated to it. In addition, the binary restriction (4.59) makes this problem combinatorial thus hard to tackle. This is classified as a non-convex mixed-integer non-linear programming (MINLP) problem, which is NP-hard. We can find the optimal solution for the association by exhaustive search, but the complexity is prohibitive and in the order $\mathcal{O}(K^L)$. In the next subsections, we define the search space (codebook) for the exhaustive search method and define a successive refinement algorithm.

4.3.1 Exhaustive Search

In exhaustive search, we run over every possible IRS-user association combination taken from a specific codebook. To create the codebook $C \in \mathbb{B}^{K\times L\times K^L}$ that can generate all possible K^L $\mathbf{\Lambda}$ combinations, we need to transform the problem into a simpler and already solved base conversion problem. First, define the rules according to $(P1)$:

1. Each IRS l is associated to a only one user k.
2. User k has at most L IRSs associated to it.
3. $\lambda_{l,k}$ is a binary variable.

We assign a number $N \in \mathbb{Z}^+$, where N can take values from $1, \ldots, K^L$ to uniquely represent each matrix $\mathbf{\Lambda}_N$. We can then find the base K representation of N, which we store as a row vector $\mathbf{r} \in \mathbb{F}_K$, where $\mathbb{F}_K$ is a finite Galois field with order K. Each element k in $\mathbf{r}, k = 1, \ldots, K$ denotes the position index of 1 in $\boldsymbol{\lambda}_l \in \mathbb{B}^{K\times 1}$ where the rest of its elements are all zeros, because of (4.57). Thus, L of the $\boldsymbol{\lambda}_l$ columns are assembled as $\mathbf{\Lambda}_N$.

4.3.2 Successive Refinement Algorithm

In a bid to observe the gains of user-associations, one low-complexity method referred to as successive refinement (SR) is used solve this optimization problem. This method has a complexity in the order of $\mathcal{O}(L)$ which is better than the well-known branch and bound algorithm for mixed integer linear programming problems which can reach a complexity as high as that of exhaustive search $\mathcal{O}(K^L)$. SR algorithm matches closely to the optimal

solution given by exhaustive search as shown next in the simulations.

Algorithm 1: Successive Refinement Algorithm to Solve $(P1)$

Result: Find desired IRS-user association $\mathbf{\Lambda}^*$
Initialize $\mathbf{\Lambda}$ association matrix based on nearest rule;
Initialize $\mathbf{v}_l^*$ using (4.22); Compute $\bar{\gamma}_k$s in (4.49) based on $\mathbf{\Lambda}$ and $\mathbf{v}_l^*$;
Set the iteration number $i = 1$, Set $state = true$;
while *state is true* **do**
  $\bar{k} = \min_k(\bar{\gamma}_k)$ bottleneck user;
  for $l = 1 : L$ **do**
    if $\lambda_{l,\bar{k}} = 0$ **then**
      Save previous association as j; $\lambda_{l,\bar{k}} = 1$, $\lambda_{l,j} = 0$;
      Update all $\bar{\gamma}_k$s ;
      $\tilde{\gamma}(l) = \min_k(\bar{\gamma}_k)$ bottleneck user SINR;
      reset $\lambda_{l,\bar{k}} = 0$, $\lambda_{l,j} = 1$;
    end
  end
  Find $\bar{l} = \max_l(\tilde{\gamma})$ best assigned IRS for bottleneck user $\bar{k}$;
  $\lambda_{\bar{l},\bar{k}} = 1$, $\lambda_{\bar{l},j} = 0$; $i = i + 1$;
  Update all $\bar{\gamma}_k$s based on current IRS-user association.;
  if $\bar{\gamma}_{\bar{k}}(i) < \bar{\gamma}_{\bar{k}}(i-1)$ **then**
    $state = false$;
    $\bar{\gamma}_{\bar{k}}(i) = \bar{\gamma}_{\bar{k}}(i-1)$;
    $\lambda_{\bar{l},\bar{k}} = 0$, $\lambda_{\bar{l},j} = 1$;
  end
end

4.4 Simulations and Discussion

The simulation parameter values are taken from Table 4.1. The BS is located at $(0, 0)m$ coordinates. The IRSs are deployed in an arc with radius $100m$. The users are set at an arc with radius $85m$ with one user set further away at a radius of $130m$. The path loss factors are computed at 2.5 GHz carrier frequency for the 3GPP Urban Micro (UMi) scenario from TR36.814 (also found in Section V of [1]). The LoS channel model was used for $\mathbf{H}_{1,l}$ and the non-LOS (NLOS) channel model was used to generate path loss factors for $\mathbf{h}_{2,l,k}$ and $\mathbf{h}_{d,k}$, where d in the expression for $\frac{10^{-C/10}}{d^{\alpha}}$ denotes the Euclidean distance between different links. We also assume penetration losses given in Tab. 4.1 for the direct and IRS-user

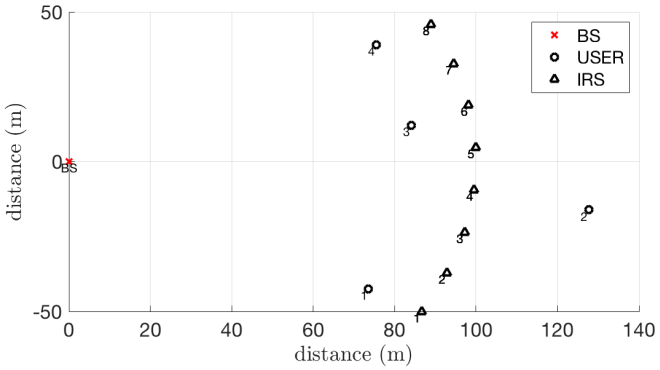

Figure 4.5: No Association Set-up

links due to obstacles in the environment, since the IRSs are envisioned to be elevated the penetration loss is significantly less than that of ground users. The distributed IRSs deployment is illustrated in Fig. 4.5 where the users and IRSs are spread on arc as described earlier and a user (numbered 2) is deliberately positioned further away from the BS and from the IRSs. This deployment scenario highlights the effect of IRSs in helping edge users, as the algorithm would need to assign more IRSs to that minimum user which is by inspection assumed to be user 2.

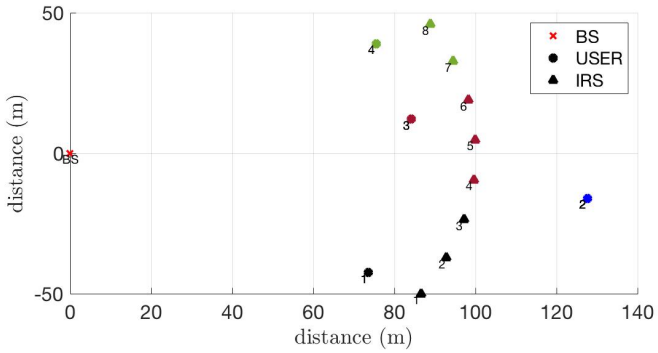

Figure 4.6: Nearest Rule Association

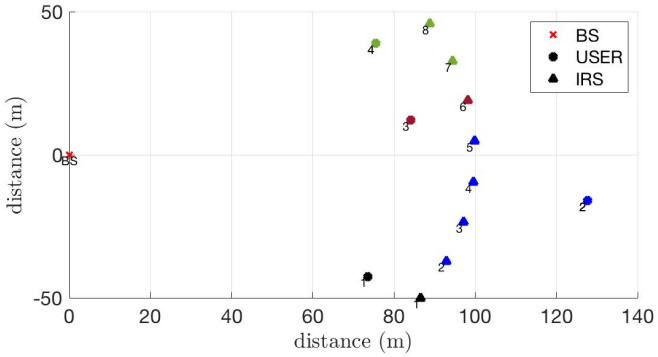

Figure 4.7: Successive Refinement Association

In Fig. 4.6, the associations are updated based on nearest rule (NR), where IRSs with the shortest distance to a particular user would be assigned. 4.7 shows the association resulting from the SR algorithm, which focuses on elevating the SINR for the bottleneck user. We noticed earlier by inspection that user 2 was the bottleneck user. Indeed, SR managed to capture that observation and assign more IRSs to user 2. We will next see that by doing so, the system achieved a higher overall minimum user average SINR. Plotting the minimum user SINR with SINR derived in (4.49) against N in Fig. 4.8, we notice that successive refinement algorithm explained in Algorithm 1 achieves a close performance to that of exhaustive search. With exhaustive search having complexity in the order of $\mathcal{O}(K^L)$ it does not scale well with many users in the system and many more IRSs. Hence the successive refinement (SR) algorithm which scales linearly with L is sufficient to solve the IRS-user association problem. Moreover, the minimum user SINR im-

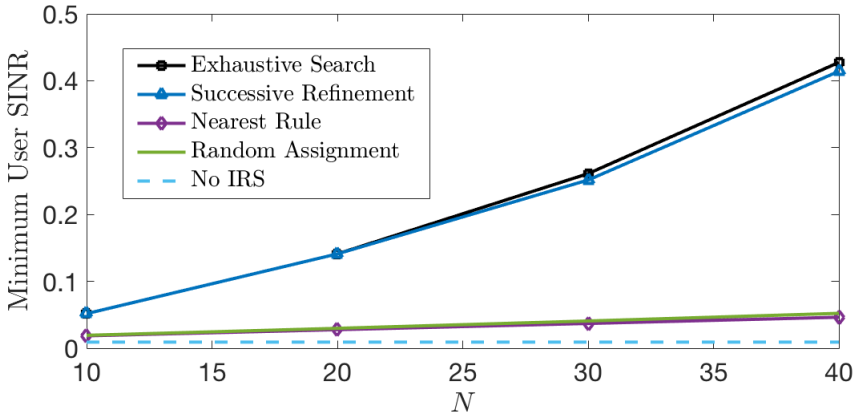

Figure 4.8: Minimum user SINR against number of elements N in each IRS with exhaustive search, successive refinement, nearest rule, random assignment, and no IRSs.

proves with increasing number of elements N but not in the order of N^2. Generally for single user systems where there is no interference, deploying IRSs commonly achieves gains in the order N^2 in the receiver's SNR which is a combination of a gain for the received power collected from the N IRS elements and a beamforming gain factor N collected after passive beamforming [12], [5]). The reason we do not see an N^2 gain is when we increase N we not only benefit the channel gain but also heighten the interference for the minimum user SINR.

Note that nearest rule (NR) and random assignment perform similarly, which is attributed to the fact that they do not necessarily help the minimum user. Thus, even when N increases, the performance remains nearly constant. However, NR and random assignment do perform better than having no IRSs in the system, as we mentioned before, non-associated IRSs still contribute to users which is seen in Lemma 4.3 and the discussion underneath it.

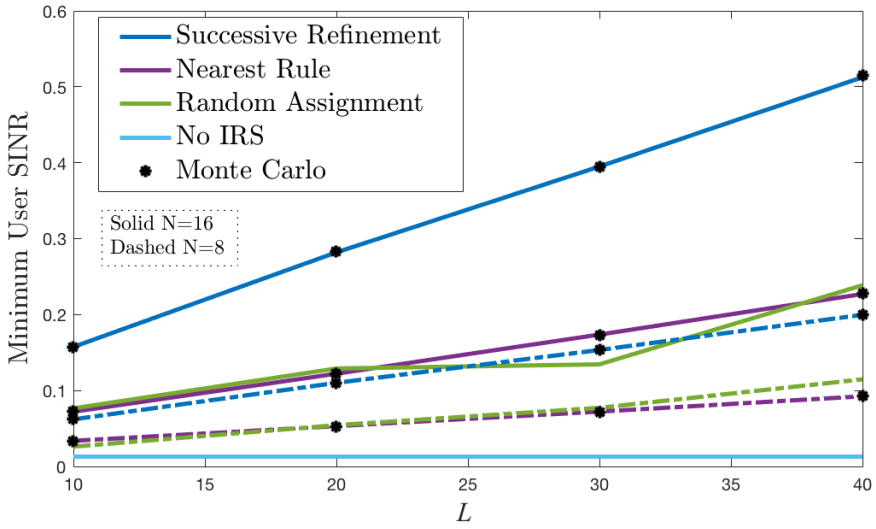

Figure 4.9: Minimum user SINR against number of distributed IRSs L. Both theoretical result in Theorem 4.7 and Monte Carlo results are plotted. Solid lines represent $N = 16$, while dashed lines represent $N = 8$.

The Monte-Carlo simulated values averaged over 1000 channel realizations match the theoretical analysis in sec. 4.2.2 as shown in Fig. 4.9. This validates Theorem 4.7, and proves the ergodic analysis is accurate. Since exhaustive search does not scale well with increasing L, it was discarded for the feasibility of this plot. However, it was already established in the previous figure that SR performs closely to that of exhaustive search which is optimal (also discussed in [29]). Increasing number of IRSs is beneficial, since there are more ways to assign them to users and improve performance. Performance can be further improved if the locations of the IRSs are carefully selected as to broaden coverage. However, that is not the focus of this work.

The solid lines represent a doubled number of IRS elements than that represented by the dashed lines and perform better for the minimum user SINR as expected. Moreover, we see that random assignment is yet again closely matched with NR but they both increase the minimum user SINR although at a slower rate than SR. Deploying more and more IRSs means even edge users are probable to have nearby IRSs assigned to them.

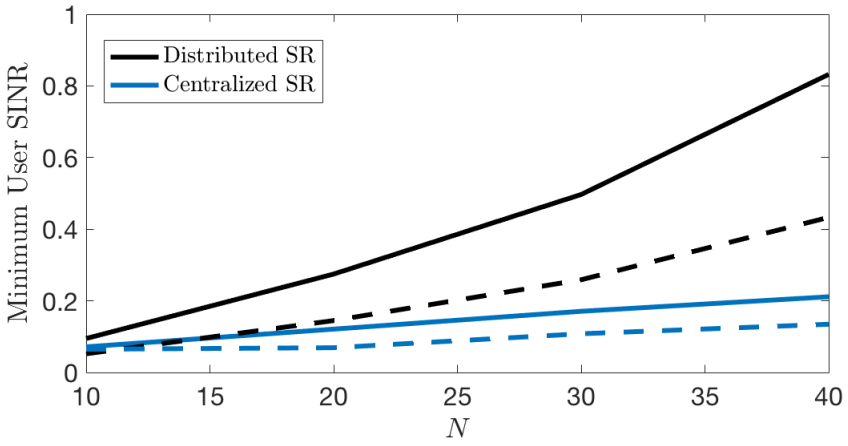

Figure 4.10: Minimum user SINR for centralized and distributed IRSs deployment scenarios against N. Solid lines represent $L = 16$, dashed lines represent $L = 8$.

Fig. 4.10 depicts the marginal improvement the minimum user SINR gains when the IRSs are distributed as opposed to having the IRSs be in one unit or in a centralized deployment. This is because of the spatial diversity the IRSs distributed placement offers. Centralized deployment of IRSs compacts the L IRSs which can be effectively represented by a single large IRS unit with NL elements. In reality, the IRSs and BS are co-located (explained in Chapter 2 sec. 2.3) because of the LoS nature of the BS-IRS l channel, the far-field assumption, and lack of nearby scatterers, which lessens the degrees of freedom especially when the IRSs are deployed close together. Having the IRSs spread out increases the degrees of freedom to at least L which enhances the effect of the IRSs on the system.

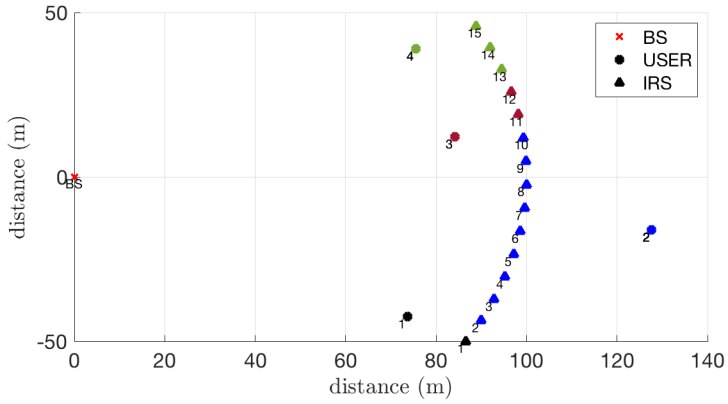

Figure 4.11: Distributed IRSs with $L = 16$ deployment scenario

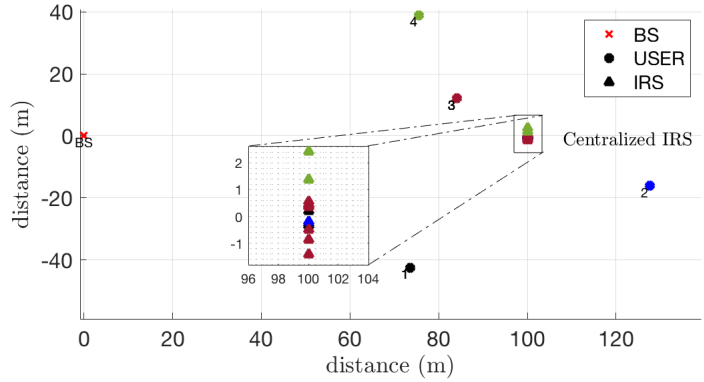

Figure 4.12: Centralized IRSs with $L = 16$ deployment scenario

To visualize this deployment scenario for $L = 16$, we plot in Fig. 4.11 and Fig. 4.12 the SR optimized association for both the distributed and centralized IRSs, respectively. One can see that the centralized IRSs are not majorly assigned to the previously inspected minimum user which was assumed to be 2, since there are high penetration losses between the BS and the other users through the direct link $\mathbf{h}_{d,k}$s. Hence, the lack of nearby IRSs is severe on the system performance, and the minimum user SINR bounces back and forth between the users furthest from the IRSs in each iteration of SR algorithm largely due to this deployment scenario.

Now, we turn our attention to the convergence of the SR algorithm. Looking at Fig. 4.13, we observe how each updated iteration of the SR algorithm matches an improvement in the minimum user SINR as expected.

For $L = 8$, the convergence actually occurs at the ninth iteration. This happens when there is no further IRS-user re-association that could improve the system's minimum SINR.

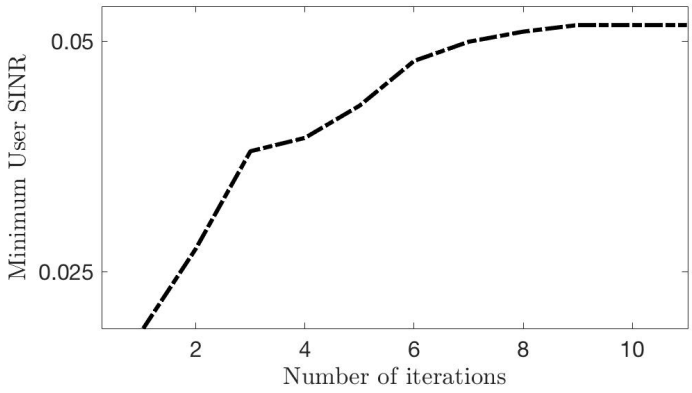

Figure 4.13: Minimum user SINR vs number of iterations in SR algorithm

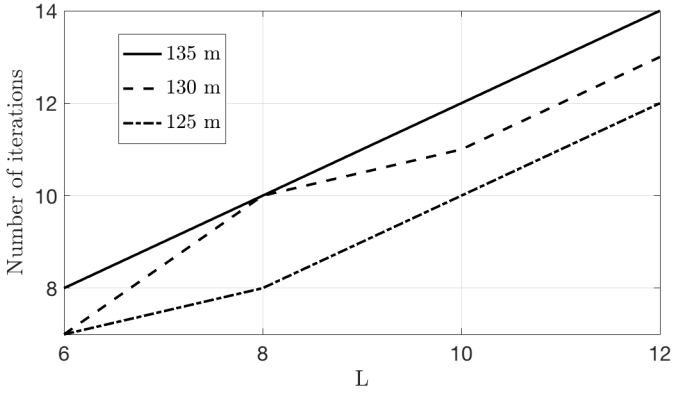

Figure 4.14: Number of iterations needed for SR to converge vs number of IRSs L. With impact of user location on convergence rate of SR algorithm

We next plot the effect of the user 2 location on the convergence rate of the SR algorithm. When user 2 is at $125m$ radius, then some IRSs would be already assigned due to nearest rule initialization, thus leading to less number of iterations needed to converge. As the second user moves away, the iterations increase as the user requires more IRSs to achieve similar per-

formance to other users. Finally, we notice that the trend remains linear in L, which is the motivation behind the suboptimal SR algorithm as opposed to other existing high-complexity methods.

Chapter 5

Conclusion

5.1 Summary

This work proposes a low-complexity CE protocol for the distributed IRSs-assisted MISO system and derives both LS and MMSE estimates of the BS-users direct channels and IRSs-users channels. The NMSE in all developed estimates is analytically studied. By leveraging the LoS nature of the BS-IRSs channel matrices, the proposed protocol achieves a better estimation accuracy per unit training time as compared to the benchmark protocol.

The work also contains the derivation of the average SINR under MRT precoding with IRS-user association parameters. Under the assumption of uncorrelated Rayleigh fading and perfect CSI, a tractable average SINR expression was derived using statistical tools. We formulate a max-min average SINR problem to optimize the IRS-user association parameters. We solve the max-min average SINR problem using a low-complexity algorithm called successive refinement. Simulations results show that successive refinement achieves similar performance as exhaustive search. Moreover, we study a deployment scenario where the effect of increasing the number of IRSs L, increasing the number of elements in each IRS N, the physical distribution of the IRSs, and user location in the system is presented.

5.2 Results

The main results of this work are summarized below:

- Chapter 3 provides a novel channel estimation (CE) protocol for distributed IRSs.
- Chapter 3 provides a comparison between a benchmark CE protocol and the proposed CE protocol for channel estimation that shows that our proposed protocol has higher estimation accuracy per unit time by about 1%.

- Chapter 4 provides an average SINR expression under MRT precoding that is validated using Monte Carlo simulations.
- Chapter 4 compares between deploying the IRSs in a centralized or distributed manner, showing the minimum user SINR gets quadrupled for the distributed case as opposed to centralized for $N = 40$.
- Chapter 4 shows a linear gain for minimum user SINR as L or N increase.
- Chapter 4 shows that the sub-optimal successive refinement algorithm performs closely to the optimal solution given by exhaustive search.

5.3 Limitations of the Study

- The NLoS part of the BS-IRS channel is not taken into consideration in the CE protocol scheme.
- MRT precoding at the BS only maximizes the SNR and often used for a single-user system, the analysis needs to consider ZF precoding at the BS.
- Strict assumptions are made to ensure a tractable analysis of the ergodic SINR. However, some assumptions are impractical such as uncorrelated Rayleigh fading or perfect CSI. We need to extend the system model to a more practical implementation and bridge the gap to industry.

5.4 Future Directions

An important future direction is to develop low-overhead CE protocols for scenarios where the NLoS paths in BS-IRS l channels are non-negligible. Another is to extend the theoretical work for average SINR for imperfect CSI case and thus make the connection to the developed CE protocol. One possible direction to take is to derive the average SINR without first imposing a solution for reflect beamforming vector at each IRS, and then optimize over both association parameters and reflect beamforming.

Another exciting direction is the study of practical implementations of IRSs from the communication perspective, for example defining path loss models based on physical optics theory, as well as modelling practical phase

shifts at the IRS elements based on electromagnetic theory. Studying frequency selective surfaces (FSS) would be intriguing as they may dramatically reduce the complexity of IRS-user associations, if the IRSs can effectively be tuned in frequency to a particular user.

Bibliography

[1] Q.-U.-A. Nadeem *et al.*, "Intelligent reflecting surface assisted multi-user miso communication: Channel estimation and beamforming design," *IEEE Open J. Commun. Soc.*, vol. PP, pp. 1–1, May 2020. → pages v, 6, 16, 21, 25, 27, 29, 31, 51

[2] H. Alwazani, Q.-U.-A. Nadeem, and A. Chaaban, "Channel estimation for distributed intelligent reflecting surfaces assisted multi-user miso systems," 2020. → pages v, 8

[3] J. G. Andrews, X. Zhang, G. D. Durgin, and A. K. Gupta, "Are we approaching the fundamental limits of wireless network densification?" *IEEE Communications Magazine*, vol. 54, no. 10, pp. 184–190, 2016. → pages 1

[4] W. Saad, M. Bennis, and M. Chen, "A vision of 6g wireless systems: Applications, trends, technologies, and open research problems," *IEEE Network*, vol. 34, no. 3, pp. 134–142, 2020. → pages 1

[5] E. Bjrnson, . zdogan, and E. G. Larsson, "Reconfigurable intelligent surfaces: Three and two critical questions," *IEEE Communications Magazine*, vol. 58, no. 12, pp. 90–96, 2020. → pages 1, 3, 5, 54

[6] 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. → pages 2

[7] D. Berry, R. Malech, and W. Kennedy, "The reflectarray antenna," *IEEE Transactions on Antennas and Propagation*, vol. 11, no. 6, pp. 645–651, 1963. → pages 2

[8] . zdogan, E. Bjrnson, and E. G. Larsson, "Intelligent reflecting surfaces: Physics, propagation, and pathloss modeling," *IEEE Wireless Communications Letters*, vol. 9, no. 5, pp. 581–585, 2020. → pages 3

[9] J. F. O'hara, C. Holloway, A. Dienstfrey, E. Kuester, and A. K. Azad, "A discussion on the interpretation and characterization of metafilms/metasurfaces: the two-dimensional equivalent of metamaterials," *Metamaterials*. [Online]. Available: https://www.osti.gov/biblio/960956 → pages 3

[10] Q. Wu and R. Zhang, "Intelligent reflecting surface enhanced wireless network via joint active and passive beamforming," *IEEE Transactions on Wireless Communications*, vol. 18, no. 11, pp. 5394–5409, 2019. → pages 5

[11] E. Bjrnson, . zdogan, and E. G. Larsson, "Intelligent reflecting surface versus decode-and-forward: How large surfaces are needed to beat relaying?" *IEEE Wireless Communications Letters*, vol. 9, no. 2, pp. 244–248, 2020. → pages 5

[12] Q. Nadeem, A. Kammoun, A. Chaaban, M. Debbah, and M. Alouini, "Asymptotic max-min sinr analysis of reconfigurable intelligent surface assisted miso systems," *IEEE Trans. Wireless Commun.*, pp. 1–1, 2020. → pages 5, 6, 11, 16, 49, 54

[13] 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. → pages 5, 6

[14] J. He, K. Yu, and Y. Shi, "Coordinated passive beamforming for distributed intelligent reflecting surfaces network," in *2020 IEEE 91st Vehicular Technology Conference (VTC2020-Spring)*, 2020, pp. 1–5. → pages 5, 6

[15] Q. Wu and R. Zhang, "Intelligent reflecting surface enhanced wireless network: Joint active and passive beamforming design," in *2018 IEEE Global Communications Conference (GLOBECOM)*, Dec 2018, pp. 1–6. → pages 6, 16

[16] 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. → pages 6

[17] D. Mishra and H. Johansson, "Channel estimation and low-complexity beamforming design for passive intelligent surface assisted miso wireless

energy transfer," in *IEEE Int. Conf. Acoustics, Speech, Signal Process. (ICASSP)*, 2019, pp. 4659–4663. → pages 6, 21

[18] T. L. Jensen and E. De Carvalho, "An optimal channel estimation scheme for intelligent reflecting surfaces based on a minimum variance unbiased estimator," in *IEEE ICASSP*, 2020, pp. 5000–5004. → pages 6, 21, 23, 25

[19] B. Zheng and R. Zhang, "Intelligent reflecting surface-enhanced ofdm: Channel estimation and reflection optimization," *IEEE Wireless Commun. Lett.*, vol. 9, no. 4, p. 518522, Apr 2020. → pages 6

[20] Y. Yang, B. Zheng, S. Zhang, and R. Zhang, "Intelligent reflecting surface meets ofdm: Protocol design and rate maximization," *IEEE Trans. Commun.*, vol. PP, pp. 1–1, Mar 2020. → pages 6

[21] Z. Wan, Z. Gao, and M.-S. Alouini, "Broadband Channel Estimation for Intelligent Reflecting Surface Aided mmWave Massive MIMO Systems," *arXiv e-prints*, p. arXiv:2002.01629, Feb. 2020. → pages 6, 16

[22] Z. He and X. Yuan, "Cascaded channel estimation for large intelligent metasurface assisted massive MIMO," *IEEE Wireless Commun. Lett.*, pp. 1–1, 2019. → pages 6

[23] C. Hu and L. Dai, "Two-Timescale Channel Estimation for Reconfigurable Intelligent Surface Aided Wireless Communications," *arXiv e-prints*, p. arXiv:1912.07990, Dec. 2019. → pages 6

[24] G. Zhou *et al.*, "Robust beamforming design for intelligent reflecting surface aided miso communication systems," *IEEE Wireless Communications Letters*, pp. 1–1, 2020. → pages 6, 16

[25] A. Taha, M. Alrabeiah, and A. Alkhateeb, "Enabling large intelligent surfaces with compressive sensing and deep learning," *CoRR*, vol. abs/1904.10136, 2019. [Online]. Available: http://arxiv.org/abs/1904.10136 → pages 6

[26] Q.-U.-A. Nadeem, A. Kammoun, A. Chaaban, M. Debbah, and M.-S. Alouini, "Intelligent reflecting surface assisted wireless communication: Modeling and channel estimation," 2019. → pages 7, 11

[27] J. He, K. Yu, and Y. Shi, "Coordinated Passive Beamforming for Distributed Intelligent Reflecting Surfaces Network," *arXiv e-prints*, p. arXiv:2002.05915, Feb. 2020. → pages 7

[28] M. A. Kishk and M.-S. Alouini, "Exploiting Randomly-located Blockages for Large-Scale Deployment of Intelligent Surfaces," *arXiv e-prints*, p. arXiv:2001.10766, Jan. 2020. → pages 7

[29] W. Mei and R. Zhang, "Performance analysis and user association optimization for wireless network aided by multiple intelligent reflecting surfaces," 2020. → pages 7, 38, 55

[30] E. Björnson, J. Hoydis, and L. Sanguinetti, "Massive MIMO networks: Spectral, energy, and hardware efficiency," *Foundations and Trends® in Signal Processing*, vol. 11, no. 3-4, pp. 154–655, 2017. [Online]. Available: http://dx.doi.org/10.1561/2000000093 → pages 11, 15, 47, 48

[31] M. K. Simon and M. Alouini, "Digital communications over fading channels (m.k. simon and m.s. alouini; 2005) [book review]," *IEEE Transactions on Information Theory*, vol. 54, no. 7, pp. 3369–3370, 2008. → pages 12

[32] A. Grami, "Chapter 12 - wireless communications," in *Introduction to Digital Communications*, A. Grami, Ed. Boston: Academic Press, 2016, pp. 493 – 527. [Online]. Available: http://www.sciencedirect.com/science/article/pii/B9780124076822000120 → pages 14

[33] Da-Shan Shiu, G. J. Foschini, M. J. Gans, and J. M. Kahn, "Fading correlation and its effect on the capacity of multielement antenna systems," *IEEE Transactions on Communications*, vol. 48, no. 3, pp. 502–513, 2000. → pages 15

[34] D. Tse and P. Viswanath, *MIMO I: spatial multiplexing and channel modeling*. Cambridge University Press, 2005, p. 290331. → pages 17, 18

[35] I. Khaled, A. E. Falou, C. Langlais, B. E. Hassan, and M. Jezequel, "Multi-User Digital Beamforming Based on Path Angle Information for mm-Wave MIMO Systems," in *WSA 2020 : 24th International ITG Workshop on Smart Antennas*, Hambourg, Germany, Feb. 2020. [Online]. Available: https://hal.archives-ouvertes.fr/hal-02474863 → pages 18

[36] X. Li, J. Fang, F. Gao, and H. Li, "Joint active and passive beamforming for intelligent reflecting surface-assisted massive mimo systems," *ArXiv*, vol. abs/1912.00728, 2019. → pages 18

[37] T. K. Y. Lo, "Maximum ratio transmission," *IEEE Transactions on Communications*, vol. 47, no. 10, pp. 1458–1461, 1999. → pages 18

[38] E. Bjrnson, E. G. Larsson, and M. Debbah, "Massive mimo for maximal spectral efficiency: How many users and pilots should be allocated?" *IEEE Transactions on Wireless Communications*, vol. 15, no. 2, pp. 1293–1308, 2016. → pages 19

[39] A. Epstein and G. V. Eleftheriades, "Synthesis of passive lossless metasurfaces using auxiliary fields for reflectionless beam splitting and perfect reflection," *Phys. Rev. Lett.*, vol. 117, p. 256103, Dec 2016. [Online]. Available: https://link.aps.org/doi/10.1103/PhysRevLett.117.256103 → pages 36

[40] H. Tataria, P. J. Smith, A. F. Molisch, S. Sangodoyin, M. Matthaiou, P. A. Dmochowski, J. Zhang, and R. S. Thoma, "Spatial correlation variability in multiuser systems," in *2018 IEEE International Conference on Communications (ICC)*, 2018, pp. 1–7. → pages 42, 45

[41] L. Yu, W. Liu, and R. Langley, "Sinr analysis of the subtraction-based smi beamformer," *IEEE Transactions on Signal Processing*, vol. 58, no. 11, pp. 5926–5932, 2010. → pages 42

[42] H. Tataria, P. J. Smith, and P. A. Dmochowski, "On the general analysis of coordinated regularized zero-forcing precoding: An application to two-tier small-cell networks," *IEEE Transactions on Communications*, vol. 65, no. 7, pp. 3133–3150, 2017. → pages 42

[43] J. Hoydis, S. ten Brink, and M. Debbah, "Massive mimo in the ul/dl of cellular networks: How many antennas do we need?" *IEEE Journal on Selected Areas in Communications*, vol. 31, no. 2, pp. 160–171, 2013. → pages 43

[44] P. Ramirez Espinosa *et al.*, "Analysis of gaussian quadratic forms with application to statistical channel modeling," 2020. → pages 45

Appendix

Appendix A

Proof of Theorem 3.1

Proof. Given the observed training signal, $\tilde{\mathbf{r}}_{0,k}^{tr}$ in (3.19), the direct channel's MMSE estimate can be found as

$$\hat{\mathbf{h}}_{d,k} = \mathbf{W}\tilde{\mathbf{r}}_{0,k}^{tr}, \tag{A.1}$$

In which $\mathbf{W}$ is found to be the solution to

$$\begin{aligned} &\min_{\mathbf{W}} \mathbb{E}[\|\hat{\mathbf{h}}_{d,k} - \mathbf{h}_{d,k}\|^2] \\ &\underset{\text{(Def.)}}{\iff} \min_{\mathbf{W}} \mathbb{E}[(\hat{\mathbf{h}}_{d,k} - \mathbf{h}_{d,k})(\hat{\mathbf{h}}_{d,k} - \mathbf{h}_{d,k})^H]) \\ &\underset{\text{(Expand)}}{\iff} \min_{\mathbf{W}} \mathbf{W}\mathbb{E}[\tilde{\mathbf{r}}_{0,k}\tilde{\mathbf{r}}_{0,k}^H]\mathbf{W}^H - 2\mathbb{E}[\mathbf{h}_{d,k}\tilde{\mathbf{r}}_{0,k}^H] + \mathbb{E}[\mathbf{h}_{d,k}\mathbf{h}_{d,k}^H], \end{aligned} \tag{A.2}$$

after differentiating this expression and setting it equal to zero, we acquire the solution for where $\mathbf{W} = \mathbb{E}[\mathbf{h}_{d,k}\tilde{\mathbf{r}}_{0,k}^{tr^H}](\mathbb{E}[\tilde{\mathbf{r}}_{0,k}^{tr}\tilde{\mathbf{r}}_{0,k}^{tr^H}])^{-1}$. Since $\mathbf{n}_k^{tr}$ and $\mathbf{h}_{d,k}$ are independent Gaussian vectors we obtain

$$\mathbb{E}[\mathbf{h}_{d,k}\tilde{\mathbf{r}}_{0,k}^{tr^H}] = \mathbb{E}[\mathbf{h}_{d,k}\mathbf{h}_{d,k}^H] = \beta_{d,k}\mathbf{R}_{BS_k}, \tag{A.3}$$

$$= \mathbb{E}\left[\mathbf{h}_{d,k}\left(\mathbf{h}_{d,k} + \frac{1}{S}(\mathbf{v}_1^{tr} \otimes \mathbf{I}_M)^H\mathbf{n}_k^{tr}\right)^H\right], \tag{A.4}$$

and $\mathbb{E}[\tilde{\mathbf{r}}_{0,k}^{tr}\tilde{\mathbf{r}}_{0,k}^{tr^H}] =$

$$\mathbb{E}[\mathbf{h}_{d,k}\mathbf{h}_{d,k}^H] + \frac{(\mathbf{v}_1^{tr} \otimes \mathbf{I}_M)^H\mathbb{E}\left[\mathbf{n}_k^{tr}\mathbf{n}_k^{tr^H}\right](\mathbf{v}_1^{tr} \otimes \mathbf{I}_M)}{S^2} \tag{A.5}$$

$$= \beta_{d,k}\mathbf{R}_{BS_k} + \frac{\sigma^2}{S^2(P_C\tau_S)}(\mathbf{v}_1^{tr} \otimes \mathbf{I}_M)^H\mathbf{I}_{SM}(\mathbf{v}_1^{tr} \otimes \mathbf{I}_M),$$

$$= \beta_{d,k}\mathbf{R}_{BS_k} + \frac{1}{S}\frac{\sigma^2\mathbf{I}_M}{P_C\tau_S}, \tag{A.6}$$

where the last two steps follow from observing that

$$\begin{aligned}\mathbb{E}\left[\mathbf{n}_k^{tr}\mathbf{n}_k^{tr^H}\right] &= \mathbb{E}\left[\mathbf{n}_{s,k}^{tr}\mathbf{n}_{s,k}^{tr^H}\otimes\mathbf{I}_S\right] \\ &= \frac{1}{(P_C\tau_S)^2}\mathbb{E}\left[\mathbf{N}_s^{tr}\mathbf{x}_{p,k}\mathbf{x}_{p,k}^H\mathbf{N}_s^{tr^H}\right]\otimes\mathbf{I}_S \\ &= \frac{P_C\tau_S}{(P_C\tau_S)^2}\mathbf{I}_S\mathbb{E}[\mathbf{N}_s^{tr}\mathbf{N}_s^{tr^H}]\otimes\mathbf{I}_S \\ &= \frac{\sigma^2}{(P_C\tau_S)}\mathbf{I}_{SM},\end{aligned} \tag{A.7}$$

and (A.6) follows from $\mathbf{v}_1^{tr^H}\mathbf{v}_1^{tr} = S$ under the optimal DFT design for $\mathbf{V}^{tr}$.

Therefore using (A.3) and (A.6) in (A.1) we obtain $(\boldsymbol{\Sigma}_0^{-1})^2$)

$$\hat{\mathbf{h}}_{d,k} = \beta_{d,k}\mathbf{R}_{BS_k}\left(\beta_{d,k}\mathbf{R}_{BS_k} + \frac{\sigma^2\mathbf{I}_M}{SP_C\tau_S}\right)^{-1}\tilde{\mathbf{r}}_{0,k}^{tr}. \tag{A.8}$$

Since $\hat{\mathbf{h}}_{d,k}$ is a complex Gaussian vector, we can compute its covariance matrix $\boldsymbol{\Psi}_{d,k}$ as follows

$$\begin{aligned}&\mathbb{E}[\hat{\mathbf{h}}_{d,k}\hat{\mathbf{h}}_{d,k}^H] = \beta_{d,k}\mathbf{R}_{BS_k}(\mathbb{E}[\tilde{\mathbf{r}}_{0,k}^{tr}\tilde{\mathbf{r}}_{0,k}^{tr^H}])^{-1}\mathbb{E}[\tilde{\mathbf{r}}_{0,k}^{tr}\tilde{\mathbf{r}}_{0,k}^{tr^H}]\quad\left(\mathbb{E}[\tilde{\mathbf{r}}_{0,k}^{tr}\tilde{\mathbf{r}}_{0,k}^{tr^H}]\right)^{-1}\beta_{d,k}\mathbf{R}_{BS_k}^H \\ &= \beta_{d,k}^2\mathbf{R}_{BS_k}\left(\beta_{d,k}\mathbf{R}_{BS_k} + \frac{\sigma^2\mathbf{I}_M}{SP_C\tau_S}\right)^{-1}\mathbf{R}_{BS_k}^H.\end{aligned} \tag{A.9}$$

This concludes the proof. □

Appendix B

Proof of Theorem 3.2

Proof. In a similar fashion, the MMSE estimate for $\mathbf{h}_{2,l,k}$ for the observed training signal in (3.22) is equivalent to

$$\hat{\mathbf{h}}_{2,l,k} = \mathbf{W}\tilde{\mathbf{r}}_{l,k}, \tag{B.1}$$

In which

$$\mathbf{W} = \mathbb{E}[\mathbf{h}_{2,l,k}\tilde{\mathbf{r}}_{l,k}^H](\mathbb{E}[\tilde{\mathbf{r}}_{l,k}\tilde{\mathbf{r}}_{l,k}^H])^{-1} \tag{B.2}$$

Since $\mathbf{n}_k^{tr}$ and $\mathbf{h}_{2,l,k}$ are independent and uncorrelated random Gaussian vectors we can solve as

$$\begin{aligned}\mathbb{E}[\mathbf{h}_{2,l,k}\tilde{\mathbf{r}}_{l,k}^H] &= \mathbb{E}\left[\mathbf{h}_{2,l,k}\left(\mathbf{h}_{2,l,k} + \frac{1}{SM}\boldsymbol{\Sigma}_l^{-1}\bar{\mathbf{H}}_{1,l}^H(\mathbf{V}_l^{tr}\otimes\mathbf{I}_M)^H\mathbf{n}_k^{tr}\right)^H\right], \\ &= \mathbb{E}[\mathbf{h}_{2,l,k}\mathbf{h}_{2,l,k}^H] = \beta_{2,l,k}\mathbf{R}_{IRS_{l,k}}, \end{aligned} \tag{B.3}$$

Moreover $\mathbb{E}[\tilde{\mathbf{r}}_{l,k}\tilde{\mathbf{r}}_{l,k}^H] =$

$$= \mathbb{E}[\mathbf{h}_{2,l,k}\mathbf{h}_{2,l,k}^H] + \frac{\boldsymbol{\Sigma}_l^{-1}\bar{\mathbf{H}}_{1,l}^H(\mathbf{V}_l^{tr}\otimes\mathbf{I}_M)^H\mathbb{E}\left[\mathbf{n}_k^{tr}\mathbf{n}_k^{tr^H}\right](\mathbf{V}_l^{tr}\otimes\mathbf{I}_M)\bar{\mathbf{H}}_{1,l}\boldsymbol{\Sigma}_l^{-1}}{S^2M^2} \tag{B.4}$$

$$= \beta_{2,l,k}\mathbf{R}_{IRS_{l,k}} + \frac{\sigma^2}{S^2M^2(P_C\tau_S)}\boldsymbol{\Sigma}_l^{-1}\bar{\mathbf{H}}_{1,l}^H(\mathbf{V}_l^{tr}\otimes\mathbf{I}_M)^H\mathbf{I}_{SM}(\mathbf{V}_l^{tr}\otimes\mathbf{I}_M)\bar{\mathbf{H}}_{1,l}\boldsymbol{\Sigma}_l^{-1}, \tag{B.5}$$

$$= \beta_{2,l,k}\mathbf{R}_{IRS_{l,k}} + \frac{\sigma^2}{S^2M^2(P_C\tau_S)}\boldsymbol{\Sigma}_l^{-1}\bar{\mathbf{H}}_{1,l}^H(\mathbf{V}_l^{tr^H}\mathbf{V}_l^{tr}\otimes\mathbf{I}_M)\bar{\mathbf{H}}_{1,l}\boldsymbol{\Sigma}_l^{-1}, \tag{B.6}$$

$$= \beta_{2,l,k}\mathbf{R}_{IRS_{l,k}} + \frac{\sigma^2}{SM^2(P_C\tau_S)}\boldsymbol{\Sigma}_l^{-1}\bar{\mathbf{H}}_{1,l}^H(\mathbf{I}_{MN})\bar{\mathbf{H}}_{1,l}\boldsymbol{\Sigma}_l^{-1}, \tag{B.7}$$

$$= \beta_{2,l,k}\mathbf{R}_{IRS_{l,k}} + \frac{\sigma^2}{SM(P_C\tau_S)}\boldsymbol{\Sigma}_l^{-1}\boldsymbol{\Sigma}_l\boldsymbol{\Sigma}_l^{-1}, \tag{B.8}$$

$$= \beta_{2,l,k}\mathbf{R}_{IRS_{l,k}} + \frac{\sigma^2}{SM(P_C\tau_S)}\boldsymbol{\Sigma}_l^{-1} \tag{B.9}$$

where some steps and details are redundant and glossed over in this proof. The expression in (B.7) is exacted by the observation that $\mathbf{V}_l^{tr^H}\mathbf{V}_l^{tr} = S\mathbf{I}_M$ under the DFT protocol for the original undistributed matrix $\mathbf{V}^{tr}$. Moreover, (B.9) is a result of the known BS-IRS l channel matrix multiplication $\bar{\mathbf{H}}_{1,l}^H\bar{\mathbf{H}}_{1,l}$ which results in $M\boldsymbol{\Sigma}_l$. Therefore, the gathered MMSE estimate using (B.9) and (B.3) in (B.1) for $\mathbf{h}_{2,l,k}$ is realized to be

$$\hat{\mathbf{h}}_{2,l,k} = \beta_{2,l,k}\mathbf{R}_{IRS_{l,k}}\left(\beta_{2,l,k}\mathbf{R}_{IRS_{l,k}} + \frac{\sigma^2\boldsymbol{\Sigma}_l^{-1}}{SM(P_C\tau_S)}\right)^{-1}\tilde{\mathbf{r}}_{l,k}, \tag{B.10}$$

The covariance matrix for $\hat{\mathbf{h}}_{2,l,k}$ follows the same steps as $\hat{\mathbf{h}}_{d,k}$'s and can be directly computed.

□

Appendix C

Proof of Lemma 4.3

Given

$$\mathbb{E}[\|\mathbf{h}_k(\boldsymbol{\lambda_k})\|^2] = \tag{C.1}$$

$$\mathbb{E}[\mathbf{h}_{d,k}^H \mathbf{h}_{d,k} + 2\sum_{l=1}^{L} \lambda_{l,k} {\mathbf{v}_l^{k_l}}^H \mathbf{H}_{0,l,k}^H \mathbf{h}_{d,k}, \tag{C.2}$$

$$+ \sum_{l=1}^{L}\sum_{\bar{l}=1}^{L} \lambda_{l,k}\lambda_{\bar{l},k} {\mathbf{v}_l^{k_l}}^H \mathbf{H}_{0,l,k}^H \mathbf{H}_{0,\bar{l},k} \mathbf{v}_{\bar{l}}^{k_l}, \tag{C.3}$$

$$+ \sum_{l=1}^{L}\sum_{\bar{l}=1}^{L} (1-\lambda_{l,k})(1-\lambda_{\bar{l},k}) {\mathbf{v}_l}^H \mathbf{H}_{0,l,k}^H \mathbf{H}_{0,\bar{l},k} \mathbf{v}_{\bar{l}}, \tag{C.4}$$

$$+ 2\sum_{l=1}^{L} (1-\lambda_{l,k}) {\mathbf{v}_l}^H \mathbf{H}_{0,l,k}^H \mathbf{h}_{d,k}, \tag{C.5}$$

$$+ 2\sum_{l=1}^{L}\sum_{\bar{l}=1}^{L} (1-\lambda_{l,k})\lambda_{\bar{l},k} {\mathbf{v}_l}^H \mathbf{H}_{0,l,k}^H \mathbf{H}_{0,\bar{l},k} \mathbf{v}_{\bar{l}}^{k_l}]. \tag{C.6}$$

With $\mathbf{v}_l^{k_l}$ defined in (4.22) and $\mathbf{h}_k(\boldsymbol{\lambda_k})$ defined in (4.8). We can separate this since expectation operation is linear, hence we obtain the sum of six expectations terms $(E_1, \ldots, E_6)$ where we define

$$E_1 = \mathbb{E}[\mathbf{h}_{d,k}^H \mathbf{h}_{d,k}] = M\beta_{d,k}, \tag{C.7}$$

which comes from Assumption 1 and the model for direct channel stated in (2.4). Moreover, E_2 is interpreted as the inner product of the cascaded channel and the direct channel which we maximized through the optimized

configuration for $\mathbf{v}_l^{k_l}$, we can thus compute

$$E_2 = 2\sum_{l=1}^{L}\lambda_{l,k}\mathbb{E}[{\mathbf{v}_l^{k_l}}^H\mathbf{H}_{0,l,k}^H\mathbf{h}_{d,k}], \tag{C.8}$$

$$= 2\sqrt{\beta_{1,l}}\sum_{l=1}^{L}\lambda_{l,k}\mathbb{E}\sum_{n=1}^{N}|b_l(n)||h_{2,l,k}(n)|\cdot\mathbf{a}_l^H\mathbf{h}_{d,k}], \tag{C.9}$$

$$= 2\sqrt{\beta_{1,l}}\sum_{l=1}^{L}\lambda_{l,k}\mathbb{E}[\sum_{n=1}^{N}|h_{2,l,k}(n)|\cdot|\mathbf{a}_l^H\mathbf{h}_{d,k}|], \tag{C.10}$$

$$= 2\sqrt{\beta_{1,l}}\sum_{l=1}^{L}\lambda_{l,k}\mathbb{E}[\sum_{n=1}^{N}|h_{2,l,k}(n)|\cdot|\mathbf{a}_l^H\mathbf{h}_{d,k}|], \tag{C.11}$$

$$= 2\sqrt{\beta_{1,l}}\sum_{l=1}^{L}\lambda_{l,k}\sum_{n=1}^{N}\frac{\pi}{4}\sqrt{\mathbf{a}_l^H\mathbf{a}_l}, \tag{C.12}$$

$$= \sqrt{\beta_{1,l}\beta_{d,k}\beta_{2,l,k}}2\sum_{l=1}^{L}\lambda_{l,k}\frac{\pi}{4}\sqrt{M}N, \tag{C.13}$$

where $\mathbf{H}_{0,l,k}$ is given as in (4.6) and $h_{2,l,k}(n)$ is the n^{th} element in $\mathbf{h}_{2,l,k}$. Given $|X| \sim \text{Rayleigh}(\sigma)$, then the mean

$$\mu_X = \sqrt{\frac{\pi}{4}}\sigma, \tag{C.14}$$

where σ is the standard deviation. The extra-looking $\frac{1}{\sqrt{2}}$ factor appears due to the fact that the Rayleigh distributed random variable is an absolute value of a *complex* normal random variable X. Note that $|h_{2,l,k}(n)|$ and $|\mathbf{a}_l^H\mathbf{h}_{d,k}|$ are statistically independent Rayleigh distributed random variables in which their mean values are given as $\sqrt{\pi\beta_{2,l,k}}/2$ and $\sqrt{\pi\beta_{dk}\mathbf{a}_l^H\mathbf{a}_l}/2$, respectively. We can split the third expectation term as follows

$$E_3 = \mathbb{E}[\sum_{l=1}^{L}\sum_{\bar{l}=1}^{L}\lambda_{l,k}\lambda_{\bar{l},k}{\mathbf{v}_l^{k_l}}^H\mathbf{H}_{0,l,k}^H\mathbf{H}_{0,\bar{l},k}\mathbf{v}_{\bar{l}}^{k_l}], \tag{C.15}$$

$$= \sum_{l=1}^{L}\lambda_{l,k}\mathbb{E}[{\mathbf{v}_l^{k_l}}^H\mathbf{H}_{0,l,k}^H\mathbf{H}_{0,l,k}\mathbf{v}_l^{k_l}]+ \tag{C.16}$$

$$\sum_{l=1}^{L}\sum_{\bar{l}\neq l}^{L}\lambda_{l,k}\lambda_{\bar{l},k}\mathbb{E}[{\mathbf{v}_l^{k_l}}^H\mathbf{H}_{0,l,k}^H\mathbf{H}_{0,\bar{l},k}\mathbf{v}_{\bar{l}}^{k_l}], \tag{C.17}$$

$$= \sum_{l=1}^{L} \lambda_{l,k} \mathbb{E}[\mathbf{v}_l^{k_l H} \mathbf{H}_{0,l,k}^H \mathbf{H}_{0,l,k} \mathbf{v}_l^{k_l}], \tag{C.18}$$

$$= \sum_{l=1}^{L} \lambda_{l,k} \mathbb{E}[\mathbf{e}^{j\angle \mathbf{h}_{d,k}^H \mathbf{H}_{0,l,k}} \mathbf{H}_{0,l,k}^H \mathbf{H}_{0,l,k} \mathbf{e}^{j\angle \mathbf{H}_{0,l,k}^H \mathbf{h}_{d,k}}], \tag{C.19}$$

$$= \sum_{l=1}^{L} \lambda_{l,k} \mathbb{E}[\|\mathbf{H}_{0,l,k} \mathbf{v}_l^{k_l}\|^2]. \tag{C.20}$$

The sum for $l \neq \bar{l}$ cancels to zero, due to independence between the different IRS l-user k channels. Now, we can use a result from CGQF to solve for this as

$$E_3 = \sum_{l=1}^{L} \lambda_{l,k} \mathbb{E}[\mathbf{e}^{j\angle \mathbf{h}_{d,k}^H \mathbf{H}_{0,l,k}} \mathbf{H}_{0,l,k}^H \mathbf{H}_{0,l,k} \mathbf{e}^{j\angle \mathbf{H}_{0,l,k}^H \mathbf{h}_{d,k}}], \tag{C.21}$$

$$= \sum_{l=1}^{L} \lambda_{l,k} \mathbb{E}[\tilde{\mathbf{v}}_l^{k_l H} \mathbf{A} \tilde{\mathbf{v}}_l^{k_l}] \tag{C.22}$$

Where $\tilde{\mathbf{v}}_l^{k_l} = \text{diag}(|\mathbf{h}_{2,l,k}|)\mathbf{e}^{j\angle \mathbf{H}_{1,l}^H \mathbf{h}_{d,k}}$, and

$$\mathbf{A} = \mathbf{H}_{1,l}^H \mathbf{H}_{1,l} = \beta_{1,l} \|\mathbf{a}_l\|^2 \mathbf{b}_l \mathbf{b}_l^H \tag{C.23}$$

is a symmetric, deterministic matrix. We can now write

$$E_3 = \sum_{l=1}^{L} \lambda_{l,k} \mathbb{E}[\tilde{\mathbf{v}}_l^{k_l H} \mathbf{A} \tilde{\mathbf{v}}_l^{k_l}] \quad = \sum_{l=1}^{L} \lambda_{l,k} \big(tr(\mathbf{A} \boldsymbol{\Sigma}_{\tilde{\mathbf{v}}_l^{k_l}}) + \boldsymbol{\mu}_{\tilde{\mathbf{v}}_l^{k_l}}^H \mathbf{A} \boldsymbol{\mu}_{\tilde{\mathbf{v}}_l^{k_l}} \big). \tag{C.24}$$

At this point, we can introduce Lemma C.1 which is relevant in untangling the expectation term $\boldsymbol{\Sigma}_{\tilde{\mathbf{v}}_l^{k_l}} = \mathbb{E}[\tilde{\mathbf{v}}_l^{k_l} \tilde{\mathbf{v}}_l^{k_l H}]$ in (C.24).

Lemma C.1. *The expectation of a random diagonal matrix* $\mathbf{D}$ *multiplied by a deterministic matrix* $\mathbf{A}$ *multiplied by* $\mathbf{D}^T$ *is a quadratic form defined to be*

$$\mathbb{E}[\mathbf{D}\mathbf{A}\mathbf{D}^T] = \mathbb{E}[\mathbf{d}_i \mathbf{A}_{i,j} \mathbf{d}_j^T], \tag{C.25}$$

Proof. We can expand $\mathbb{E}[\mathbf{D}\mathbf{A}\mathbf{D}^T] = \mathbb{E}[[\mathbf{D}\mathbf{A}\mathbf{D}^T]_{i,j}] = \mathbb{E}[\mathbf{d}_i \mathbf{A}_{i,j} \mathbf{d}_j^T]$ and observe the following for the diagonal and off-diagonal terms of the expression such that,

$$\mathbb{E}[\mathbf{d}_i \mathbf{A}_{i,i} \mathbf{d}_i^T] = (\text{Var}(\mathbf{d}_i) + \mathbb{E}[\mathbf{d}_i]^2) \mathbf{A}_{i,i}, \tag{C.26}$$

for the diagonal entries, and for $i \neq j$,

$$\mathbb{E}[\mathbf{d}_i \mathbf{A}_{i,j} \mathbf{d}_j^T] = \mathbb{E}\{\mathbf{d}_i\} \mathbf{A}_{i,j} \mathbb{E}[\mathbf{d}_j]. \tag{C.27}$$

□

Using Lemma C.1, the covariance matrix $\boldsymbol{\Sigma}_{\tilde{\mathbf{v}}_l^{k_l}}$ is found to be

$$\boldsymbol{\Sigma}_{\tilde{\mathbf{v}}_l^{k_l}} = \mathbb{E}[\tilde{\mathbf{v}}_l^{k_l} \tilde{\mathbf{v}}_l^{k_l H}] \tag{C.28}$$

$$= \mathbb{E}[\text{diag}(|\mathbf{h}_{2,l,k}|) \mathbf{e}^{j\angle \mathbf{H}_{1,l}^H \mathbf{h}_{d,k}} \mathbf{e}^{j\angle \mathbf{h}_{d,k}^H \mathbf{H}_{1,l}} \text{diag}(|\mathbf{h}_{2,l,k}|)] \tag{C.29}$$

$$= \mathbb{E}_{\mathbf{h}_{2,l,k}}[\text{diag}(|\mathbf{h}_{2,l,k}|) \mathbb{E}_{\mathbf{h}_{d,k}}[\mathbf{e}^{j\angle \mathbf{b}_l \mathbf{a}_l^H \mathbf{h}_{d,k}} \mathbf{e}^{j\angle \mathbf{h}_{d,k}^H \mathbf{a}_l \mathbf{b}_l^H}] \text{diag}(|\mathbf{h}_{2,l,k}|)] \tag{C.30}$$

$$= \mathbb{E}_{\mathbf{h}_{2,l,k}}[\text{diag}(|\mathbf{h}_{2,l,k}|) \mathbf{e}^{j\angle \mathbf{b}_l} \mathbf{e}^{j\angle \mathbf{b}_l^H}] \text{diag}(|\mathbf{h}_{2,l,k}|)], \tag{C.31}$$

$$= \mathbb{E}_{\mathbf{h}_{2,l,k}}[\text{diag}(|\mathbf{h}_{2,l,k}|) \mathbf{e}^{j\angle \mathbf{H}_{1,l}^H \mathbf{H}_{1,l}} \text{diag}(|\mathbf{h}_{2,l,k}|)] \tag{C.32}$$

$$= (1 - \frac{\pi}{4} + \frac{\pi}{4}) \beta_{2,l,k} \mathbf{e}^{j\angle \mathbf{H}_{1,l}^H \mathbf{H}_{1,l}} \odot \mathbf{I}_N + \frac{\pi \beta_{2,l,k}}{4} \mathbf{e}^{j\angle \mathbf{H}_{1,l}^H \mathbf{H}_{1,l}} \odot (\mathbf{1}_N - \mathbf{I}_N), \tag{C.33}$$

$$= \beta_{2,l,k} \mathbf{I}_N + \frac{\pi \beta_{2,l,k}}{4} \mathbf{e}^{j\angle \mathbf{H}_{1,l}^H \mathbf{H}_{1,l}} \odot (\mathbf{1}_N - \mathbf{I}_N), \tag{C.34}$$

$$= \beta_{2,l,k} \mathbf{I}_N + \frac{\pi \beta_{2,l,k}}{4} \mathbf{e}^{j\angle \mathbf{b}_l \mathbf{b}_l^H} \odot (\mathbf{1}_N - \mathbf{I}_N). \tag{C.35}$$

In the last step, we recall the model of $\mathbf{H}_{1,l}$ in (2.5), some of these steps rely on concepts defined earlier such as the mean of a Rayleigh distribution shown in (C.14). Moreover, the mean $\boldsymbol{\mu}_{\tilde{\mathbf{v}}_l^{k_l}}$ in the quadratic form (C.24) is derived as

$$\boldsymbol{\mu}_{\tilde{\mathbf{v}}_l^*} = \mathbb{E}[\tilde{\mathbf{v}}_l^{k_l}] = \mathbb{E}[\text{diag}(|\mathbf{h}_{2,l,k}|) \mathbf{e}^{j\angle \mathbf{H}_{1,l}^H \mathbf{h}_{d,k}}] \tag{C.36}$$

$$= \mathbb{E}[\text{diag}(|\mathbf{h}_{2,l,k}|)] \mathbb{E}[\mathbf{e}^{j\angle \mathbf{H}_{1,l}^H \mathbf{h}_{d,k}}]] \tag{C.37}$$

$$= \mathbf{0}_{N \times 1}. \tag{C.38}$$

This follows from the independence between the direct and IRS-user channels, and from the fact that the phase distribution of a complex normal is uniform. Since $\mathbf{h}_{d,k}$ follows a circularly symmetric Gaussian distribution, the mean of the phase term will equal the center point, which is zero. Denoting $\boldsymbol{\theta} = \angle \mathbf{H}_{1,l}^H \mathbf{h}_{d,k}$, we can expand the definition of the expectation to achieve

$$\mathbb{E}[e^{j\boldsymbol{\theta}}] = \int_0^{2\pi} \frac{e^{j\boldsymbol{\theta}}}{2\pi} d\boldsymbol{\theta} = \left(\frac{e^{j2\pi}}{j2\pi} - \frac{1}{j2\pi} \right) \mathbf{1}_{N \times 1} = \mathbf{0}_{N \times 1}. \tag{C.39}$$

As for the fourth term, recall that $\mathbf{v}_l$ is the non-associated l^{th} IRS beamforming vector for which $\lambda_{l,k} = 0$. Where it is instead optimized for another user t, such that one can refer to (4.22) to create for an IRS l associated with a particular user t

$$\mathbf{v}_l = e^{j\angle \mathrm{diag}(\mathbf{h}_{2,l,t}\mathbf{b}_l)} e^{j\angle \mathbf{a}_l \mathbf{h}_{d,t})} \tag{C.40}$$

Using Assumption 1 and observing that the covariance matrix $\mathbf{\Sigma}_{\mathbf{v}_l} = \mathbf{I}_N$, we obtain

$$E_4 = \mathbb{E}[\sum_{l=1}^{L}\sum_{\bar{l}=1}^{L}(1-\lambda_{l,k})(1-\lambda_{\bar{l},k})\mathbf{v}_l^H \mathbf{H}_{0,l,k}^H \mathbf{H}_{0,\bar{l},k}\mathbf{v}_{\bar{l}}], \tag{C.41}$$

$$= \sum_{l=1}^{L}(1-\lambda_{l,k})\mathbb{E}[\mathbf{v}_l^H \mathbf{H}_{0,l,k}^H \mathbf{H}_{0,l,k}\mathbf{v}_l] = \beta_{2,l,k} tr(\mathbf{A}), \tag{C.42}$$

with the structure of $\mathbf{A}$ stated in (C.23) and similar to the process of computing E_3 in (C.15).

The fifth term is equal to zero due to the conditional expectation of the direct channel $\mathbf{h}_{d,k}$ for user k and independence to other channels for user t, where $t \neq k$. In realizing this independence, we can write

$$E_5 = \mathbb{E}[2\sum_{l=1}^{L}(1-\lambda_{l,k})\mathbf{v}_l^H \mathbf{H}_{0,l,k}^H \mathbf{h}_{d,k}\}, \tag{C.43}$$

$$= 2\sum_{l=1}^{L}(1-\lambda_{l,k})\mathbb{E}[\mathbf{v}_l^H \mathbf{H}_{0,l,k}^H \mathbb{E}[\mathbf{h}_{d,k}|\mathbf{h}_{d,t}, \mathbf{h}_{2,l,k}, \mathbf{h}_{2,l,t}]] = 0 \tag{C.44}$$

Finally, the sixth term is also equal to zero because conditioning over the other independent random channels yields

$$E_6 = \mathbb{E}[2\sum_{l=1}^{L}\sum_{\bar{l}\neq l}^{L}(1-\lambda_{l,k})\lambda_{\bar{l},k}\mathbf{v}_l^H \mathbf{H}_{0,l,k}^H \mathbf{H}_{0,\bar{l},k}\mathbf{v}_{\bar{l}}^{k_l}], \tag{C.45}$$

$$= 2\sum_{l=1}^{L}\sum_{\bar{l}\neq l}^{L}(1-\lambda_{l,k})\lambda_{\bar{l},k}\mathbb{E}[\mathbf{v}_l^H \mathbf{H}_{0,l,k}^H \mathbf{H}_{0,\bar{l},k}\mathbf{v}_{\bar{l}}^{k_l}], \tag{C.46}$$

$$= 2\sum_{l=1}^{L}\sum_{\bar{l}\neq l}^{L}(1-\lambda_{l,k})\lambda_{\bar{l},k}\mathbb{E}[\mathbf{v}_l^H \mathbf{H}_{0,l,k}^H \mathbf{H}_{1,\bar{l}}|\mathrm{diag}(\mathbf{h}_{2,l,k})|\times \tag{C.47}$$

$$\mathbb{E}_{\mathbf{h}_{d,k}}[\mathbf{e}^{j\angle \mathbf{H}_{1,l}^H \mathbf{h}_{d,k}}|\mathbf{h}_{d,t}, \mathbf{h}_{2,l,k}, \mathbf{h}_{2,l,t}]] = 0.$$

Thus, we complete the proof and find that collecting these terms yields the result in Lemma (4.3), stated here for completeness.

$$\begin{aligned}\mathbb{E}[\|\mathbf{h}_k(\boldsymbol{\lambda}_k)\|^2] = M\beta_{d,k} + \sum_{l=1}^{L} \lambda_{l,k}\Bigg(\sqrt{M}N\sqrt{\beta_{1,l}\beta_{2,l,k}\beta_{d,k}}\frac{\pi}{2} \\ + tr(\mathbf{H}_{1,l}^H\mathbf{H}_{1,l}\boldsymbol{\Sigma}_{\tilde{\mathbf{v}}_l^*}) - \beta_{2,l,k}tr(\mathbf{H}_{1,l}^H\mathbf{H}_{1,l})\Bigg) + \\ \sum_{l=1}^{L}\beta_{2,l,k}tr(\mathbf{H}_{1,l}^H\mathbf{H}_{1,l}).\end{aligned} \tag{C.48}$$

Appendix D

Proof of Lemma 4.6

In Lemma (4.6), we defined $\mathbf{R}_k$ as the covariance matrix of $\mathbf{h}_k(\boldsymbol{\lambda}_k)$, now we are interested in computing the steps needed to achieve Lemma 4.6, note that

$$\begin{aligned}\mathbf{R}_k =&\mathbb{E}[\mathbf{h}_{d,k}\mathbf{h}_{d,k}^H + 2\sum_{l=1}^{L}\lambda_{l,k}\mathbf{h}_{d,k}\mathbf{v}_l^{k_l\,H}\mathbf{H}_{0,l,k}^H \qquad \text{(D.1)}\\ &+\sum_{l=1}^{L}\sum_{\bar{l}=1}^{L}\lambda_{l,k}\lambda_{\bar{l},k}\mathbf{H}_{0,l,k}\mathbf{v}_l^{k_l}\mathbf{v}_{\bar{l}}^{k_l\,H}\mathbf{H}_{0,\bar{l},k}^H\\ &+\sum_{l=1}^{L}\sum_{\bar{l}=1}^{L}(1-\lambda_{l,k})(1-\lambda_{\bar{l},k})\mathbf{H}_{0,l,k}\mathbf{v}_l\mathbf{v}_{\bar{l}}^H\mathbf{H}_{0,\bar{l},k}^H],\\ &=\mathbf{R}_{k,1}+\mathbf{R}_{k,2}+\mathbf{R}_{k,3}+\mathbf{R}_{k,4} \qquad \text{(D.2)}\end{aligned}$$

where other terms are zeros as shown in Appendix C. Since the expecation is a linear operator, we find that

$$\mathbf{R}_{k,1}=\mathbb{E}[\mathbf{h}_{d,k}\mathbf{h}_{d,k}^H]=\beta_{d,k}\mathbf{I}_M \qquad \text{(D.3)}$$

which comes from Assumption 1 and the model for $\mathbf{h}_{d,k}$ defined in (2.4).

Now for $\mathbf{R}_{k,2}$, let us use the definition of $\mathbf{H}_{0,l,k}^H$ in (4.6) to write out

$$\mathbf{R}_{k,2}=\mathbb{E}[2\sum_{l=1}^{L}\lambda_{l,k}\mathbf{h}_{d,k}\mathbf{v}_l^{k_l\,H}\sqrt{\beta_{1,l}}\mathrm{diag}(\mathbf{h}_{2,l,k}^H)\mathbf{b}_l\mathbf{a}_l^H] \qquad \text{(D.4)}$$

without loss of generality, we can drop the unimportant terms and define

$$\tilde{\mathbf{R}}_{k,2}=\mathbb{E}[\mathbf{h}_{d,k}\mathbf{v}_l^{k_l\,H}\mathrm{diag}(\mathbf{h}_{2,l,k}^H)\mathbf{b}_l\mathbf{a}_l^H] \qquad \text{(D.5)}$$

One way to tackle such a problem is to dissect it into its element form as follows

$$[\tilde{\mathbf{R}}_{k,2}]_{m,\bar{m}}=\mathbb{E}[h_{d,k}(m)\mathbf{v}_l^{k_l\,H}\mathrm{diag}(\mathbf{h}_{2,l,k}^H)\mathbf{b}_l a_l(\bar{m})^*] \qquad \text{(D.6)}$$

where $a_l(\bar{m})^*$ is the $\bar{m}^{th}$ element in $\mathbf{a}_l^H$ and $h_{d,k}(m)$ is the m^{th} element in $\mathbf{h}_{d,k}$. We can further simplify by noticing that a row vector times a diagonal matrix times a column vector yields sum of product of their elements and using the optimized reflection vector as in (4.22) to obtain

$$\begin{aligned}[\tilde{\mathbf{R}}_{k,2}]_{m,\bar{m}} =&\mathbb{E}[h_{d,k}(m)\sum_{n=1}^{N} e^{j\angle \mathbf{h}_{d,k}^H \mathbf{a}_l} e^{j\angle b_l(n)^* h_{2,l,k}(n)} \\ &\times h_{2,l,k}^*(n) b_l(n) a_l(\bar{m})^*].\end{aligned} \tag{D.7}$$

Here, $h_{2,l,k}(n)$ is the n^{th} element in $\mathbf{h}_{2,l,k}$, and $b_l(n)$ is the n^{th} element in $\mathbf{b}_l$. Thus, the phases cancel, the expectation of $\mathbb{E}[|h_{2,l,k}(n)|] = \sqrt{\pi\beta_{2,l,k}/4}$ since it is the mean of a Rayleigh distribution defined in (C.14), and $|b_l(n)| = 1, \forall n$ to yield

$$[\tilde{\mathbf{R}}_{k,2}]_{m,\bar{m}} = \sum_{n=1}^{N} \sqrt{\frac{\pi\beta_{2,l,k}}{4}} a_l(\bar{m})^* \mathbb{E}[h_{d,k}(m) e^{j\angle \mathbf{h}_{d,k}^H \mathbf{a}_l}]. \tag{D.8}$$

Define $\bar{\mathbf{R}}_{k,2}$ as follows to focus on the remaining expectation,

$$[\bar{\mathbf{R}}_{k,2}]_{m,\bar{m}} = \mathbb{E}[h_{d,k}(m) e^{j\angle \mathbf{h}_{d,k}^H \mathbf{a}_l}], \tag{D.9}$$

$$= \mathbb{E}[|h_{d,k}(m)| e^{j\angle h_{d,k}(m) \mathbf{h}_{d,k}^H \mathbf{a}_l}]. \tag{D.10}$$

Define $\theta_{h_{d,k}} = \angle h_{d,k}(m) \mathbf{h}_{d,k}^H \mathbf{a}_l$ and expand

$$\theta_{h_{d,k}} = \angle h_{d,k}(m) \sum_{i=1}^{M} h_{d,k}^*(i) a_l(i) \tag{D.11}$$

$$= \angle h_{d,k}(m) h_{d,k}^*(m) a_l(i) + h_{d,k}(m) \sum_{i \neq m}^{M} h_{d,k}^*(i) a_l(i) \tag{D.12}$$

$$= \angle |h_{d,k}(m)|^2 a_l(m) \left(1 + \frac{\sum_{i \neq m}^{M} h_{d,k}^*(i) a_l(i))}{h_{d,k}^*(m) a_l(m)}\right) \tag{D.13}$$

$$= \angle |h_{d,k}(m)|^2 a_l(m)) + \angle \left(1 + \frac{\sum_{i \neq m}^{M} h_{d,k}^*(i) a_l(i))}{h_{d,k}^*(m) a_l(m)}\right) \tag{D.14}$$

$$= \angle a_l(m) + \angle \left(1 + \frac{\sum_{i \neq m}^{M} h_{d,k}^*(i) a_l(i))}{h_{d,k}^*(m) a_l(m)}\right) \tag{D.15}$$

All these steps came from the fact that the angle of product is sum of angles. And if we have an angle of a scalar times a complex number it is equivalent to the angle of the complex number. Let

$$X = \frac{\sum_{i\neq m}^{M} h_{d,k}^{*}(i)a_l(i))}{h_{d,k}^{*}(m)a_l(m)} \tag{D.16}$$

$$Y = 1 + X \tag{D.17}$$

where $X \sim \mathcal{CN}(0, \frac{M-1}{|h_{d,k}(m)|^2})$ and $Y \sim \mathcal{CN}(1, \frac{M-1}{|h_{d,k}(m)|^2})$. We can update

$$[\bar{\mathbf{R}}_{k,2}]_{m,\bar{m}} = e^{j\angle a_l(m)}\mathbb{E}\{|h_{d,k}(m)|e^{j\angle Y}\}. \tag{D.18}$$

At this point we have

$$[\tilde{\mathbf{R}}_{k,2}]_{m,\bar{m}} = N\sqrt{\frac{\pi}{4}}e^{j\angle a_l(\bar{m})^* a_l(m)}\mathbb{E}[|h_{d,k}(m)|e^{j\angle Y}], \tag{D.19}$$

$$= N\sqrt{\frac{\pi}{4}}e^{j\angle a_l(\bar{m})^* a_l(m)}C_m. \tag{D.20}$$

Computing the expectation, the diagonal entries of $\mathbf{R}_{k,2}$ are equal to their average values, now we use (C.12) which is equivalent to $E_2 = tr(\mathbf{R}_{k,2})$ to deduce that $C_m = \sqrt{\frac{\beta_{d,k}\pi}{4M}}$ and obtain

$$[\mathbf{R}_{k,2}]_{m,\bar{m}} = 2\sum_{l=1}^{L}\lambda_{l,k}\sqrt{\beta_{1,l}\beta_{2,l,k}\beta_{d,k}}N\frac{\pi}{4\sqrt{M}}e^{j\angle a_l^*(\bar{m})a_l(m)}. \tag{D.21}$$

As for $\mathbf{R}_{k,3}$, we can check that

$$\mathbf{R}_{k,3} = \mathbb{E}[\sum_{l=1}^{L}\sum_{\bar{l}=1}^{L}\lambda_{l,k}\lambda_{\bar{l},k}\mathbf{H}_{0,l,k}\mathbf{v}_l^{k_l}{\mathbf{v}_{\bar{l}}^{k_l}}^H\mathbf{H}_{0,\bar{l},k}^H] \tag{D.22}$$

$$= \sum_{l=1}^{L}\lambda_{l,k}\mathbb{E}[\mathbf{H}_{0,l,k}\mathbf{v}_l^{k_l}{\mathbf{v}_l^{k_l}}^H\mathbf{H}_{0,l,k}^H] \tag{D.23}$$

$$= \sum_{l=1}^{L}\lambda_{l,k}\beta_{1,l}\mathbf{a}_l\mathbb{E}[\mathbf{b}_l^H\mathrm{diag}(\mathbf{h}_{2,l,k})\mathbf{v}_l^{k_l}{\mathbf{v}_l^{k_l}}^H\mathrm{diag}(\mathbf{h}_{2,l,k}^H)\mathbf{b}_l]\mathbf{a}_l^H \tag{D.24}$$

$$= \sum_{l=1}^{L}\lambda_{l,k}\beta_{1,l}\tilde{\mathbf{R}}_{k,3} \tag{D.25}$$

The summation reduces due to independence between the different IRS-users channels. Using $\mathbf{v}_l^{k_l} = \exp(j\angle\text{diag}(\mathbf{h}_{2,l,k}^H)\mathbf{b}_l)\exp(j\angle\mathbf{a}_l^H\mathbf{h}_{d,k})$, we can find an expression for $\tilde{\mathbf{R}}_{k,3}$ as follows

$$\tilde{\mathbf{R}}_{k,3} = \mathbb{E}[\mathbf{b}_l^H\text{diag}(\mathbf{h}_{2,l,k})\mathbf{v}_l^{k_l}{\mathbf{v}_l^{k_l}}^H\text{diag}(\mathbf{h}_{2,l,k}^H)\mathbf{b}_l], \tag{D.26}$$

$$= \sum_{n=1}^{N}\sum_{\bar{n}=1}^{N}\mathbb{E}[b_l^*(n)h_{2,l,k}(n)v_l^{k_l}(n){v_l^{k_l}}^*(\bar{n})h_{2,l,k}^*(\bar{n})b_l(\bar{n})], \tag{D.27}$$

where $b_l^*(n), h_{2,l,k}(n)$, and $v_l^{k_l}(n)$ are elements in $\mathbf{b}_l^H, \mathbf{h}_{2,l,k}$, and $\mathbf{v}_l^{k_l}$, respectively. Note that we can decompose this double sum as

$$\tilde{\mathbf{R}}_{k,3} = \sum_{n=1}^{N}\mathbb{E}[b_l^*(n)h_{2,l,k}(n)v_l^{k_l}(n){v_l^{k_l}}^*(n)h_{2,l,k}^*(n)b_l(n)] + \tag{D.28}$$

$$\sum_{n=1}^{N}\sum_{\bar{n}\neq n}^{N}\mathbb{E}[b_l^*(n)h_{2,l,k}(n)v_l^{k_l}(n){v_l^{k_l}}^*(\bar{n})h_{2,l,k}^*(\bar{n})b_l(\bar{n})].$$

The first sum yields $N\mathbb{E}[|h_{2,l,k}(n)|^2] = NVar(|h_{2,l,k}(n)|) + N\mathbb{E}[|h_{2,l,k}(n)|]^2$ in which the variance of a Rayleigh distributed random variable $|h_{2,l,k}(n)|$ is given as $Var(|h_{2,l,k}(n)|) = (1-\frac{\pi}{4})\beta_{2,l,k}$, the mean is given by (C.14) which results in $\mathbb{E}[|h_{2,l,k}(n)|]^2 = \frac{\pi}{4}\beta_{2,l,k}$. Putting it all together, $N\mathbb{E}[|h_{2,l,k}(n)|^2] = N\beta_{2,l,k}$.

The argument of the second sum is $b_l^*(n)b_l(n)\mathbb{E}[|h_{2,l,k}(n)|]\mathbb{E}[|h_{2,l,k}(\bar{n})|] = b_l^*(n)b_l(n)\beta_{2,l,k}\frac{\pi}{4}$. Following the result for $\mathbf{\Sigma}_{\tilde{\mathbf{v}}_l^{k_l}}$ in (C.35), and plugging (D.28) in (D.25) yields

$$\mathbf{R}_{k,3} = \sum_{l=1}^{L}\lambda_{l,k}\mathbf{H}_{1,l}\mathbf{\Sigma}_{\tilde{\mathbf{v}}_l^{k_l}}\mathbf{H}_{1,l}^H \tag{D.29}$$

Finally, $\mathbf{R}_{k,4}$ is written as

$$\mathbf{R}_{k,4} = \mathbb{E}[\sum_{l=1}^{L}\sum_{\bar{l}=1}^{L}(1-\lambda_{l,k})(1-\lambda_{\bar{l},k})\mathbf{H}_{0,l,k}\mathbf{v}_l\mathbf{v}_{\bar{l}}^H\mathbf{H}_{0,\bar{l},k}^H] \tag{D.30}$$

$$= \sum_{l=1}^{L}(1-\lambda_{l,k})\mathbb{E}[\mathbf{H}_{0,l,k}\mathbf{v}_l\mathbf{v}_l^H\mathbf{H}_{0,l,k}^H] \tag{D.31}$$

$$= \sum_{l=1}^{L}\beta_{1,l}(1-\lambda_{l,k})\mathbb{E}[\mathbf{a}_l\mathbf{b}_l^H\text{diag}(\mathbf{h}_{2,l,k})\mathbf{v}_l\mathbf{v}_l^H\text{diag}(\mathbf{h}_{2,l,k}^H)\mathbf{b}_l\mathbf{a}_l^H] \tag{D.32}$$

$$= \sum_{l=1}^{L} \beta_{1,l}(1-\lambda_{l,k})\tilde{\mathbf{R}}_{k,4} \tag{D.33}$$

The summation reduces due to independence between the different IRS-users channels. Using $\mathbf{v}_l = \exp(j\angle\text{diag}(\mathbf{h}_{2,l,t}^H)\mathbf{b}_l)\exp(j\angle\mathbf{a}_l^H\mathbf{h}_{d,t})$, for $t \neq k$ and defining $\tilde{\mathbf{R}}_{k,4}$ we get

$$\tilde{\mathbf{R}}_{k,4} = \mathbb{E}[\mathbf{a}_l\mathbf{b}_l^H\text{diag}(\mathbf{h}_{2,l,k})e^{j\angle\text{diag}(\mathbf{h}_{2,l,t}^H)\mathbf{b}_l}e^{j\angle\mathbf{b}_l{}^H\text{diag}(\mathbf{h}_{2,l,t})}\text{diag}(\mathbf{h}_{2,l,k}^H)\mathbf{b}_l\mathbf{a}_l^H] \tag{D.34}$$

$$= \mathbf{a}_l\mathbf{b}_l^H\mathbb{E}[\text{diag}(\mathbf{h}_{2,l,k})e^{j\angle\text{diag}(\mathbf{h}_{2,l,t}^H)\mathbf{b}_l}e^{j\angle\mathbf{b}_l{}^H\text{diag}(\mathbf{h}_{2,l,t})}\text{diag}(\mathbf{h}_{2,l,k}^H)|\mathbf{h}_{2,l,t}], \tag{D.35}$$

$$= \|\mathbf{b}_l\|^2\mathbf{a}_l\mathbf{a}_l^H. \tag{D.36}$$

Notice that $\mathbb{E}[\text{diag}(\mathbf{h}_{2,l,k})e^{j\angle\text{diag}(\mathbf{h}_{2,l,t}^H)\mathbf{b}_l}e^{j\angle\mathbf{b}_l{}^H\text{diag}(\mathbf{h}_{2,l,t})}\text{diag}(\mathbf{h}_{2,l,k}^H)|\mathbf{h}_{2,l,t}]$ can be seen as random diagonal matrix $\text{diag}(\mathbf{h}_{2,l,t}^H)$ multiplied by a deterministic matrix $e^{j\angle\text{diag}(\mathbf{h}_{2,l,t}^H)\mathbf{b}_l}e^{j\angle\mathbf{b}_l{}^H\text{diag}(\mathbf{h}_{2,l,t})}$ multiplied by the same diagonal matrix. Therefore, we have the right set-up to apply Lemma C.1 and get the following: the off-diagonals are equal to zero, while the diagonal terms are equal to $\beta_{2,l,k}$. Thus, we achieve a scaled identity for this expectation term and which simplifies our expression when plugging $\tilde{\mathbf{R}}_{k,4}$ for $\mathbf{R}_{k,4}$ to find

$$\mathbf{R}_{k,4} = \sum_{l=1}^{L}(1-\lambda_{l,k})\beta_{1,l}\beta_{2,l,k}\|\mathbf{b}_l\|^2\mathbf{a}_l\mathbf{a}_l^H \tag{D.37}$$

$$= \sum_{l=1}^{L}(1-\lambda_{l,k})\beta_{2,l,k}\mathbf{H}_{1,l}\mathbf{H}_{1,l}^H \tag{D.38}$$

Thus, we have completed the proof of Lemma 4.6, and the expression for $\mathbf{R}_k$ is validated.

$$\mathbf{R}_k = \beta_{d,k}\mathbf{I}_M + \sum_{l=1}^{L}\lambda_{l,k}\Bigg(2\sqrt{\beta_{1,l}\beta_{2,l,k}\beta_{d,k}}\frac{N\pi}{4\sqrt{M}}\mathbf{e}^{j\angle\mathbf{H}_{1,l}\mathbf{H}_{1,l}^H} \tag{D.39}$$

$$+ \mathbf{H}_{1,l}\mathbf{\Sigma}_{\tilde{\mathbf{v}}_l^{k_l}}\mathbf{H}_{1,l}^H - \beta_{2,l,k}\mathbf{H}_{1,l}\mathbf{H}_{1,l}^H\Bigg) + \sum_{l=1}^{L}\beta_{2,l,k}\mathbf{H}_{1,l}\mathbf{H}_{1,l}^H.$$