Channel Estimation in RIS-Enabled mmWave Wireless Systems: A Variational Inference Approach

by

Firas Fredj

A Thesis submitted to The Faculty of Graduate Studies of

The University of Manitoba

in partial fulfillment of the requirements for the degree of

Master of Science

Department of Electrical and Computer Engineering

University of Manitoba

Winnipeg

Summer 2023

Try not to become a man of success. Rather become a man of value.

ALBERT EINSTEIN

Abstract

Obtaining the channel state information (CSI) is a challenging problem in reconfigurable intelligent surfaces (RIS)-assisted wireless communication systems due to the passive nature of the RIS elements. In this thesis, we study a variational inference (VI)-based CSI estimation approach in a fully passive RIS-aided mmWave single-user single-input multiple-output (SIMO) communication system. Specifically, first, we introduce the VI-based estimation framework in a communication system along with the neural variational networks used to estimate the CSI in a communication system based on the training signals. Then, we propose a VI-based joint channel estimation method to estimate the user-equipment (UE)-to-RIS (UE-RIS) and RIS-to-base station (RIS-BS) channels using uplink training signals in a passive RIS setup. However, updating the phase-shifts based on the instantaneous CSI (I-CSI) leads to a high signaling overhead especially due to the short coherence block of the UE-RIS channel. Therefore, to reduce the signaling complexity for updating the phase-shifts, we propose a VI-based method to estimate the RIS-BS channel along with the covariance matrix of the UE-RIS channel that remains quasi-static for a longer period than the instantaneous UE-RIS channel. In the VI framework, we approximate the posterior of the channel gains/channel covariance matrix with convenient distributions given the received uplink training signals. The parameters of the approximated distributions are generated by deep neural networks trained using variational loss functions derived using the lower bound on the log-likelihood of the received signal. Then, the learned distributions, which are close to the true posterior distributions in terms of Kullback-Leibler divergence, are leveraged to obtain the maximum a posteriori (MAP) estimation of the considered CSI. The simulation results demonstrate that MAP channel estimation using approximated posteriors

yields a capacity that is close to the one achieved with true posteriors, thus demonstrating the effectiveness of the proposed methods. Furthermore, our results show that estimating the channel covariance matrix improves the spectral efficiency by reducing the pilot signaling required to obtain the phase-shifts for the RIS elements in a channel-varying environment.

Keywords: Reconfigurable Intelligent Surfaces (RIS), channel estimation, instantaneous channel state information, statistical channel state information, Variational Inference (VI), mmWave communications, spatial channel covariance estimation

Acknowledgement

First and foremost, I would like to express my sincere appreciation to my supervisor Professor Ekram Hossain for his guidance and mentorship during this research. I am truly grateful for the opportunity to learn and grow and his supervision.

I am profoundly grateful to Professor Amine Mezghani for his exceptional insights and invaluable contributions throughout the course of this work. I am truly fortunate to have had the opportunity to work under his guidance and am sincerely appreciative of the knowledge and growth this collaboration has facilitated. I would like to thank Ms. Amal Feriani for the assistance provided in navigating the Machine Learning intricacies of this work. Her expertise and guidance were crucial in overcoming challenges and achieving the desired outcomes.

With immense appreciation, I express my thanks to the committee members: Professor Pradeepa Yahampath and Professor Amine Mezghani for their valuable feedback to improve the quality of this work.

Finally, I would like to express my gratitude to my family and friends for their support and throughout this journey. Your encouragement and belief in me have been a driving force, and I am truly thankful for the stability you've provided as I worked on this endeavor.

Table of Contents

List of Figures

List of Tables

List of Abbreviations

AOA	Angle of arrival
AOD	Angle of Departure
BS	Base station
CCM	Channel covariance matrix
CSI	Channel state information
DFT	Discrete fourier transform
ELBO	Evidence lower bound
I-CSI	Instantaneous channel state information
KL	Kullback-Leibler
MAP	Maximum a posteriori
mmWave	Millimeter-wave
PDF	Probability density function
RIS	Reconfigurable intelligent surfaces
S-CSI	Statistical channel state information
SIMO	Single-input multiple-output
SNR	Signal-to-noise ratio
UE	User equipment
VI	Variational inference

Chapter 1

Introduction

1.1 Overview

Millimeter-wave (mmWave) communication is one of the emerging technologies for 5G/6G communication systems and beyond to meet the high data rate and spectral efficiency requirements [2]. Although mmWave communications offers a significant gain in throughput thanks to the increased available bandwidth, they are more susceptible to blockages due to rapid signal attenuation and severe path loss. In this context, reconfigurable intelligent surfaces (RISs) have been proposed to mitigate the challenges in mmWave communication systems and also enable smart and reconfigurable wireless environments [1,3]. A RIS is a two-dimensional (2D) array consisting of a large number of passive or semi-passive low-cost reflecting elements that redirect the impinging electromagnetic waves following a specific phase shifts pattern [4, 5] to create a favorable environment for the propagation of the signals. By manipulating the signals' phases and amplitudes, the RIS can create constructive or destructive interference, amplify or attenuate the signals, and improve the communication link quality and coverage [6].

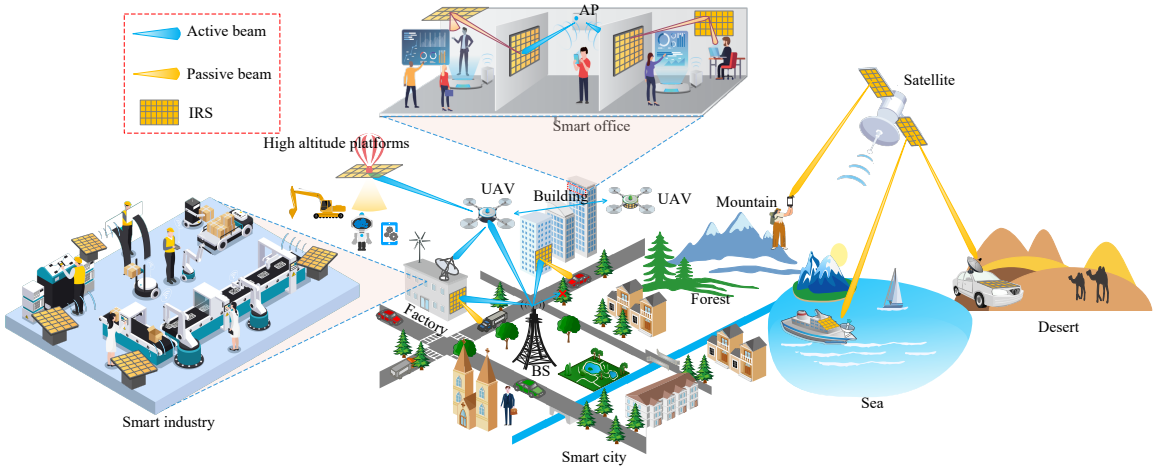


Figure 1.1: Illustration of RIS application/deployment scenarios in future wireless networks. [1]

1.1.1 Use Cases and Potential Applications

RIS technology has many potential benefits, including improving the signal-to-noise ratio (SNR), increasing coverage and capacity, reducing power consumption, and enhancing security and privacy [7–10]. In contrast to conventional wireless communication techniques employed at the transceivers, RIS enables the reshaping of the wireless propagation channels. This results in enhanced signal transmission and the ability to bypass blockages, offering a variety of improvements over the conventional wireless communication techniques such as extending the coverage by establishing a line-of-sight (LOS) link bypassing the signal blockage between the transceivers. Additionally, RIS can be densly integrated into the actual communication systems because of the low-cost of its components and its flexibility to adapt to different environments and frequencies in a scalable manner by relying on passive reflecting elements without Radio Frequency chains as the case is for relayers.

These features make RIS a promising technology for the future wireless systems,

such as in indoor and urban environments where signal quality is typically weaker due to blockages as shown in Fig. 1.1. For instance, RIS can be utilized to extend the coverage area of satellite communication signals by adjusting the phase-shifts to redirect the signal beams towards specific regions that are unreachable by the satellite [11]. Additionally with dynamic signal control, it enables adaptive beamforming and reduces interference, addressing challenges in dense city areas. RIS holds the potential to revolutionize smart cities by providing efficient, sustainable, and reliable wireless connectivity for various applications such as transportation and public services [12]. By dynamically adjusting signals, RIS improves signal strength, directionality, and mitigates indoor signal attenuation. This enables higher data rates, reliable connectivity, and ultra-low latency for mission-critical applications in smart buildings and IoT-centric indoor environments [13].

1.1.2 Challenges in RISs Communication Systems

RIS presents several challenges in their practical implementation that must be addressed for its successful implementation. The main challenges are discussed as follows:

- **CSI acquisition:** To achieve the promised performance of RIS, an accurate CSI is essential [14]. However, the channel estimation process of the RIS related channels is challenging due to the passive nature of the reflecting elements that lack active components of Radio Frequency chains for baseband signal processing, limiting their ability to transmit/receive pilot signals for channel estimation as in conventional systems. Furthermore, the extensive quantity of passive elements results in a substantial number of channel coefficients to be estimated, consequently increasing the system overhead associated with channel estimation [15,16]. In this thesis, we focus on the CSI estimation (instantaneous and statistical) in mmWave RIS-assisted single user communication system.

- **Phase-shifts optimization:** In practical applications, the effectiveness of passive beamforming and reflection optimization in RIS heavily relies on obtaining accurate CSI. However, perfect CSI is often challenging to achieve. In scenarios where imperfect instantaneous CSI (I-CSI) and/or statistical CSI (S-CSI) are available, it becomes essential to design the passive reflection of RIS in conjunction with active transceivers. This joint design ensures robust communication performance, even in the presence of CSI errors, and enhances the overall system reliability and efficiency. In this thesis, we propose closed-form expressions of the phase-shifts maximizing the capacity based on the instantaneous and statistical CSI estimated to evaluate the proposed estimation methods.

1.2 Related Work and Contributions

This section presents an overview of previous works on CSI estimation RIS-assisted wireless communication systems. In first part, we discuss the literature review for estimating the I-CSI. Later, we present the recent works concerning the use of S-CSI to optimize the phase-shifts along with their acquisition.

Instantaneous CSI estimation

In a RIS-aided communication system, the channel between the user equipment (UE) and the base station (BS) through the RIS is usually called a *cascaded* channel [17]. The cascaded channel is composed of two separate channels: *i*) the channel between the UE and the RIS (UE-RIS), and *ii*) the channel between the RIS and the BS (RIS-BS).

Channel estimation in RIS-aided systems has been extensively investigated in the literature along with the reflection optimization when it operates in the reflection mode [18–23]. For instance, a compressed sensing-based method was proposed to solve the channel estimation problem in a single-user narrowband setup that exploits

the mmWave sparse channels [18]. Additionally, a channel estimation scheme was developed for a RIS-aided multi-user broadband communication system by leveraging a shared RIS-BS channel between the users, which improves the training efficiency [19]. In mmWave communication, the channel is modeled by using a small number of paths compared to the number of antennas at the transceivers where each path is distinguished by a direction of departure (DOD) and a direction of arrival (DOA) [24–26]. For the RIS-BS and UE-RIS channels that result from the large number of elements in the RIS, a non-iterative channel estimation framework can be adopted through the estimation of the DODs for the RIS-BS channel and the DOAs for the UE-RIS channel in a first stage, then the cascaded channel BS-UE-RIS can be directly estimated using the estimated DODs and DOAs [20]. However, the training overhead is still considerably high due to the challenge of estimating the UE-RIS-BS link with passive elements in the RIS. To reduce the training overhead, a semi-passive setup of the RIS where the RIS includes a small number of active sensing elements is used to estimate the UE-RIS and the RIS-BS channels in a first coherence block. Using the property of quasi-static channel for the RIS-BS channels (since the RIS and BS are in fixed positions [27]), only the UE-RIS channel is estimated in the training time of the subsequent coherence blocks [21]. In the same context of semi-passive RISs, a variational inference (VI)-based method was developed to reduce the training overhead and estimate the channels using only the uplink training signals, unlike previous works which require the downlink and uplink signals for channel estimation [22]. Furthermore, the decomposition of the cascaded UE-RIS-BS channel into two separate channels (i.e. UE-RIS channel and RIS-BS channel) has been studied in RIS-aided systems with fully passive RIS elements. For instance, it was shown that the received signal follows the parallel factor tensor model [23] which is used to develop channel estimation methods based on the Khatri-Rao factorization of the cascaded channel and iterative alternating estimation scheme.

Statistical CSI for passive beamforming

The aforementioned works focused on estimating the I-CSI of the RIS-related channels. However, the channel conditions in mmWave frequencies can change rapidly since the mmWave signals are more susceptible to blockages and attenuation [28], and therefore, the coherence block of the channels is shorter than that in sub-6GHz bands. Consequently, the optimization of the phase-shifts based on the I-CSI may not be practical in real scenarios as it will result in high training overhead and signaling complexity to update the phase-shifts of the RIS in every coherence block. Recently, various works have been proposed to overcome this challenge by considering the statistical CSI (S-CSI) to solve the active and passive beamforming in RIS-assisted wireless systems [29–31]. In particular, a two-timescale beamforming design was proposed to reduce the training overhead and signal processing for acquiring the I-CSI with a specific transmission protocol [29]. The main idea relies on optimizing the phase-shifts based on the statistical CSI while computing the downlink beamforming vectors based on the I-CSI of the effective channel between the UEs and the BS through the RIS (i.e. UE-RIS-BS channel). A more sophisticated algorithm was proposed in [30] to cover a more general fading channel with discrete phase-shifts in both single-user and multi-user cases. In mmWave scenarios, the statistical CSI was exploited for joint hybrid and passive precoder design using block-coordinate descent-based algorithms to maximize the ergodic capacity [31]. Furthermore, the two-timescale recently attracted the interest of the researchers. In [8], a tradeoff between the spectral efficiency and the energy efficiency has been studied with partial CSI assuming the full knowledge of the RIS-BS channel link and the statistics of the UE-RIS channel links under both continuous and discrete phase-shifts. In [32], a closed-form expression of the ergodic data rate has been developed where the RIS reflecting elements are designed based on the S-CSI.

The S-CSI plays a crucial role in the two-timescale beamforming designs to reduce the training overhead and the signaling complexity. In RIS-aided mmWave wireless

communication systems, the two-timescale based algorithms assume the signal's direction to be constant over a long period of time. In such setting, the directions can be estimated using multiple signal classification (MUSIC) [33, 34] which a well known algorithm for DOAs estimation. However, this strong assumption may be violated in the practical scenario where the angles change over time in a subspace. Typically, the S-CSI is characterized by the spatial channel covariance matrix (CCM). The covariance matrix can be estimated using samples of the estimated channels over a long period of time, or through simulations illustrating the detailed propagation models of the environment [35]. However, these procedures require a large amount of data to obtain an accurate estimate and exhibits a severe training overhead due to the large number of elements at the RIS. Therefore, the structure of the CCM needs to be leveraged in order [35] to reduce the training overhead and improve the S-CSI efficiency. In this context, a CCM estimation method was proposed in [36] by exploiting the low-rank and the semi-definite three-level Toeplitz structure of the covariance matrix of the cascaded channel in an RIS-assisted network.

1.3 Motivation

The vast majority of the existing work focuses on the cascaded channel estimation in RIS-assisted wireless communication [18,19]. In addition, a semi-passive RIS is adopted to estimate the separate channels of the RIS related channels [22] which increases the deployment costs of the RIS due to the integration of Radio Frequency chains. The aforementioned methods proposed estimation methods [18–23] rely basically on the assumption of foreknowledge of the DOAs and DODs, however estimating the DOAs exhibits additional training overhead and complexity to the I-CSI estimation. In Chapter 2, we tackle the problem of jointly estimating the UE-RIS and RIS-BS channels based only on the uplink training signal with a fully passive RIS setup using a VI-based

framework.

Although full knowledge, or estimating the instantaneous channels may lead to the optimal phase shift optimization, it is a challenging task in the real world. First, the coherence time in mmWave channels can be drastically shorter than that in sub-6GHz channels [28]. Hence, the channel estimation and phase optimization need to be performed repeatedly in each coherence block which entails a significant amount of training overhead and tremendous computational resources accompanied by spectral inefficiency due to the pilots sent in each coherence block. Furthermore, the system optimization based on CSI requires frequent transmissions of control signals from the BS to the RIS which involves a considerable amount of signaling overhead. Therefore, to mitigate the overhead due to the traditional channel estimation approaches, one promising direction for the RIS phase-shift design is to use only the statistical CSI of the UE-RIS channel that is considered wide-sense stationary with an invariant covariance $\boldsymbol{R}_{\boldsymbol{h}} = \mathbb{E}[\boldsymbol{h}\boldsymbol{h}^H]$ [37], and the CSI of the RIS-BS channel link which remains quasi-static given the static positions of the RIS and the BS. Therefore, no frequent updates will be required, thus reducing the signaling overhead and enhancing the efficiency of the RIS-aided communication system.

However a limited research has been conducted, so far, on the joint estimation of these UE-RIS covariance and RIS-BS channel. Therefore, in Chapter 3, by leveraging the static nature of the RIS-BS channel and the sparse structure of UE-RIS covariance matrix, we propose a joint channel-covariance estimation method in which the VI-based framework is used by enforcing a complex Laplace priors on the estimated quantities. In addition, we propose a closed-form expression of the phase-shifts given the RIS-BS channel and the UE-RIS covariance that maximize the capacity.

1.4 Scholastic Outputs and Achievements

Table 1.1 presents a summary of my academic achievements. The content of this thesis includes material from the first and second works entitled "Variational Inference-Based Channel Estimation for Reconfigurable Intelligent Surface-Aided Wireless Systems" and "Channel Estimation in RIS-Enabled mmWave Wireless Systems: A Variational Inference Approach"

Table 1.1: Summary of publications and achievements

1. **F. Fredj**, A. Feriani, A. Mezghani, E.Hossain, "Variational Inference-Based Channel Estimation for Reconfigurable Intelligent Surface-Aided Wireless Systems" in *IEEE ICC 2023-IEEE International Conference on Communications*.
2. **F. Fredj**, A. Feriani, A. Mezghani, E.Hossain, "Channel Estimation in RIS-Enabled mmWave Wireless Systems: A Variational Inference Approach" submitted to the *IEEE Transactions on Wireless Communications*.
3. **F. Fredj**, Y. Al-Eryani, S. Maghsudi, M. Akrout and E. Hossain, "Distributed Beamforming Techniques for Cell-Free Wireless Networks Using Deep Reinforcement Learning," in *IEEE Transactions on Cognitive Communications and Networking*, doi: 10.1109/TCCN.2022.3165810.

1.5 Thesis Organization and Notations

Notations: The list of symbols that will be later used within the body of the paper is given in Table 1.2. Scalars, vectors and matrices are denoted by x, $\boldsymbol{x}$, and $\boldsymbol{X}$, respectively. $\boldsymbol{X}^*$ and $\boldsymbol{X}^H$ denote the complex conjugate and conjugate transpose of $\boldsymbol{X}$. The i-th element of a vector a is $\boldsymbol{a}_i$, while the (i,j)-th element of a matrix $\boldsymbol{A}$ is $\boldsymbol{A}_{i,j}$. The $n \times n$ identity matrix is written as $\boldsymbol{I}_n$. The diag($\boldsymbol{a}$) is the diagonal matrix

with the elements of the vector $\boldsymbol{a}$ on the main diagonal. The element-wise product of $\boldsymbol{X}$ and $\boldsymbol{Y}$ is written as $\boldsymbol{X} \circ \boldsymbol{Y}$, while the Khatri-Rao product between $\boldsymbol{X}$ and $\boldsymbol{Y}$ is written as $\boldsymbol{X} \odot \boldsymbol{Y}$. $\boldsymbol{X} \otimes \boldsymbol{Y}$ denotes the kronecker product between $\boldsymbol{X}$ and $\boldsymbol{Y}$. $\text{Tr}(\boldsymbol{X})$ and $|\boldsymbol{X}|$ represent the trace and determinant of the matrix $\boldsymbol{X}$, respectively, and $|x|$ represents the absolute value of a complex number x. The complex Gaussian random vector is denoted as $\boldsymbol{x} \sim \mathcal{CN}(\boldsymbol{m}, \boldsymbol{\Sigma})$ with mean $\boldsymbol{m}$ and covariance matrix $\boldsymbol{\Sigma}$, whereas a complex Laplace random variable x is denoted as $x \sim \mathcal{CL}(m, b)$ with mean m, scale b and probability density function (PDF) given by:

$$p(x) = \frac{1}{2\pi b^2} e^{-\frac{|x-m|}{b}}. \tag{1.1}$$

A Gamma distributed random variable with unit scale is denoted as $x \sim \text{Gamma}(k)$ with shape k, while an Exponentially distributed random variable with rate α is denoted by $x \sim \text{Exp}(\alpha)$.

Table 1.2: List of symbols

System model	
M	Number of antennas at the BS
N	Number of RIS elements
N_p	Number of pilots per UE-RIS coherence block
N_b	Number of coherence blocks used for training
ρ	Signal-to-noise ratio (SNR)
Q	Number of paths of UE-RIS channel
P	Number of paths of RIS-BS channel
θ_n	The phase-shift of the n-th element of the RIS
$\boldsymbol{h}, \boldsymbol{G}$	UE-RIS and RIS-BS channels in the time domain
$\boldsymbol{h}^{\mathsf{vir}}, \boldsymbol{G}^{\mathsf{vir}}$	UE-RIS and RIS-BS channels in the angular domain
$\boldsymbol{R}_{\boldsymbol{h}}, \boldsymbol{d}$	The channel covariance matrix and angular correlation vector of the UE-RIS link
$\boldsymbol{\Phi}$	RIS configurations used for uplink training
$\boldsymbol{F}_N, \boldsymbol{F}_M$	Discrete Fourier Transform (DFT) matrices
Variational Inference	
$\mathcal{L}^{\mathsf{I-CSI}}, \mathcal{L}^{\mathsf{S-CSI}}$	ELBO functions
$p(\boldsymbol{h}, \boldsymbol{G}\|\boldsymbol{Y})$	True posterior
$q_{\boldsymbol{\lambda}_1}(\boldsymbol{h}^{\mathsf{vir}}\|\boldsymbol{Y})$	Auxiliary posterior of UE-RIS channel in the angular domain
$q_{\boldsymbol{\lambda}_2}(\boldsymbol{G}^{\mathsf{vir}}\|\boldsymbol{Y})$	Auxiliary posterior of RIS-BS channel in the angular domain
$q_{\boldsymbol{\lambda}_1}(\boldsymbol{d}\|\boldsymbol{Y})$	Auxiliary posterior of the angular correlation vector
$p(\boldsymbol{h}^{\mathsf{vir}})$	Prior of the UE-RIS channel in the angular domain
$p(\boldsymbol{G}^{\mathsf{vir}})$	Prior of the UE-RIS channel in the angular domain
$p(\boldsymbol{d})$	Prior of the angular correlation
$\mathcal{F}$	Encoder predicts the statistical parameters $\boldsymbol{\lambda}_1$
$\boldsymbol{\mathcal{W}}_1$	Weights of Encoder $\mathcal{F}$
$\mathcal{G}$	Encoder predicts the statistical parameters $\boldsymbol{\lambda}_2$
$\boldsymbol{\mathcal{W}}_2$	Weights of Encoder $\mathcal{G}$

Chapter 2

Joint Channel Estimation for RIS-Assisted System

RIS have emerged as a transformative technology with the potential to revolutionize wireless communication systems. By enabling dynamic control over signal propagation, RIS holds the promise of enhancing coverage, increasing spectral efficiency, and mitigating interference. However, to fully unlock the capabilities of RIS, accurate CSI estimation becomes imperative, as it empowers RIS to intelligently optimize signal paths, harnessing its true potential for next-generation wireless communication systems. In this chapter, we propose a joint channel estimation method to estimate separately the I-CSI of the RIS related channels via variational inference. The main contributions embodied by this chapter are as follows:

- We propose a VI-based estimation framework for estimation in a communication context. We infer over latent variables by approximating the posterior distribution based on the observed signal with parameterized convenient distributions through the minimization of the variational loss which based on minimizing the statistical distribution and the approximated distributions. We use neural networks to model

the approximated distribution where the input is the observed signal and the outputs are the parameters of the approximated distribution

- We adopt the VI-based estimation framework in an RIS-assisted network where we infer over the UE-RIS and RIS-BS channels to approximated the intractable channels posteriors.

- In a first part, we derive our solution based on complex Gaussian prior representing the Rayleigh fading channel models. Then, we leverage the sparse structure in the angular domain of the mmWave channels where we model the sparsity by a complex Laplace distribution. The proposed method bypass the foreknowledge of the number of paths of the channel model to estimate the sparse channels.

- We showcase the effectiveness of our channel estimation framework in terms of capacity and approximation error. Furthermore, we highlight the gain of performance obtained by using the structured channels of mmWave channels compared to the Rayleigh fading channels. Finally, we investigate the effect of sparsity of the mmWave on the performance of the proposed VI-based method.

The rest of the chapter is organized as follows: The VI-based estimation framework is presented in Section 2.1 introducing the variational loss and the neural variational inference. Section 2.2 introduces the system model of a single user RIS-assisted network. In Section 2.3, the derivation the variational loss considering the Rayleigh fading channel models and the path-based channel models for mmWave communication is described. In Section 2.2, we propose a suboptimal closed-form expression of the phase-shifts for passive beamforming to evaluate the the obtained estimates based on the capacity in Section. Finally, exhaustive numerical results are shown in Section 2.5.

2.1 Variational Inference

2.1.1 Introduction

The variational methods are a class of systematic approaches that approximate complex and intractable probability distributions with convenient tractable ones. VI is a specific case of variational methods that infers the marginal distributions or likelihood functions of hidden variables in a statistical model. Specifically, VI is a probabilistic inference technique that offers a flexible and computationally efficient approach for approximating complex posterior distributions [38] [39, 40]. It is particularly useful when direct inference or exact calculations are intractable or computationally expensive. At its core, VI seeks to approximate the true posterior distribution by finding the closest distribution within a family of tractable distributions, known as the variational family. The goal is to find the member of the variational family that minimizes the Kullback-Leibler (KL) divergence to the true posterior. This optimization problem is formulated as maximizing the evidence lower bound (ELBO), which serves as an approximation to the log marginal likelihood of the observed data. By maximizing the ELBO, VI strikes a balance between fitting the observed data and preserving the prior assumptions encoded in the model. The VI algorithm iteratively updates the parameters of the variational distribution to minimize the KL divergence and improve the approximation to the true posterior. This iterative procedure can be seen as a form of coordinate ascent optimization [41]. VI offers several advantages, including scalability to large datasets and complex models, flexibility in model design, and the ability to incorporate prior knowledge and regularization. However, it also introduces a trade-off between accuracy and computational efficiency, as the variational approximation may not capture the full complexity of the true posterior. Nonetheless, VI has found widespread applications in various communication problems, such as joint data detection and phase

noise estimation [42], blind detection under orthogonal frequency-division multiplexing (OFDM) communication systems [43], and the design of an end-to-end communication system leveraging the probabilistic structure of the channels and noise [44], providing a powerful tool for approximating posterior distributions in complex probabilistic models.

For instance, we consider a communication model with two unknown inputs denoted $\boldsymbol{z}_1$ and $\boldsymbol{z}_2$ (e.g. BS-RIS and UE-RIS channels links) and an observed output $\boldsymbol{Y}$, and we assume the output is obtained following a certain probability $p(\boldsymbol{Y}|\boldsymbol{z}_1, \boldsymbol{z}_2)$. If the goal is to infer $\{\boldsymbol{z}_1, \boldsymbol{z}_2\}$ based on the evidence $\boldsymbol{Y}$, we have interest in deriving the probability $p(\boldsymbol{z}_1, \boldsymbol{z}_2|\boldsymbol{Y})$. When the direct evaluation of the posterior distribution $p(\boldsymbol{z}_1, \boldsymbol{z}_2|\boldsymbol{Y})$ is infeasible, VI allows us to approximate the posterior $p(\boldsymbol{z}_1, \boldsymbol{z}_2|\boldsymbol{Y})$ with a parameterized tractable distribution $q_{\boldsymbol{\lambda}}(\boldsymbol{z}_1, \boldsymbol{z}_2|\boldsymbol{Y})$.

2.1.2 Deriving the ELBO

We will delve into the derivation of the evidence lower bound (ELBO) within the context of VI. The ELBO plays a central role as the objective function that guides the optimization process in approximating the posterior distribution. By understanding the derivation of the ELBO, we gain insights into the mathematical foundation of VI and how it balances the trade-off between fidelity to the true posterior and computational tractability. We will step through the derivation, highlighting key mathematical manipulations and concepts along the way, to arrive at a comprehensive understanding of this crucial component of VI.

To do so, we start by expressing the log-likelihood function of $\boldsymbol{Y}$:

$$\begin{aligned}\log p(\boldsymbol{Y}) &= \int_{\boldsymbol{z}_1,\boldsymbol{z}_2} q_{\boldsymbol{\lambda}}(\boldsymbol{z}_1,\boldsymbol{z}_2|\boldsymbol{Y}) \cdot \log p(\boldsymbol{Y})\, d\boldsymbol{z}_1\, d\boldsymbol{z}_2 \\ &= \int_{\boldsymbol{z}_1,\boldsymbol{z}_2} q_{\boldsymbol{\lambda}}(\boldsymbol{z}_1,\boldsymbol{z}_2|\boldsymbol{Y}) \log\left[\frac{p(\boldsymbol{z}_1,\boldsymbol{z}_2,\boldsymbol{Y})}{q_{\boldsymbol{\lambda}}(\boldsymbol{z}_1,\boldsymbol{z}_2|\boldsymbol{Y})} \cdot \frac{q_{\boldsymbol{\lambda}}(\boldsymbol{z}_1,\boldsymbol{z}_2|\boldsymbol{Y})}{p(\boldsymbol{z}_1,\boldsymbol{z}_2|\boldsymbol{Y})}\right] d\boldsymbol{z}_1\, d\boldsymbol{z}_2 \\ &= \mathbb{E}_{\boldsymbol{z}_1,\boldsymbol{z}_2\sim q_{\boldsymbol{\lambda}}(\boldsymbol{z}_1,\boldsymbol{z}_2|\boldsymbol{Y})}\left[\log\frac{p(\boldsymbol{z}_1,\boldsymbol{z}_2,\boldsymbol{Y})}{q_{\boldsymbol{\lambda}}(\boldsymbol{z}_1,\boldsymbol{z}_2|\boldsymbol{Y})}\right] + \underbrace{\mathbb{E}_{\boldsymbol{z}_1,\boldsymbol{z}_2\sim q_{\boldsymbol{\lambda}}(\boldsymbol{z}_1,\boldsymbol{z}_2|\boldsymbol{Y})}\left[\log\frac{q_{\boldsymbol{\lambda}}(\boldsymbol{z}_1,\boldsymbol{z}_2|\boldsymbol{Y})}{p(\boldsymbol{z}_1,\boldsymbol{z}_2|\boldsymbol{Y})}\right]}_{\text{Kullback-Leibler divergence}}.\end{aligned} \tag{2.1}$$

The second term in Eq. 2.1 is the KL divergence $D_{KL}\big(q_{\boldsymbol{\lambda}}(\boldsymbol{z}_1,\boldsymbol{z}_2|\boldsymbol{Y})||p(\boldsymbol{z}_1,\boldsymbol{z}_2|\boldsymbol{Y})\big)$ illustrating the statistical distance in the distribution space that measures how close the approximated distribution $q_{\boldsymbol{\lambda}}(\boldsymbol{z}_1,\boldsymbol{z}_2|\boldsymbol{Y})$ to the exact posterior $p(\boldsymbol{z}_1,\boldsymbol{z}_2|\boldsymbol{Y})$. The minimization of the KL divergence yields a better approximation of the exact posterior, thus we can derive the estimated quantities $\hat{\boldsymbol{z}}_1$ and $\hat{\boldsymbol{z}}_2$ using the approximated distribution $q_{\boldsymbol{\lambda}}(\boldsymbol{z}_1,\boldsymbol{z}_2|\boldsymbol{Y})$. However, the KL-divergence cannot be derived directly since the PDF of the posterior is intractable. Thus, a lower bound, i.e. ELBO, on the log-likelihood of the observations can be derived given that the KL-divergence is a non-negative term:

$$\begin{aligned}\log p(\boldsymbol{Y}) &\geq \mathbb{E}_{\boldsymbol{z}_1,\boldsymbol{z}_2\sim q_{\boldsymbol{\lambda}}(\boldsymbol{z}_1,\boldsymbol{z}_2|\boldsymbol{Y})}\left[\log\frac{p(\boldsymbol{z}_1,\boldsymbol{z}_2,\boldsymbol{Y})}{q_{\boldsymbol{\lambda}}(\boldsymbol{z}_1,\boldsymbol{z}_2|\boldsymbol{Y})}\right] \\ &\triangleq -\mathcal{L}(\boldsymbol{Y};\boldsymbol{\lambda}).\end{aligned} \tag{2.2}$$

Given that $\log p(\boldsymbol{Y})$ in Eq. 2.1 is an unknown constant, maximizing the ELBO $-\mathcal{L}(\boldsymbol{Y};\boldsymbol{\lambda})$ is equivalent to minimizing the KL divergence, which solves the approximation problem of the posterior $p(\boldsymbol{z}_1,\boldsymbol{z}_2|\boldsymbol{Y})$. Assuming that $q_{\boldsymbol{\lambda}}(\boldsymbol{z}_1,\boldsymbol{z}_2|\boldsymbol{Y})$ belongs to a family of tractable distributions defined by a set of parameters $\boldsymbol{\lambda}$, the VI approach consists in optimizing the parameters $\boldsymbol{\lambda}$ of the approximated distribution $q_{\boldsymbol{\lambda}}(\boldsymbol{z}_1,\boldsymbol{z}_2|\boldsymbol{Y})$ that minimizes the objective function $\mathcal{L}(\boldsymbol{Y};\boldsymbol{\lambda})$.

We further assume that the approximated distribution can be factorized as $q_{\boldsymbol{\lambda}}(\boldsymbol{z}_1,\boldsymbol{z}_2|\boldsymbol{Y}) = q_{\boldsymbol{\lambda}_1}(\boldsymbol{z}_1|\boldsymbol{Y}) \cdot q_{\boldsymbol{\lambda}_2}(\boldsymbol{z}_2|\boldsymbol{Y})$ and we optimize the independent distributions by

minimizing $\mathcal{L}(\boldsymbol{Y};\boldsymbol{\lambda}_1,\boldsymbol{\lambda}_2)$ where $\boldsymbol{\lambda}=\{\boldsymbol{\lambda}_1,\boldsymbol{\lambda}_2\}$. This independence assumption is referred to as the *mean-field approximation* [39]. It facilitates the optimization of each individual distribution independently, typically ELBO objective. While this approach offers computational efficiency and tractability, it is essential to note that the assumption of independence among variables may not always hold in complex models. Nonetheless, the mean field approximation remains a widely adopted method in VI, enabling efficient approximation of posterior distributions in various domains of research and application [41]. Specifically, it is equivalent to assuming a low correlation between $\boldsymbol{z}_1$ and $\boldsymbol{z}_2$ conditioned on $\boldsymbol{Y}$. Hence, the objective function is simplified to a general form given by:

$$\begin{aligned}\mathcal{L}(\boldsymbol{\lambda}_1,\boldsymbol{\lambda}_2) &= \mathbb{E}_{\boldsymbol{z}_1,\boldsymbol{z}_2\sim q_{\boldsymbol{\lambda}}(\boldsymbol{z}_1,\boldsymbol{z}_2|\boldsymbol{Y})}\left[\log\frac{q_{\boldsymbol{\lambda}}(\boldsymbol{z}_1,\boldsymbol{z}_2|\boldsymbol{Y})}{p(\boldsymbol{z}_1,\boldsymbol{z}_2,\boldsymbol{Y})}\right]\\ &= \underbrace{\mathbb{E}_{\boldsymbol{z}_1\sim q_{\boldsymbol{\lambda}_1}(\boldsymbol{z}_1|\boldsymbol{Y})}\left[\log\frac{q_{\boldsymbol{\lambda}_1}(\boldsymbol{z}_1|\boldsymbol{Y})}{p(\boldsymbol{z}_1)}\right]}_{\mathcal{L}_1} + \underbrace{\mathbb{E}_{\boldsymbol{z}_2\sim q_{\boldsymbol{\lambda}_2}(\boldsymbol{z}_2|\boldsymbol{Y})}\left[\log\frac{q_{\boldsymbol{\lambda}_2}(\boldsymbol{z}_2|\boldsymbol{Y})}{p(\boldsymbol{z}_2)}\right]}_{\mathcal{L}_2}\\ &\quad \underbrace{-\mathbb{E}_{\boldsymbol{z}_1,\boldsymbol{z}_2\sim q_{\boldsymbol{\lambda}}(\boldsymbol{z}_1,\boldsymbol{z}_2|\boldsymbol{Y})}\left[\log p(\boldsymbol{Y}|\boldsymbol{z}_1,\boldsymbol{z}_2)\right]}_{\mathcal{L}_3}.\end{aligned} \tag{2.3}$$

Note that $\mathcal{L}_1$ and $\mathcal{L}_2$ in Eq. 2.3 represent the KL divergence between the variational distributions $q_{\boldsymbol{\lambda}_1}(\boldsymbol{z}_1|\boldsymbol{Y})$ and $q_{\boldsymbol{\lambda}_2}(\boldsymbol{z}_2|\boldsymbol{Y})$ and their actual priors $p(\boldsymbol{z}_1)$ and $p(\boldsymbol{z}_2)$, respectively. Regarding $\mathcal{L}_3$, it corresponds to the reconstruction error of the estimated pilot signal $\hat{\boldsymbol{Y}}$ with the variational distributions $q_{\boldsymbol{\lambda}_1}(\boldsymbol{z}_1|\boldsymbol{Y})$ and $q_{\boldsymbol{\lambda}_2}(\boldsymbol{z}_2|\boldsymbol{Y})$. Hence, minimizing the objective function $\mathcal{L}(\boldsymbol{\lambda}_1,\boldsymbol{\lambda}_2)=\mathcal{L}_1+\mathcal{L}_2+\mathcal{L}_3$ ensures that the generated posterior distributions are close to the prior distributions and the reconstructed signal $\hat{\boldsymbol{Y}}$ is similar to the received signal.

After deriving the evidence lower bound (ELBO) in VI, one common approach is to use neural networks to parameterize the approximate posterior distribution. Such technique is referred to as *neural variational inference* (neural VI) [45].

2.1.3 Neural Variational Inference

Neural VI, also known as VI with neural networks, combines the power of neural networks with VI to tackle complex probabilistic modeling problems. Neural networks offer remarkable expressive capabilities, enabling them to learn intricate patterns and relationships from data. By incorporating neural networks into VI, we can leverage their flexibility and representational capacity to approximate complex posterior distributions. This is particularly beneficial in scenarios with high-dimensional data or intricate latent structures where traditional VI methods may struggle. Neural networks serve as function approximators, mapping the observed data to the parameters of the variational distribution. This allows for efficient and scalable inference, as the computational burden is amortized across the entire dataset. By harnessing the strengths of neural networks, neural VI provides a powerful and versatile framework for approximating posterior distributions.

Applied to the inference problem which consists on maximizing the ELBO in Eq. 2.3, we introduce two neural networks denoted by *Encoder* $\mathcal{F}$ and *Encoder* $\mathcal{G}$ as shown in Fig. 2.1 that predict the parameters $\boldsymbol{\lambda}_1$ and $\boldsymbol{\lambda}_2$ of the auxiliary distributions $q_{\boldsymbol{\lambda}_1}(\boldsymbol{z}_1|\boldsymbol{Y})$ and $q_{\boldsymbol{\lambda}_2}(\boldsymbol{z}_2|\boldsymbol{Y})$, respectively:

$$\boldsymbol{\lambda}_1 = \mathcal{F}_{\boldsymbol{\mathcal{W}}_1}(\boldsymbol{Y}) \tag{2.4}$$

$$\boldsymbol{\lambda}_2 = \mathcal{G}_{\boldsymbol{\mathcal{W}}_2}(\boldsymbol{Y}) \tag{2.5}$$

where $\boldsymbol{\mathcal{W}}_1$ and $\boldsymbol{\mathcal{W}}_2$ are the weights of the encoders. In particular, the neural networks takes the observed data $\boldsymbol{Y}$ as input and outputs the parameters of distributions $q_{\boldsymbol{\lambda}_1}(\boldsymbol{z}_1|\boldsymbol{Y})$ and $q_{\boldsymbol{\lambda}_2}(\boldsymbol{z}_2|\boldsymbol{Y})$ that approximate the true posterior distributions over the latent variables $\boldsymbol{z}_1$ and $\boldsymbol{z}_2$. The neural networks learns to encode the data into a meaningful representation that captures the latent information. The parameters of the two neural networks *Encoder* $\mathcal{F}$ and *Encoder* $\mathcal{G}$ are learned by maximizing the ELBO ex-

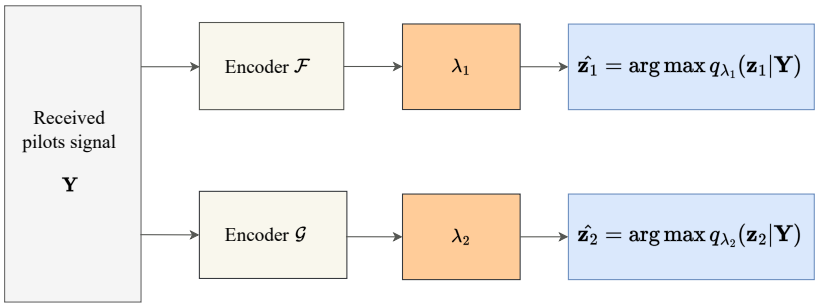


Figure 2.1: Neural Networks

pressed in Eq. 2.3:

$$\boldsymbol{\mathcal{W}}_1^*, \boldsymbol{\mathcal{W}}_2^* = \arg \min_{\boldsymbol{\mathcal{W}}_1, \boldsymbol{\mathcal{W}}_2} \mathcal{L}(\boldsymbol{Y}; \mathcal{F}_{\boldsymbol{\mathcal{W}}_1}(\boldsymbol{Y}); \mathcal{G}_{\boldsymbol{\mathcal{W}}_2}(\boldsymbol{Y})) \tag{2.6}$$

2.2 System Model

We consider a RIS-assisted single-user communication system with M antennas at the BS, N passive reflecting elements at the RIS and a single-antenna user, as illustrated in Fig. 2.2. Considering the uplink transmission, the UE-RIS and RIS-BS channels are denoted by $\boldsymbol{h} \in \mathbb{C}^N$ and $\boldsymbol{G} \in \mathbb{C}^{M \times N}$, respectively. We ignore the direct UE-BS link taking into account that it can be estimated using the conventional single-input multiple-output (SIMO) channel estimation methods by turning off the RIS. Furthermore, we adopt a block-fading channel model where the RIS-related channels $\boldsymbol{G}$ and $\boldsymbol{h}$ are considered quasi-static within a coherence time denoted by $T_{\boldsymbol{G}}$ and $T_{\boldsymbol{h}}$, respectively. Hence, the received signal at the BS can be expressed as follows:

$$\boldsymbol{y} = \sqrt{\rho}\, \boldsymbol{G}\, \mathrm{diag}(\boldsymbol{v})\, \boldsymbol{h}\, x + \boldsymbol{w}, \tag{2.7}$$

where ρ, $x \in \mathbb{C}$ and $\boldsymbol{w} \in \mathbb{C}^M$ are, respectively, the SNR, the transmitted signal, and the additive white noise, i.e. $\boldsymbol{w} \sim \mathcal{CN}(\boldsymbol{0}, \boldsymbol{I}_M)$. The phase shifts contributed by the RIS are

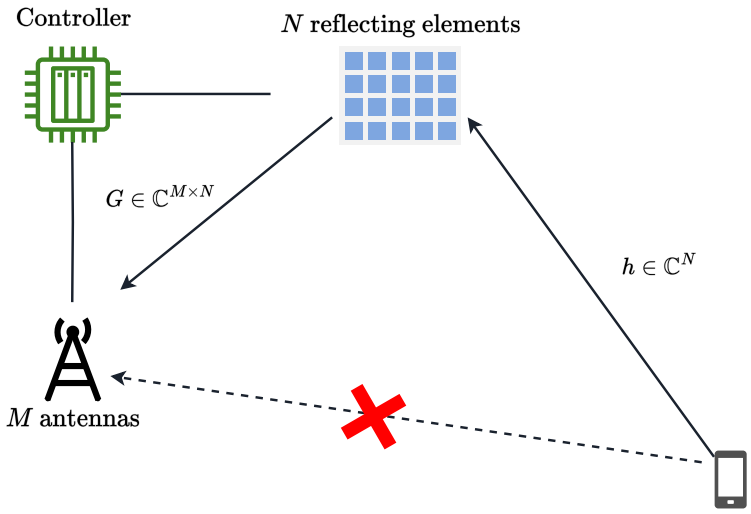


Figure 2.2: RIS-aided wireless communication system.

represented by the diagonal matrix diag($\boldsymbol{v}$), where $\boldsymbol{v} = [e^{j\theta_1}, \ldots, e^{j\theta_N}]$ with $\theta_n \in [0, 2\pi)$ is the phase shift of the n-th element in the RIS.

We aim at optimizing the phase shifts of the impinging signals at RIS based on the estimated CSI. We employ the VI technique to jointly estimate the channel links between the UE and the RIS $\boldsymbol{h}$, as well as between the RIS and the BS $\boldsymbol{G}$, based on pilot transmissions in the uplink.

The training signal is obtained by sending N_p pilot signals by the user to the BS through the UE-RIS-BS channel, denoted by $\mathbb{C}^{N_p}$. For different transmissions, different configurations of the RIS are maintained for each pilot signal, denoted by $\boldsymbol{v}_l$. The received training signal is given by:

$$\boldsymbol{Y} = \sqrt{\rho}\, \boldsymbol{G} \left(\boldsymbol{\Phi} \circ (\boldsymbol{h}\boldsymbol{x}^{\boldsymbol{T}}) \right) + \boldsymbol{W}, \tag{2.8}$$

where $\boldsymbol{Y} = \left[\boldsymbol{y}_1, \ldots, \boldsymbol{y}_l\right] \in \mathbb{C}^{M \times N_p}$ is the concatenation of the N_p training signals, $\boldsymbol{x} = \left[x_1, \ldots, x_{N_p}\right]^T$ denotes the pilots sent by the user, $\boldsymbol{\Phi} = \left[\boldsymbol{v}_1, \ldots, \boldsymbol{v}_{N_p}\right]$ is the concatenation of the phase-shifts vectors used where $\boldsymbol{v}_l$ is assigned to the l-th pilot signal, and $\boldsymbol{W} = \left[\boldsymbol{w}_1, \ldots, \boldsymbol{w}_{N_p}\right]$ is the noise matrix.

2.3 Variational Inference-Based I-CSI Estimation

In this section, we employ the VI-based estimation framework to estimate the RIS related channels separately. We first derive the ELBO for Rayleigh fading channels with complex Gaussian prior, then we derive the ELBO for mmWave channel enforcing a complex Laplace prior on the channels to model the sparsity characteristic of the channels.

2.3.1 Derivation of ELBO for Rayleigh Fading Channels

By applying the VI framework, we approximate the intractable true posterior distribution $p(\boldsymbol{h}, \boldsymbol{G}|\boldsymbol{Y})$ by a tractable parameterized distribution denoted $q_{\boldsymbol{\lambda}}(\boldsymbol{h}, \boldsymbol{G}|\boldsymbol{Y})$ that maximizes the ELBO function. Assuming a low-correlation between the channels $\boldsymbol{h}$ and $\boldsymbol{G}$ conditioned on the training signal $\boldsymbol{Y}$, by using the mean-field approximation, the auxiliary distribution is factorized as $q_{\boldsymbol{\lambda}}(\boldsymbol{h}, \boldsymbol{G}|\boldsymbol{Y}) = q_{\boldsymbol{\lambda}_1}(\boldsymbol{h}|\boldsymbol{Y}) \cdot q_{\boldsymbol{\lambda}_2}(\boldsymbol{G}|\boldsymbol{Y})$. We use neural networks *Encoder* $\mathcal{F}$ and *Encoder* $\mathcal{G}$ to predict the parameters of the auxiliary distributions $\boldsymbol{\lambda}_1$ and $\boldsymbol{\lambda}_2$, respectively. The ELBO function is given by:

$$\begin{aligned}\mathcal{L}^{\mathsf{I-CSI-Rayleigh}}(\boldsymbol{\lambda}_1, \boldsymbol{\lambda}_2) = &\underbrace{\mathbb{E}_{\boldsymbol{h}\sim q_{\boldsymbol{\lambda}_1}(\boldsymbol{h}|\boldsymbol{Y})}\left[\log \frac{q_{\boldsymbol{\lambda}_1}(\boldsymbol{h}|\boldsymbol{Y})}{p(\boldsymbol{h})}\right]}_{\mathcal{L}_1^{\mathsf{I-CSI-Rayleigh}}} + \underbrace{\mathbb{E}_{\boldsymbol{G}\sim q_{\boldsymbol{\lambda}_2}(\boldsymbol{G}|\boldsymbol{Y})}\left[\log \frac{q_{\boldsymbol{\lambda}_2}(\boldsymbol{G}|\boldsymbol{Y})}{p(\boldsymbol{G})}\right]}_{\mathcal{L}_2^{\mathsf{I-CSI-Rayleigh}}} \\ &\underbrace{-\mathbb{E}_{\boldsymbol{h},\boldsymbol{G}\sim q_{\boldsymbol{\lambda}}(\boldsymbol{h},\boldsymbol{G}|\boldsymbol{Y})}\left[\log p(\boldsymbol{Y}|\boldsymbol{h}, \boldsymbol{G})\right]}_{\mathcal{L}_3^{\mathsf{I-CSI-Rayleigh}}}. \end{aligned} \tag{2.9}$$

Considering the Rayleigh fading channel for $\boldsymbol{h}$ and $\boldsymbol{G}$, the prior of the channels are assumed to be independent and identically distributed (i.i.d) complex Gaussian random variables with variances $\sigma_{\boldsymbol{h}}^2$ and $\sigma_{\boldsymbol{G}}^2$:

$$p(\boldsymbol{h}) = \prod_{i=1}^{N} p\left(\boldsymbol{h}_i\right) \qquad ; \qquad p\big(\boldsymbol{h}_i\big) \sim \mathcal{CN}(0, \sigma_{\boldsymbol{h}}^2) \tag{2.10}$$

$$p(\boldsymbol{G}) = \prod_{i=1}^{M} \prod_{j=1}^{N} p\left(\boldsymbol{G}_{i,j}\right) \qquad ; \qquad p\big(\boldsymbol{G}_{i,j}\big) \sim \mathcal{CN}(0, \sigma_{\boldsymbol{G}}^2) \tag{2.11}$$

Additionally, the approximated posterior distributions $q_{\boldsymbol{\lambda}_1}(\boldsymbol{h}|\boldsymbol{Y})$ and $q_{\boldsymbol{\lambda}_2}(\boldsymbol{G}|\boldsymbol{Y})$ are assumed to be complex Gaussian distributions with independent elements:

$$q_{\boldsymbol{\lambda}_1}(\boldsymbol{h}|\boldsymbol{Y}) \sim \mathcal{CN}(\boldsymbol{m}_i, \boldsymbol{\gamma}_i) \tag{2.12}$$

$$q_{\boldsymbol{\lambda}_2}(\boldsymbol{G}|\boldsymbol{Y}) \sim \mathcal{CN}(\boldsymbol{M}_{i,j}, \boldsymbol{\Gamma}_{i,j}) \tag{2.13}$$

where $\boldsymbol{\lambda}_1 = \{\boldsymbol{m}, \boldsymbol{\gamma}\}$ and $\boldsymbol{\lambda}_2 = \{\boldsymbol{M}, \boldsymbol{\Gamma}\}$ are the parameters of the auxiliary distributions of means and variances, respectively, which are obtained by maximizing the ELBO function in Eq. 2.9.

We recall that $\mathcal{L}_1^{\mathsf{I-CSI-Rayleigh}}$ has a closed-form expression since it is the KL divergence between two complex Gaussian distributions $p(\boldsymbol{h})$ and $q_{\boldsymbol{\lambda}_1}(\boldsymbol{h}|\boldsymbol{Y})$:

$$\begin{aligned}
\mathcal{L}_1^{\mathsf{I-CSI-Rayleigh}} &= \mathbb{E}_{\boldsymbol{h}\sim q_{\boldsymbol{\lambda}_1}(\boldsymbol{h}|\boldsymbol{Y})}\left[\log \frac{q_{\boldsymbol{\lambda}_1}(\boldsymbol{h}|\boldsymbol{Y})}{p(\boldsymbol{h})}\right] \\
&= \frac{1}{\sigma_{\boldsymbol{h}}^2}\Big(\mathrm{Tr}\big(\mathrm{diag}(\boldsymbol{\gamma}_{\boldsymbol{h}})\big) + \boldsymbol{m}^H\boldsymbol{m}\Big) - \log\Big|\mathrm{diag}(\boldsymbol{\gamma})\Big| + N\log\sigma_{\boldsymbol{h}}^2 - N \\
&= \frac{1}{\sigma_{\boldsymbol{h}}^2}\sum_{i=1}^{N}\boldsymbol{\gamma}_i + \frac{\boldsymbol{m}^H\boldsymbol{m}}{\sigma_{\boldsymbol{h}}^2} - \sum_{i=1}^{N}\log\boldsymbol{\gamma}_i + N\log\sigma_{\boldsymbol{h}}^2 - N.
\end{aligned} \tag{2.14}$$

Similarly, the closed-form expression of $\mathcal{L}_2^{\mathsf{I-CSI-Rayleigh}}$ is expressed as follows:

$$\begin{aligned}
\mathcal{L}_2^{\mathsf{I-CSI-Rayleigh}} &= \mathbb{E}_{\boldsymbol{G}\sim q_{\boldsymbol{\lambda}_2}(\boldsymbol{G}|\boldsymbol{Y})}\left[\log \frac{q_{\boldsymbol{\lambda}_2}(\boldsymbol{G}|\boldsymbol{Y})}{p(\boldsymbol{G})}\right] \\
&= \frac{1}{\sigma_{\boldsymbol{G}}^2}\Bigg(\mathrm{Tr}\Big(\mathrm{diag}\big(\mathrm{vec}(\boldsymbol{\Gamma})\big)\Big) + \mathrm{vec}(\boldsymbol{M})^H\mathrm{vec}(\boldsymbol{M})\Bigg) - \log\Big|\mathrm{diag}\big(\mathrm{vec}(\boldsymbol{\Gamma})\big)\Big| \\
&\quad + NM\log\sigma_{\boldsymbol{G}}^2 - NM \\
&= \frac{1}{\sigma_{\boldsymbol{G}}^2}\sum_{i=1}^{M}\sum_{j=1}^{N}(\boldsymbol{\Gamma}_{i,j} - \log\boldsymbol{\Gamma}_{i,j}) + \frac{\mathrm{vec}(\boldsymbol{M})^H\mathrm{vec}(\boldsymbol{M})}{\sigma_{\boldsymbol{G}}^2} + NM\log\sigma_{\boldsymbol{G}}^2 - NM.
\end{aligned} \tag{2.15}$$

Regarding the reconstruction loss $\mathcal{L}_3^{\mathsf{I-CSI-Rayleigh}}$, we note that $p(\boldsymbol{Y}|\boldsymbol{G},\boldsymbol{h}) = \prod_{l=1}^{N_p} p(\boldsymbol{y}_l|\boldsymbol{G},\boldsymbol{h})$ where $\boldsymbol{y}_l$ is the l-th received pilot signal $p(\boldsymbol{y}_l|\boldsymbol{G},\boldsymbol{h}) \sim \mathcal{CN}(\sqrt{\rho}\ \boldsymbol{G}\ \text{diag}(\boldsymbol{v}_l)\ \boldsymbol{h}\ x_l,\ \sigma_{\boldsymbol{w}}^2\boldsymbol{I}_M)$ due to the independent realizations of the noise for each transmission of the pilot signal x_l. To derive the expression of $\mathcal{L}_3^{\mathsf{I-CSI-Rayleigh}}$, we apply the expectation over $\boldsymbol{h}$ in a first step and obtain a closed-form expression, then we use the property $\mathbb{E}[\boldsymbol{G}^H\boldsymbol{G}] = \text{diag}(\boldsymbol{\tau}) + \boldsymbol{M}^H\boldsymbol{M}$ where $\text{diag}(\boldsymbol{\tau})$ represents the covariance matrix of the columns of $\boldsymbol{G}$ and $\boldsymbol{\tau}_i = \sum_{m=1}^{M} \boldsymbol{\Gamma}_{m,i}$ to compute the expectation over $\boldsymbol{G}$:

$$
\begin{aligned}
\mathcal{L}_3^{\mathsf{I-CSI-Rayleigh}} &= -\mathbb{E}_{\boldsymbol{h},\boldsymbol{G}\sim q_{\boldsymbol{\lambda}}(\boldsymbol{h},\boldsymbol{G}|\boldsymbol{Y})}\big[\log p(\boldsymbol{Y}|\boldsymbol{h},\boldsymbol{G})\big] \\
&= -\sum_{l=1}^{L} \mathbb{E}_{\boldsymbol{h},\boldsymbol{G}\sim q_{\boldsymbol{\lambda}}(\boldsymbol{h},\boldsymbol{G}|\boldsymbol{Y})}\big[\log p(\boldsymbol{y}_l|\boldsymbol{h},\boldsymbol{G})\big] \\
&= \mathbb{E}_{\boldsymbol{h},\boldsymbol{G}\sim q_{\boldsymbol{\lambda}}(\boldsymbol{h},\boldsymbol{G}|\boldsymbol{Y})}\Bigg[\frac{1}{\sigma_{\boldsymbol{w}}^2}\sum_{l=1}^{N_p}\big(\boldsymbol{y}_l - \sqrt{\rho}\boldsymbol{G}\text{diag}(\boldsymbol{v}_l)\boldsymbol{h}x_l\big)^H \\
&\qquad \times \big(\boldsymbol{y}_l - \sqrt{\rho}\boldsymbol{G}\text{diag}(\boldsymbol{v}_l)\boldsymbol{h}x_l\big)\Bigg] + C_1 \\
&= \frac{1}{\sigma_{\boldsymbol{w}}^2}\Big[\rho||\boldsymbol{x}||_2^2\text{Tr}\big(\boldsymbol{M}^H\boldsymbol{M}\text{diag}(\boldsymbol{\gamma})\big) + \sum_{l=1}^{L}||\boldsymbol{y}_l - \sqrt{\rho}\boldsymbol{M}\text{diag}(\boldsymbol{v}_l)\boldsymbol{m}x_l||_2^2 \\
&\qquad + \rho||\boldsymbol{x}||_2^2\text{Tr}\big(\text{diag}(\boldsymbol{\tau})\ \text{diag}(\boldsymbol{\gamma})\big) + \rho||\boldsymbol{x}||_2^2\boldsymbol{m}^H\text{diag}(\boldsymbol{\tau})\boldsymbol{m}\Big] + C_1, \quad (2.16)
\end{aligned}
$$

where C_1 is a constant term. The detailed derivation can be found in the Appendix.

2.3.2 Derivation of ELBO For mmWave Channels

Due to the large number of elements in the RIS and the high path loss, the channels are more sparse in the angular domain [46–48]. Specifically, mmWave channels are often modeled as sparse in the angular domain, meaning that only a small number of paths contribute to the received signal, and the other paths are negligible because of the high path loss and directional nature of mmWave signals. Therefore, we model the sparse channels as channels whose elements follow the complex Laplace distributions in the

angular domain. In particular, by applying Discrete Fourier Transform (DFT) to the channels $\boldsymbol{G}$ and $\boldsymbol{h}$, the obtained channels are:

$$\boldsymbol{G}^{\mathsf{vir}} = \boldsymbol{F}_M \boldsymbol{G} \boldsymbol{F}_N; \tag{2.17}$$

$$\boldsymbol{h}^{\mathsf{vir}} = \boldsymbol{F}_N \boldsymbol{h}, \tag{2.18}$$

where $\boldsymbol{F}_N$ and $\boldsymbol{F}_M$ are the DFT matrices of size $N \times N$ and $M \times M$, respectively. Then, $\boldsymbol{G}^{\mathsf{vir}}$ and $\boldsymbol{h}^{\mathsf{vir}}$ are the channels in the angular domain where the elements are i.i.d and drawn from a complex Laplace distribution with zero mean and scales $\alpha_{\boldsymbol{G}^{\mathsf{vir}}}$ and $\alpha_{\boldsymbol{h}^{\mathsf{vir}}}$, respectively, i.e. $\boldsymbol{G}^{\mathsf{vir}}_{i,j} \sim \mathcal{CL}(0, \alpha_{\boldsymbol{G}^{\mathsf{vir}}})$ and $\boldsymbol{h}^{\mathsf{vir}}_i \sim \mathcal{CL}(0, \alpha_{\boldsymbol{h}^{\mathsf{vir}}})$. Given that $\boldsymbol{F}_N^{-1} = \frac{1}{N}\boldsymbol{F}_N^H$ for any DFT matrix of size $N \times N$, the received training signal for the l-th time slot is expressed as follows:

$$\boldsymbol{y}_l = \frac{\sqrt{\rho}}{MN^2} \boldsymbol{F}_M^H \boldsymbol{G}^{\mathsf{vir}} \boldsymbol{F}_N^H \operatorname{diag}(\boldsymbol{v}_l) \boldsymbol{F}_N^H \boldsymbol{h}^{\mathsf{vir}} x_l + \boldsymbol{w}_l, \qquad l = 1, \ldots, N_p. \tag{2.19}$$

By applying the VI framework, we approximate the intractable true posterior distribution $p(\boldsymbol{h}^{\mathsf{vir}}, \boldsymbol{G}^{\mathsf{vir}} | \boldsymbol{Y})$ by a tractable parameterized distribution denoted $q_{\boldsymbol{\lambda}}(\boldsymbol{h}^{\mathsf{vir}}, \boldsymbol{G}^{\mathsf{vir}} | \boldsymbol{Y})$ that maximize the ELBO function. Under the assumption of a low correlation between the channels $\boldsymbol{h}^{\mathsf{vir}}$ and $\boldsymbol{G}^{\mathsf{vir}}$ conditioned on the training signal $\boldsymbol{Y}$, the auxiliary distribution can be factorized as $q_{\boldsymbol{\lambda}}(\boldsymbol{h}^{\mathsf{vir}}, \boldsymbol{G}^{\mathsf{vir}} | \boldsymbol{Y}) = q_{\boldsymbol{\lambda}_1}(\boldsymbol{h}^{\mathsf{vir}} | \boldsymbol{Y}) \cdot q_{\boldsymbol{\lambda}_2}(\boldsymbol{G}^{\mathsf{vir}} | \boldsymbol{Y})$. We use neural networks to predict the parameters of the auxiliary distributions $\boldsymbol{\lambda}_1$ and $\boldsymbol{\lambda}_2$.

We assume that the auxiliary distributions follow complex Laplace distributions with independent elements:

$$q_{\boldsymbol{\lambda}_1}\big(\boldsymbol{h}^{\mathsf{vir}}_i | \boldsymbol{Y}\big) \sim \mathcal{CL}\big(\boldsymbol{m}_{i,j}, \boldsymbol{b}_i\big) \qquad \forall i; \tag{2.20}$$

$$q_{\boldsymbol{\lambda}_2}\big(\boldsymbol{G}^{\mathsf{vir}}_{i,j} | \boldsymbol{Y}\big) \sim \mathcal{CL}\big(\boldsymbol{M}_{i,j}, \boldsymbol{B}_{i,j}\big) \qquad \forall i, j, \tag{2.21}$$

where $\boldsymbol{\lambda}_1 = \{\boldsymbol{m}, \boldsymbol{b}\}$ and $\boldsymbol{\lambda}_2 = \{\boldsymbol{M}, \boldsymbol{B}\}$ are the parameters of the auxiliary distributions of means and scales which are obtained by maximizing the ELBO function, which is

given in general form in Eq. 2.3, expressed as follows:

$$\mathcal{L}^{\mathsf{I-CSI-mmWave}}(\boldsymbol{\lambda}_1,\boldsymbol{\lambda}_2) = \underbrace{\mathbb{E}_{\boldsymbol{h}^{\mathsf{vir}}\sim q_{\boldsymbol{\lambda}_1}(\boldsymbol{h}^{\mathsf{vir}}|\boldsymbol{Y})}\left[\log\frac{q_{\boldsymbol{\lambda}_1}(\boldsymbol{h}^{\mathsf{vir}}|\boldsymbol{Y})}{p(\boldsymbol{h}^{\mathsf{vir}})}\right]}_{\mathcal{L}_1^{\mathsf{I-CSI-mmWave}}} + \underbrace{\mathbb{E}_{\boldsymbol{G}^{\mathsf{vir}}\sim q_{\boldsymbol{\lambda}_2}(\boldsymbol{G}^{\mathsf{vir}}|\boldsymbol{Y})}\left[\log\frac{q_{\boldsymbol{\lambda}_2}(\boldsymbol{G}^{\mathsf{vir}}|\boldsymbol{Y})}{p(\boldsymbol{G}^{\mathsf{vir}})}\right]}_{\mathcal{L}_2^{\mathsf{I-CSI-mmWave}}}$$
$$\underbrace{-\mathbb{E}_{\boldsymbol{h}^{\mathsf{vir}},\boldsymbol{G}^{\mathsf{vir}}\sim q_{\boldsymbol{\lambda}}(\boldsymbol{h}^{\mathsf{vir}},\boldsymbol{G}^{\mathsf{vir}}|\boldsymbol{Y})}\big[\log p(\boldsymbol{Y}|\boldsymbol{h}^{\mathsf{vir}},\boldsymbol{G}^{\mathsf{vir}})\big]}_{\mathcal{L}_3^{\mathsf{I-CSI-mmWave}}}. \tag{2.22}$$

The first loss $\mathcal{L}_1^{\mathsf{I-CSI-mmWave}}$ is the KL-divergence between the auxiliary distribution and the prior of $\boldsymbol{h}^{\mathsf{vir}}$, which can be expressed as follows:

$$\begin{aligned}\mathcal{L}_1^{\mathsf{I-CSI-mmWave}}(\boldsymbol{\lambda}_1) &= \mathbb{E}_{\boldsymbol{h}^{\mathsf{vir}}\sim q_{\boldsymbol{\lambda}_1}(\boldsymbol{h}^{\mathsf{vir}}|\boldsymbol{Y})}\big[\log q_{\boldsymbol{\lambda}_1}(\boldsymbol{h}^{\mathsf{vir}}|\boldsymbol{Y})\big] - \mathbb{E}_{\boldsymbol{h}^{\mathsf{vir}}\sim q_{\boldsymbol{\lambda}_1}(\boldsymbol{h}^{\mathsf{vir}}|\boldsymbol{Y})}\big[\log p(\boldsymbol{h}^{\mathsf{vir}})\big]\\ &= \sum_{i=1}^{N}\mathbb{E}_{\boldsymbol{h}_i^{\mathsf{vir}}\sim q_{\boldsymbol{\lambda}_1}(\boldsymbol{h}_i^{\mathsf{vir}}|\boldsymbol{Y})}\big[\log q_{\boldsymbol{\lambda}_1}(\boldsymbol{h}_i^{\mathsf{vir}}|\boldsymbol{Y})\big] - \mathbb{E}_{\boldsymbol{h}_i^{\mathsf{vir}}\sim q_{\boldsymbol{\lambda}_1}(\boldsymbol{h}_i^{\mathsf{vir}}|\boldsymbol{Y})}\big[\log p(\boldsymbol{h}_i^{\mathsf{vir}})\big]\\ &= \sum_{i=1}^{N} H\Big(q_{\boldsymbol{\lambda}_1}\big(\boldsymbol{h}_i^{\mathsf{vir}}|\boldsymbol{Y}\big), p\big(\boldsymbol{h}_i^{\mathsf{vir}}\big)\Big) - H\Big(q_{\boldsymbol{\lambda}_1}\big(\boldsymbol{h}_i^{\mathsf{vir}}|\boldsymbol{Y}\big)\Big),\end{aligned} \tag{2.23}$$

where $H\Big(q_{\boldsymbol{\lambda}_1}\big(\boldsymbol{h}_i^{\mathsf{vir}}|\boldsymbol{Y}\big)\Big)$ is the entropy of $q_{\boldsymbol{\lambda}_1}\big(\boldsymbol{h}_i^{\mathsf{vir}}|\boldsymbol{Y}\big)$ and $H\Big(q_{\boldsymbol{\lambda}_1}\big(\boldsymbol{h}_i^{\mathsf{vir}}|\boldsymbol{Y}\big), p\big(\boldsymbol{h}_i^{\mathsf{vir}}\big)\Big)$ is the cross entropy between $q_{\boldsymbol{\lambda}_1}\big(\boldsymbol{h}_i^{\mathsf{vir}}|\boldsymbol{Y}\big)$ and $p\big(\boldsymbol{h}_i^{\mathsf{vir}}\big)$. The entropy of the complex Laplace distribution is:

$$H\Big(q_{\boldsymbol{\lambda}_1}\big(\boldsymbol{h}_i^{\mathsf{vir}}|\boldsymbol{Y}\big)\Big) = \log(2\pi\boldsymbol{b}_i^2) + 2. \tag{2.24}$$

The proof can be found in the Appendix. The cross-entropy between two Laplace distributions can be obtained using Monte-Carlo method to approximate the expectation over $\boldsymbol{h}^{\mathsf{vir}}$. Therefore, it is given by:

$$\begin{aligned}H\Big(q_{\boldsymbol{\lambda}_1}\big(\boldsymbol{h}_i^{\mathsf{vir}}|\boldsymbol{Y}\big), p\big(\boldsymbol{h}_i^{\mathsf{vir}}\big)\Big) &= \mathbb{E}_{\boldsymbol{h}_i^{\mathsf{vir}}\sim q_{\boldsymbol{\lambda}_1}(\boldsymbol{h}_i^{\mathsf{vir}}|\boldsymbol{Y})}\Big[-\log p\big(\boldsymbol{h}_i^{\mathsf{vir}}\big)\Big]\\ &= \log(2\pi\alpha_{\boldsymbol{h}^{\mathsf{vir}}}^2) + \mathbb{E}_{\boldsymbol{h}_i^{\mathsf{vir}}\sim q_{\boldsymbol{\lambda}_1}(\boldsymbol{h}_i^{\mathsf{vir}}|\boldsymbol{Y})}\left[\frac{\big|\boldsymbol{h}_i^{\mathsf{vir}}\big|}{\alpha_{\boldsymbol{h}^{\mathsf{vir}}}}\right]\\ &\approx \log(2\pi\alpha_{\boldsymbol{h}^{\mathsf{vir}}}^2) + \frac{1}{D}\sum_{d=1}^{D}\frac{\big|\widehat{\boldsymbol{h}_i^{\mathsf{vir}}}^{(d)}\big|}{\alpha_{\boldsymbol{h}^{\mathsf{vir}}}},\end{aligned} \tag{2.25}$$

where the d-th sample is computed as $\widehat{\boldsymbol{h}_i^{\mathsf{vir}}}^{(d)} = \boldsymbol{m}_i + \boldsymbol{b}_i \times \mathcal{CL}(0,1)$. Hence, the $\mathcal{L}_1^{\mathsf{l-CSI-mmWave}}$ is expressed as:

$$\mathcal{L}_1^{\mathsf{l-CSI-mmWave}}(\boldsymbol{\lambda}_1) = \frac{1}{D}\sum_{i=1}^{N}\sum_{d=1}^{D}\frac{\left|\widehat{\boldsymbol{h}_i^{\mathsf{vir}}}^{(d)}\right|}{\alpha_{\boldsymbol{h}^{\mathsf{vir}}}} - \sum_{i=1}^{N}\log(2\pi\boldsymbol{b}_i^2) + N\log(2\pi\alpha_{\boldsymbol{h}^{\mathsf{vir}}}^2) - 2N. \quad (2.26)$$

Similarly, we derive $\mathcal{L}_2^{\mathsf{l-CSI-mmWave}}$ using the independence propriety of the elements $q_{\boldsymbol{\lambda}_2}(\boldsymbol{G}_{i,j}^{\mathsf{vir}}|\boldsymbol{Y})$ and $p(\boldsymbol{G}_{i,j}^{\mathsf{vir}})$:

$$\begin{aligned}\mathcal{L}_2^{\mathsf{l-CSI-mmWave}}(\boldsymbol{\lambda}_2) &= \mathbb{E}_{\boldsymbol{G}^{\mathsf{vir}}\sim q_{\boldsymbol{\lambda}_2}(\boldsymbol{G}^{\mathsf{vir}}|\boldsymbol{Y})}\big[\log q_{\boldsymbol{\lambda}_2}(\boldsymbol{G}^{\mathsf{vir}}|\boldsymbol{Y})\big] - \mathbb{E}_{\boldsymbol{G}^{\mathsf{vir}}\sim q_{\boldsymbol{\lambda}_2}(\boldsymbol{G}^{\mathsf{vir}}|\boldsymbol{Y})}\big[\log p(\boldsymbol{G}^{\mathsf{vir}})\big] \\ &= \frac{1}{D}\sum_{i=1}^{M}\sum_{j=1}^{N}\sum_{d=1}^{D}\frac{\left|\widehat{\boldsymbol{G}_{i,j}^{\mathsf{vir}}}^{(d)}\right|}{\alpha_{\boldsymbol{G}^{\mathsf{vir}}}} - \sum_{i=1}^{M}\sum_{j=1}^{N}\log(2\pi\boldsymbol{B}_{i,j}^2) + NM\log(2\pi\alpha_{\boldsymbol{G}^{\mathsf{vir}}}^2) \\ &\quad - 2NM, \end{aligned} \quad (2.27)$$

where the Monte-Carlo samples are computed as $\widehat{\boldsymbol{G}_{i,j}^{\mathsf{vir}}}^{(d)} = \boldsymbol{M}_{i,j} + \boldsymbol{B}_{i,j} \times \mathcal{CL}(0,1)$. The third loss consists of the expectation over the auxiliary distribution of the log-likelihood of the received training signal. It can be derived in closed-form as in Eq. 2.28:

$$\begin{aligned}\mathcal{L}_3^{\mathsf{l-CSI}}(\boldsymbol{\lambda}) &= -\mathbb{E}_{\boldsymbol{h}^{\mathsf{vir}},\boldsymbol{G}^{\mathsf{vir}}\sim q_{\boldsymbol{\lambda}}(\boldsymbol{h}^{\mathsf{vir}},\boldsymbol{G}^{\mathsf{vir}}|\boldsymbol{Y})}\left[\log p(\boldsymbol{Y}|\boldsymbol{h}^{\mathsf{vir}},\boldsymbol{G}^{\mathsf{vir}})\right] \\ &= -\sum_{l=1}^{N_p}\mathbb{E}_{\boldsymbol{h}^{\mathsf{vir}},\boldsymbol{G}^{\mathsf{vir}}\sim q_{\boldsymbol{\lambda}}(\boldsymbol{h}^{\mathsf{vir}},\boldsymbol{G}^{\mathsf{vir}}|\boldsymbol{Y})}\left[\log p(\boldsymbol{y}_l|\boldsymbol{h}^{\mathsf{vir}},\boldsymbol{G}^{\mathsf{vir}})\right] \\ &= \sum_{l=1}^{N_p}\Bigg[(\boldsymbol{y}_l - \frac{\sqrt{\rho}}{MN^2}\boldsymbol{F}_M^H\boldsymbol{M}\boldsymbol{F}_N^H\mathrm{diag}(\boldsymbol{v}_l)\boldsymbol{F}_N^H\boldsymbol{m}x_l)^H \quad (2.28) \\ &\quad\times(\boldsymbol{y}_l - \frac{\sqrt{\rho}}{MN^2}\boldsymbol{F}_M^H\boldsymbol{M}\boldsymbol{F}_N^H\mathrm{diag}(\boldsymbol{v}_l)\boldsymbol{F}_N^H\boldsymbol{m}x_l) \\ &\quad+\frac{\rho|x_l|^2}{MN^4}\cdot\mathrm{Tr}\big(\boldsymbol{\Lambda}\boldsymbol{F}_N^H\mathrm{diag}(\boldsymbol{v}_l)\boldsymbol{F}_N^H\boldsymbol{Q}\boldsymbol{F}_N\mathrm{diag}(\boldsymbol{v}_l)^H\boldsymbol{F}_N\big) \quad (2.29) \\ &\quad+\frac{\rho|x_l|^2}{MN^4}\,\mathrm{Tr}\big(\boldsymbol{M}^H\boldsymbol{M}\boldsymbol{F}_N^H\mathrm{diag}(\boldsymbol{v}_l)\boldsymbol{F}_N^H\boldsymbol{Q}\boldsymbol{F}_N\mathrm{diag}(\boldsymbol{v}_l)^H\boldsymbol{F}_N\big) \\ &\quad+\frac{\rho|x_l|^2}{MN^4}\,\boldsymbol{m}^H\boldsymbol{F}_N\mathrm{diag}(\boldsymbol{v}_l)^H\boldsymbol{F}_N\boldsymbol{\Lambda}\boldsymbol{F}_N^H\mathrm{diag}(\boldsymbol{v}_l)\boldsymbol{F}_N^H\boldsymbol{m}\Bigg] + C_1. \quad (2.30)\end{aligned}$$

where C_1 is a constant, $\boldsymbol{Q}$ and $\boldsymbol{\Lambda}$ are the covariance matrix over the columns of $\boldsymbol{G}^{\mathsf{vir}}$ and covariance matrix of $\boldsymbol{h}^{\mathsf{vir}}$, respectively, which are diagonal matrices due to the independence of the elements according to the auxiliary distributions. The main diagonal elements are as follows (see proof in the Appendix.):

$$\boldsymbol{\Lambda}_{i,i} = 6\boldsymbol{b}_i^2;\ \boldsymbol{Q}_{i,i} = 6\sum_{m=1}^{M} \boldsymbol{B}_{m,i}^2. \tag{2.31}$$

2.4 Optimization of the phase-shifts based on I-CSI

The primary evaluation metric is the capacity of the RIS-assisted network obtained using the phase-shifts derived from the estimated quantities. In this subsection, we derive closed-form expressions of the phase-shifts of the RIS that maximize the capacities using the I-CSI. Although the proposed solution is not an optimal method to obtain the phase-shifts given the I-CSI, it is used to evaluate the performance of the estimated I-CSI in terms of the capacity.

For the considered uplink RIS-assisted mmWave system, the received signal at the BS can be expressed as follows:

$$\boldsymbol{y} = \sqrt{\rho}\, \boldsymbol{G}\, \mathrm{diag}(\boldsymbol{v})\, \boldsymbol{h}\, x + \boldsymbol{w}, \tag{2.32}$$

where x is the transmitted symbol satisfying $\mathbb{E}(|x|^2) = 1$, ρ is the SNR, and $\boldsymbol{w} \sim \mathcal{CN}(\boldsymbol{0}, \boldsymbol{I}_M)$ denotes the additive white noise. The ergodic capacity is expressed by:

$$C = \log_2\big(1 + \rho\, ||\boldsymbol{G}\, \mathrm{diag}(\boldsymbol{v})\, \boldsymbol{h}||_2^2\big). \tag{2.33}$$

Based on the I-CSI, i.e. $\boldsymbol{h}$ and $\boldsymbol{G}$, we configure the phase-shifts to maximize the capacity C, which is equivalent to solving the following problem:

$$\begin{aligned} &\max_{\{\theta_i\}}\ ||\boldsymbol{G}\, \mathrm{diag}(\boldsymbol{v})\, \boldsymbol{h}||_2^2 \\ &\text{Subject to: } \boldsymbol{v}_i = e^{\theta_i}. \end{aligned} \tag{2.34}$$

Given the singular value decomposition (SVD) of $\boldsymbol{G} = \boldsymbol{U}\boldsymbol{S}\boldsymbol{V}^H$, the problem is equivalent to maximizing $||\boldsymbol{S}\boldsymbol{V}^{\boldsymbol{H}}\ \text{diag}(\boldsymbol{v})\ \boldsymbol{h}||_2^2$ which is expressed as follows:

$$||\boldsymbol{S}\boldsymbol{V}^H \text{diag}(\boldsymbol{v})\boldsymbol{h}||_2^2 = \sum_{i=1}^{M} \left| \sum_{k=1}^{N} \boldsymbol{S}_{i,i}\boldsymbol{V}_{ki}^* \boldsymbol{h}_k \boldsymbol{v}_k \right|^2 \tag{2.35}$$

$$= \sum_{i=1}^{M} \left| \sum_{k=1}^{N} \boldsymbol{S}_{i,i}|\boldsymbol{V}_{ki}||\boldsymbol{h}_k| e^{\theta_k - \angle \boldsymbol{V}_{ki} + \angle \boldsymbol{h}_k} \right|^2 . \tag{2.36}$$

The solution we propose is to align the phase-shifts θ_k to the phases of the largest right singular vector of $\boldsymbol{G}$, denoted as $\boldsymbol{\vartheta}^{\max}$, and the phases of the channel vector $\boldsymbol{h}$, especially when the UE-RIS channel $\boldsymbol{G}$ is low-rank, and therefore, most of the values of $\boldsymbol{S}_{ii}$ are zeros and do not contribute into the summation. Specifically, the suboptimal phase-shifts are obtained as follows:

$$\theta_k^* = -(\angle \boldsymbol{h}_k - \angle \boldsymbol{\vartheta}_k^{\max}). \tag{2.37}$$

2.5 Simulations and Results

In this section, we evaluate the performance of the proposed I-CSI estimation methods in RIS-aided SIMO mmWave wireless communication systems. The approach is dubbed Joint Channel Estimation (JCE).

2.5.1 Evaluation Metrics and Baselines

the main performance metric is the capacity of the RIS-aided SIMO communication system. We leverage the estimated quantities, specifically the UE-RIS and RIS-BS channels to calculate the phase-shifts and determine the achieved capacity defined as $C = \log_2(1 + \rho||\boldsymbol{G}\ \text{diag}(\boldsymbol{v})\ \boldsymbol{h}||_2^2)$. Moreover, we evaluate the normalized mean square error (NMSE) defined by $\text{NMSE} = ||\hat{\boldsymbol{X}} - \boldsymbol{X}||^2/||\boldsymbol{X}||^2$, where Frobenius norm is used for the channel matrix $\boldsymbol{G}$ and l_2 norm for the channel vector $\boldsymbol{h}$.

We compare our approach against the following baselines:

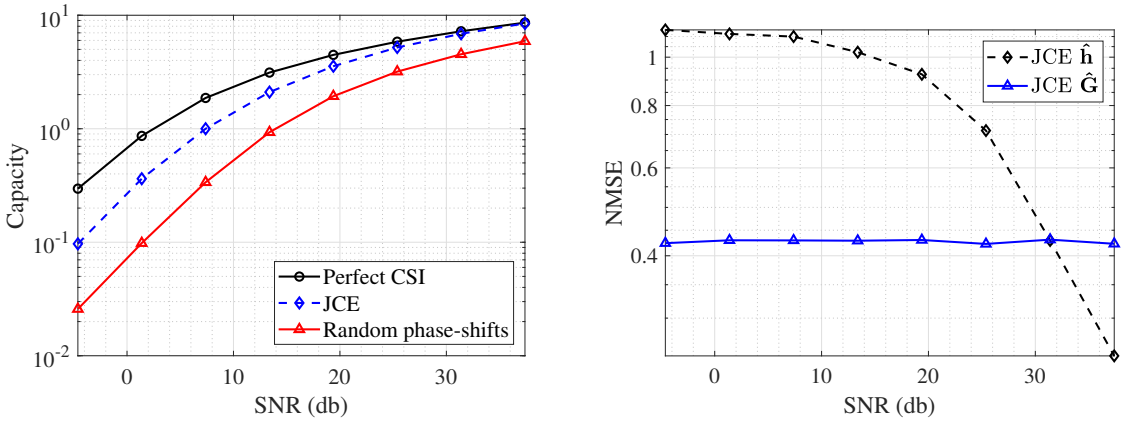
- **Perfect CSI**: this is an upper bound where the capacity is obtained based on the optimal phase shifts using the true channels $\boldsymbol{G}$ and $\boldsymbol{h}$;
- **Random phase-shifts**: this represents a lower bound for our method.
- **MO-EST**: This method is based on alternating minimization and manifold optimization [49].

2.5.2 Simulation Setup

We consider a RIS composed of $N = 64$ passive elements and a BS equipped with $M = 4$ antennas. To estimate the UE-RIS and the RIS-BS channels, we send $N_p = 50$ pilot symbols over an uplink SIMO RIS-assisted mmWave communication system, and obtain the training signals which are fed to the trained neural networks *Encoder* $\mathcal{F}$ and *Encoder* $\mathcal{G}$. The encoders are fully connected neural networks where each network consists of an input layer, two 300 unit hidden layers with Relu activation and an output layer with two heads: the first outputs the mean after a Tanh activation and the second uses Softmax activation for the scale. Adam optimizer [50] is used to train the neural networks with 0.001 as an initial learning rate. The neural networks are trained by maximizing the ELBO functions using unlabeled of 10^4 samples. The non-closed form expectations within the ELBO functions are evaluated under 1000 samples. The methods are evaluated based on 2500 Monte-Carlo samples.

2.5.3 Performance of Joint Channel Estimation: Gaussian prior and Laplace prior in the angular domain

Figures 2.3(a) and 2.3(b) depict the channel estimation performance with Rayleigh channel model with $\sigma^2_{\boldsymbol{G}} = \sigma^2_{\boldsymbol{h}} = 1$. In Fig. 2.3(a), we illustrate the capacity as a



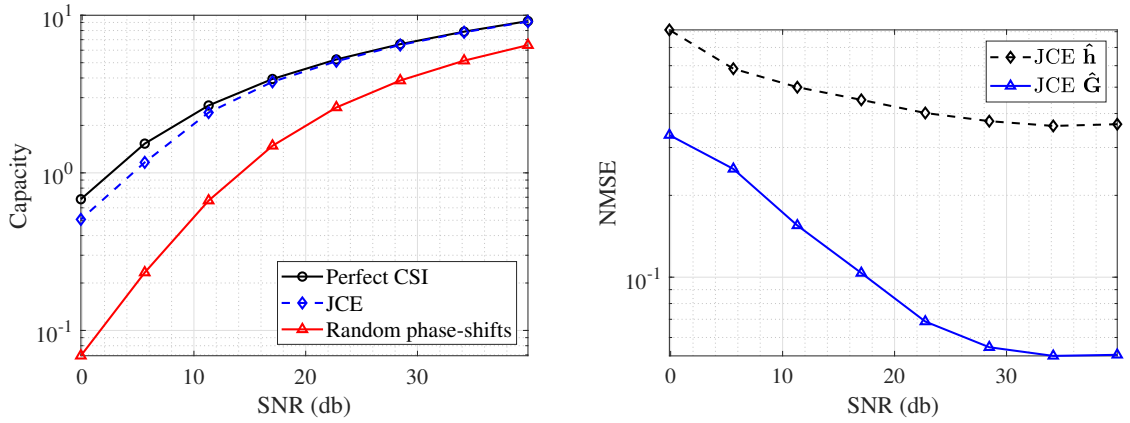
(a) Achieved capacity

(b) NMSE

Figure 2.3: Performance of the VI-based RIS joint channel estimation using complex Gaussian channels.

function of the signal-to-noise ratio (SNR) for different methods of the capacities. The phase shifts derived from the estimated channels are able to achieve a better capacity than the random selection of the RIS configuration which validates that the neural networks were able to effectively learn the channels. Also, we observe that, with high SNR, the capacity gets closer to the exact capacity. In Fig. 2.3(b), we evaluate the NMSE and we observe that with increasing SNR, the NMSE decreases for the UE-RIS channel while being constant for the RIS-BS due to the large number of elements the channel matrix is composed of.

In Fig. 2.4(a), we evaluate the performance of the sparse channel models where the elements of the channels in the angular domain are sampled from complex Laplace distribution with unit scale, i.e. $\alpha_{\bar{\boldsymbol{G}}} = 1$ and $\alpha_{\bar{\boldsymbol{h}}} = 1$. We observe that the capacity obtained using the proposed method is closer to the upper bound which is the capacity with the optimal phase shifts. Moreover, Fig. 2.4(b) shows that the NMSE decreases with the SNR. The performance with sparse channels is better than the performance we



(a) Achieved capacity (b) NMSE

Figure 2.4: Performance of the VI-based RIS joint channel estimation using complex Laplace channels.

obtained using Gaussian models thanks to the structure of sparsity added to the models. Since the model applied is more structured than Gaussian channels which has a larger entropy than sparse channels. The structuring of the model improves the performance of the estimation process.

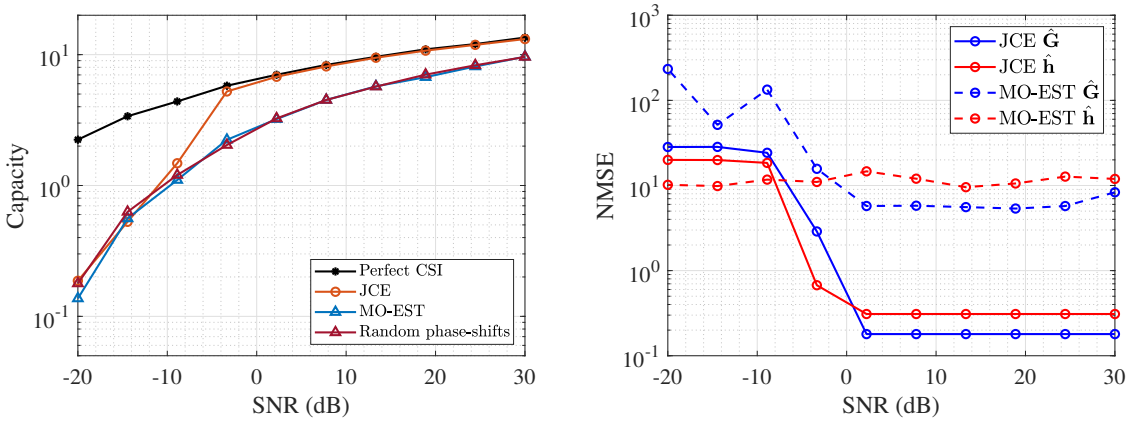
2.5.4 Performance of Joint Channel Estimation

We evaluate the performance of the proposed JCE method using Saleh-Valenzuela model [51] [52] that is modeled as follows:

$$\boldsymbol{G} = \sqrt{\frac{MN}{P}} \sum_{p=1}^{P} \alpha_p \boldsymbol{a}_{\mathrm{BS}}(\theta_p) \boldsymbol{a}_{\mathrm{RIS}}(\phi_p, \varphi_p); \tag{2.38}$$

$$\boldsymbol{h} = \sqrt{\frac{N}{Q}} \sum_{q=1}^{Q} \beta_q \boldsymbol{a}_{\mathrm{RIS}}(\phi_q, \varphi_q), \tag{2.39}$$

where α_p, θ_p, and ϕ_p/φ_p denote the complex gain, angle of arrivals (AOA), and azimuth/elevation of angle of departure (AOD) of the p-th path of RIS-BS channel. Sim-



(a) Achieved capacity (b) NMSE

Figure 2.5: Performance of JCE method with path-based model.

ilarly, β_q and ϕ_q/φ_q denote the complex gain and azimuth/elevation AOA of the q-th path of the UE-RIS channel, respectively. Besides, $\boldsymbol{a}_{\text{BS}}$ and $\boldsymbol{a}_{\text{RIS}}$ denote the receive and transmit array response vectors at the BS and the RIS, respectively. Then, the array response vector of the half-wavelength spaced uniform linear array at the BS is given by:

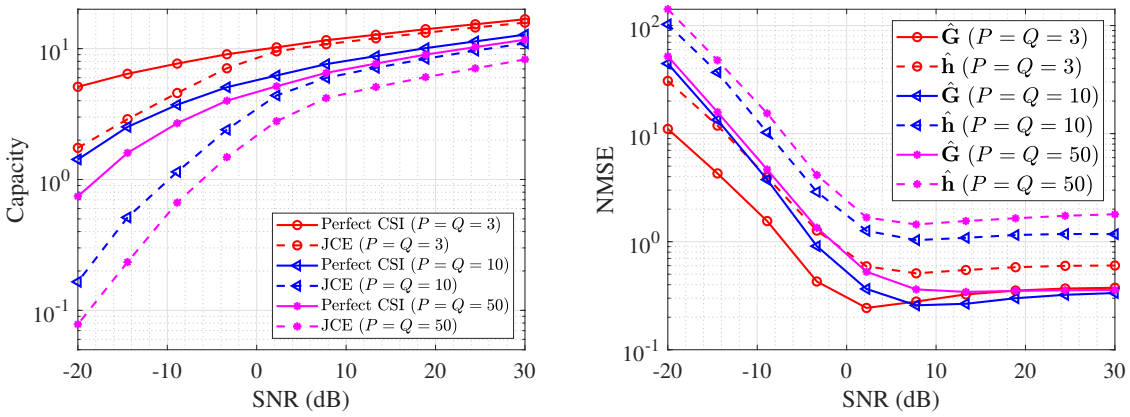
$$\boldsymbol{a}_{\text{BS}}(\theta_p) = \frac{1}{\sqrt{M}}\left[1, e^{j\pi\cos\theta_p}, \ldots, e^{j\pi(M-1)\cos\theta_p}\right]^T. \tag{2.40}$$

In addition, the array response vector of the planar array at the RIS involving N elements is given by:

$$\boldsymbol{a}_{\text{RIS}}(\phi,\varphi) = \frac{1}{\sqrt{N}}\begin{bmatrix} 1 \\ e^{j\pi\sin\phi\sin\varphi} \\ \vdots \\ e^{j\pi\sqrt{N}\sin\phi\sin\varphi} \end{bmatrix} \otimes \begin{bmatrix} 1 \\ e^{j\pi\cos\varphi} \\ \vdots \\ e^{j\pi\sqrt{N}\cos\varphi} \end{bmatrix}. \tag{2.41}$$

The angles $\theta_p, \phi_p, \varphi_p, \phi_q, \varphi_q$ are generated uniformly from $[0, 2\pi)$.

Fig. 2.5(a) illustrates the capacity as a function of the SNR ρ. The phase-shifts derived from the estimated channels are able to achieve a better capacity than the



(a) Achieved capacity

(b) NMSE

Figure 2.6: Performance of JCE method with different number of paths.

random selection of the RIS configuration which validates that the neural networks were able to effectively learn the channels. Moreover, our method outperforms the MO-EST method primarily due to the ability of neural networks to capture the sparse structure of the channels at high dimensions. In particular, the JCE method demonstrates a notable improvement with a gain of 3.70 dB at -3.33 dB SNR compared to the MO-EST method, and achieves a gain of 1.35 dB at 30 dB SNR.

Next, we investigate the estimation error of both channels EU-RIS and RIS-BS. As depicted in Fig. 2.5(b), the NMSE decreases with increasing SNR. Notably, our learning-based approach significantly outperforms the MO-EST baseline. In addition, the proposed method presents a lower computation time than the iterative algorithm MO-EST by leveraging the significantly lower inference time of the neural networks. Specifically and at 20 dB of SNR, the neural networks predict the auxiliary parameters within 0.20 seconds, whereas MO-EST requires 1.45 seconds to estimate the channels.

2.5.5 Effects of Sparsity on the Estimation

Furthermore, we evaluate the JCE method under different number of paths investigating the effect of the level of sparsity on the estimation performance. Specifically, Fig. 2.6(a) presents the capacity as a function of the SNR for three scenarios: $P = Q = 3$, $P = Q = 10$, and $P = Q = 50$. The numerical results reveal that, under high SNR, the capacity achieved based on phase-shifts derived from the estimated channels converges towards the exact capacity obtained when employing phase-shifts derived from the perfect CSI. Furthermore, we observe a notable impact of channel sparsity on the estimation performance in terms of capacity. Specifically, as the channel sparsity increases, signifying a reduced number of propagation paths, the achieved capacity becomes increasingly closer to the exact capacity due to the improvement of estimation of the channels. This behavior can be attributed to the sparsity-inducing nature of the variational loss function employed by the encoders, which leverages a Laplace prior to enforcing a sparse structure over the channels. Consequently, the proposed JCE method demonstrates superior performance for scenarios involving more sparse channels compared to those with less sparsity.

Fig. 2.6(b) depicts the evaluation of the NMSE to assess the performance of the proposed method. Notably, as the SNR increases, a clear trend emerges where the NMSE consistently decreases. Additionally, the degree of sparsity in the channel in the angular domain $\boldsymbol{h}^{\mathsf{vir}}$ plays a critical role. More specifically, the NMSE exhibits a significant degradation when the number of paths increases, with the most substantial performance deterioration occurring when $P = Q = 50$ paths are considered. This degradation can be attributed to the fact that 50 paths approach the dimensionality of the channel vector $\boldsymbol{h} \in \mathbb{C}^{64}$. Conversely, for the RIS-BS channel $\boldsymbol{G}^{\mathsf{vir}}$, the NMSE experiences a minor degradation as the number of paths varies. This behavior stems from the larger dimensionality of the RIS-BS channel matrix, $M \times N = 256$, in relation

to the maximum number of paths, mitigating the impact of variations in the number of paths. Importantly, these findings highlight the superior efficiency of the proposed method in scenarios characterized by higher levels of sparsity, effectively bypassing the need for a priori knowledge of the specific number of paths.

2.6 Summary

We propose a VI-based channel estimation method in fully passive RIS-aided mmWave single-user SIMO communication systems. The channel estimation problem has been formulated as a bayesian inference problem where we rely on the received training signal and the knowledge on the prior of the channels to obtain the posterior of the channels. Using a VI framework, we approximate the intractable posterior of the channels with convenient distributions. The parameters of the approximated distributions are generated by neural networks *Encoder* $\mathcal{F}$ and *Encoder* $\mathcal{G}$ trained using variational loss functions derived using a lower bound, i.e. ELBO, on the log-likelihood of the received signal. Then, the learned distributions, which are close to the true posterior distributions in terms of Kullback Leibler divergence, are leveraged to obtain the MAP estimation of the UE-RIS and RIS-BS channels. We derive the ELBO under two channel priors. The first channel prior models Rayleigh fading channels with Gaussian prior, whereas the second one represents sparse channels in the angular domain with Laplace prior. The simulation results demonstrate that MAP channel estimates using the approximated posteriors yield a capacity which is close to the one achieved with the true posteriors, thus demonstrating the effectiveness of the proposed method. In addition, we observe an improvement of the channels estimation with Laplace prior compared to the Gaussian prior. Moreover, it has been shown that the performance of the proposed approach degrades with the increase of number of paths compared to the channels size due to the Laplace prior that enforces the sparsity of the channels in the

angular domain.

Chapter 3

Joint Channel-Covariance Estimation for RIS-Assisted System

The estimation of the S-CSI in RIS-assisted networks offers a broader perspective on signal propagation dynamics compared to focusing solely on I-CSI. User mobility introduces a dynamic and evolving channel environment for the UE-RIS link making the I-CSI susceptible to inaccuracies due to its limited snapshot view. I-CSI necessitates frequent updates and high training overhead to maintain accuracy, which can consume valuable resources and impede overall efficiency. In contrast, statistical CSI estimation offers a comprehensive view of the channel's behavior over multiple time instances, reducing the frequency of updates and mitigating the training overhead burden. Furthermore, the signaling complexity of adjusting phase shifts for I-CSI can become prohibitive. By focusing on statistical CSI, we alleviate this complexity by allowing the RIS to adapt more smoothly and efficiently to overarching channel trends, minimizing the need for rapid and intricate phase adjustments. Therefore, in this chapter, we focus on estimating the long-term CSI that remain constant for extended periods. Specifically, we aim to estimate the RIS-BS channel which remains static due to the fixed positions

of the RIS and BS, and the UE-RIS CCM using the VI-based estimation framework. The major contributions embodied by this chapter can be summerized as follows:

- To the best our knowledge, this is the first work that studies the joint estimation of the RIS-BS channel and the UE-RIS CCM in a single user RIS-assisted networks. We adopt the VI-based to separately estimate the RIS-BS channel and the UE-RIS covariance leveraging the sparse structure of the mmWave channels and the mmWave channel covariance matrices considering the low mobility of the user where the AOAs of the UE-RIS channel are deemed to be in a subspace.
- We develop a closed-form expression of the phase-shifts for the passive beamforming based on the RIS-BS channel and the UE-RIS CCM.
- We show the effectiveness of the proposed methods by benchmarking the capacity with the capacity using the different ground truths of the I-CSI and the covariance matrix.

The rest of the chapter is organized as follows: The system model is described in Section 3.1. Section 3.2 presents the uplink training scheme adopted to obtain the training signal. In Section 3.3, we solve the joint channel-covariance estimation problem by deriving the variational losses that will be feeded to the neural networks as loss. Further, Section 3.4 discusses the phase-shifts optimization based on the RIS-BS channel and the UE-RIS CCM. Lastly, numerical results are presented in Section 3.5 before drawing out some concluding remarks in Section 3.5

3.1 System Model, Assumptions, and Problem Formulation

Consider a BS equipped with M antennas elements serving a single antenna user in the uplink. The BS is assisted by an RIS which has a N reflective elements. The UE-RIS and RIS-BS channel links are denoted by $\boldsymbol{h} \in \mathbb{C}^N$ and $\boldsymbol{G} \in \mathbb{C}^{M \times N}$, respectively. We omit considering the direct link between the user and the BS since it can be estimated using conventional SIMO channel estimation methods by deactivating the RIS. We adopt the block-fading channel model with coherence time $T_{\boldsymbol{h}}$ for the UE-RIS channel and $T_{\boldsymbol{G}}$ for the RIS-BS channel. Due to the fixed positions of the RIS and the BS, we assume $T_{\boldsymbol{G}} >> T_{\boldsymbol{h}}$. The channels are modeled using the channel model described in 2.5.4.

3.2 Uplink training

The RIS-BS channel may be considered quasi-static since the physical locations of the RIS and the BS do not change over time. The quasi-static nature of the RIS-BS channel can be advantageous for the design and optimization of a RIS-aided wireless communication system. We propose a transmission protocol to effectively estimate the RIS-BS channel and the covariance matrix of the UE-RIS channel, as shown in Fig. 3.1. The considered time interval consists of $T_{\boldsymbol{G}}$ time slots and can be divided into a data communication phase and a training phase. The RIS-BS channel link $\boldsymbol{G}$ and the CCM of UE-RIS channel denoted as $\boldsymbol{R}_{\boldsymbol{h}}$ are assumed to be constant during this time. The training time can be divided into N_b coherence blocks of the UE-RIS channel link where for each coherence block of channel $\boldsymbol{h}$ we use $N_p < T_{\boldsymbol{h}}$ time slots to send the training symbols. The remaining time slots can be used for data transmissions where passive beamforming with imperfect CSI methods can be used [53]. Therefore, we get $N_p \times N_b$ time slots used to send the training symbols.

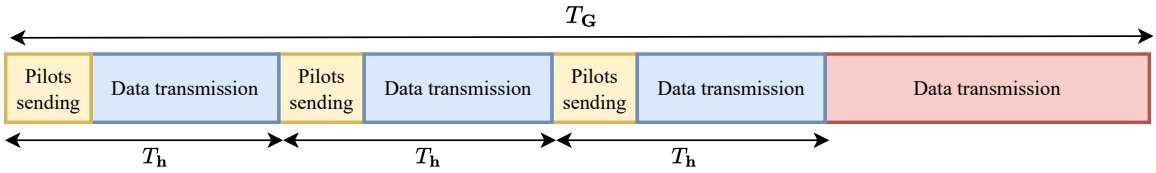


Figure 3.1: Transmission protocol.

In each coherence block, by sending N_p pilot signals while altering the configuration for each pilot, the received signal at the BS can be expressed as:

$$\boldsymbol{Y} = \sqrt{\rho}\, \boldsymbol{G}\, \mathrm{diag}(\boldsymbol{h})\boldsymbol{\Phi} + \boldsymbol{W}, \tag{3.1}$$

where $\boldsymbol{\Phi} = [\boldsymbol{v}_1, \ldots, \boldsymbol{v}_{N_p}] \in \mathbb{C}^{N \times N_p}$ is the RIS configuration used for training, $\boldsymbol{W} = [\boldsymbol{w}_1, \ldots, \boldsymbol{w}_{N_p}]$ is the noise matrix where $\boldsymbol{w}_l \sim \mathcal{CN}(\boldsymbol{0}, \boldsymbol{I}_N)$. The vectorized form of $\boldsymbol{Y}$ can be expressed as follows:

$$\tilde{\boldsymbol{y}} = \mathrm{vec}(\boldsymbol{Y}) \tag{3.2}$$

$$= \sqrt{\rho}(\boldsymbol{\Phi}^T \odot \boldsymbol{G})\boldsymbol{h} + \boldsymbol{w}, \tag{3.3}$$

where $\boldsymbol{w} = \mathrm{vec}(\boldsymbol{W}) \sim \mathcal{CN}(\boldsymbol{0}, \boldsymbol{I}_{MN_p})$. We define the combined training received signal as $\tilde{\boldsymbol{Y}} = [\tilde{\boldsymbol{y}}_1, \ldots, \tilde{\boldsymbol{y}}_{N_b}]$. The covariance matrix for the received training signal $\tilde{\boldsymbol{y}}$, given that the RIS-BS channel remains quasi-static, is expressed as:

$$\boldsymbol{R}_{\tilde{\boldsymbol{y}}} = \mathbb{E}[\tilde{\boldsymbol{y}}\tilde{\boldsymbol{y}}^H] = \rho(\boldsymbol{\Phi}^T \odot \boldsymbol{G})\boldsymbol{R}_{\boldsymbol{h}}(\boldsymbol{\Phi}^T \odot \boldsymbol{G})^H + \boldsymbol{I}_{MN_p}. \tag{3.4}$$

In various scenarios, the UE-RIS channel is highly correlated because of the small set of AOAs contributing to the propagation within a small region in the angular domain [37]. Therefore, the covariance matrix $\boldsymbol{R}_{\boldsymbol{h}} = \mathbb{E}[\boldsymbol{h}\boldsymbol{h}^H]$ is considered as a low-rank matrix. Formally, we express the covariance matrix as follows:

$$\boldsymbol{R}_{\boldsymbol{h}} = \boldsymbol{F}_N^H \boldsymbol{D} \boldsymbol{F}_N, \tag{3.5}$$

where $\boldsymbol{D} = \text{diag}(\boldsymbol{d})$ is a diagonal matrix with a sparse main diagonal denoted as $\boldsymbol{d}$ representing the angular correlation. We focus on estimating the sparse vector $\boldsymbol{d}$ and RIS-BS channel in the angular domain denoted as $\boldsymbol{G}^{\mathsf{vir}} = \frac{1}{MN}\boldsymbol{F}_M^H \boldsymbol{G} \boldsymbol{F}_N^H$ rather than estimating the full covariance matrix $\boldsymbol{R}_{\boldsymbol{h}}$ which typically is a large matrix of size $N \times N$.

3.3 Variational Inference-Based Joint Channel-Covariance Estimation

We aim to jointly estimate the RIS-BS channel link $\boldsymbol{G}$, and the covariance matrix of the UE-RIS channel $\boldsymbol{h}$ which is same as estimating the vector $\boldsymbol{d}$, based on the received training signal.

As discussed in Section 2.3.2, the channel between the RIS and the BS, denoted as $\boldsymbol{G}$, exhibits sparsity in the angular domain. To represent this sparsity, a complex Laplace distribution is employed to model the sparse matrix. Additionally, the vector $\boldsymbol{d}$, which represents a sparse positive real-valued vector, is modeled using a complex Exponential distribution:

$$\boldsymbol{G}_{i,j}^{\mathsf{vir}} \sim \mathcal{CL}(0, \alpha_{\boldsymbol{G}^{\mathsf{vir}}}); \tag{3.6}$$

$$\boldsymbol{d}_i \sim \text{Exp}(\alpha_{\boldsymbol{d}}). \tag{3.7}$$

Applying the VI framework, we approximate the intractable true posterior distribution $p(\boldsymbol{G}^{\mathsf{vir}}, \boldsymbol{d}|\tilde{\boldsymbol{Y}})$ by two separate tractable parameterized distributions denoted by $q_{\boldsymbol{\lambda}_1}(\boldsymbol{d}|\tilde{\boldsymbol{Y}})$ and $q_{\boldsymbol{\lambda}_2}(\boldsymbol{G}^{\mathsf{vir}}|\tilde{\boldsymbol{Y}})$ using the mean-field approximation. Moreover, the parameters of the chosen auxiliary distributions are returned by *Encoder* $\mathcal{F}$ and *Encoder* $\mathcal{G}$. The training signal $\tilde{\boldsymbol{Y}}$ is preprocessed such that the input to the neural networks is defined by $\tilde{\boldsymbol{Y}}\tilde{\boldsymbol{Y}}^H/N_b - \boldsymbol{I}_{MN_p}$.

The auxiliary distribution for the RIS-BS channel link in the angular domain $\boldsymbol{G}^{\mathsf{vir}}$ is assumed to follow the complex Laplace distribution with independent elements, and

the elements of $\boldsymbol{d}$ follow a Gamma distribution with unit scale:

$$q_{\boldsymbol{\lambda}_1}(\boldsymbol{d}_i|\tilde{\boldsymbol{Y}}) \sim \text{Gamma}(\boldsymbol{k}_i); \tag{3.8}$$

$$q_{\boldsymbol{\lambda}_2}(\boldsymbol{G}^{\mathsf{vir}}_{i,j}|\tilde{\boldsymbol{Y}}) \sim \mathcal{CL}(\boldsymbol{M}_{i,j}, \boldsymbol{B}_{i,j}), \tag{3.9}$$

where $\boldsymbol{\lambda_1} = \{\boldsymbol{k}\}$ and $\boldsymbol{\lambda_2} = \{\boldsymbol{M}, \boldsymbol{B}\}$ are the parameters of the auxiliary distributions which are obtained by maximizing the ELBO function, which is given in general form in Eq. 2.3, expressed as follows:

$$\begin{aligned}\mathcal{L}^{\mathsf{S-CSI}}(\boldsymbol{\lambda}_1, \boldsymbol{\lambda}_2) = &\underbrace{\mathbb{E}_{\boldsymbol{d}\sim q_{\boldsymbol{\lambda}_1}(\boldsymbol{d}|\tilde{\boldsymbol{Y}})}\left[\log \frac{q_{\boldsymbol{\lambda}_1}(\boldsymbol{d}|\tilde{\boldsymbol{Y}})}{p(\boldsymbol{d})}\right]}_{\mathcal{L}_1^{\mathsf{S-CSI}}} + \underbrace{\mathbb{E}_{\boldsymbol{G}^{\mathsf{vir}}\sim q_{\boldsymbol{\lambda}_2}(\boldsymbol{G}^{\mathsf{vir}}|\tilde{\boldsymbol{Y}})}\left[\log \frac{q_{\boldsymbol{\lambda}_1}(\boldsymbol{G}^{\mathsf{vir}}|\tilde{\boldsymbol{Y}})}{p(\boldsymbol{G}^{\mathsf{vir}})}\right]}_{\mathcal{L}_2^{\mathsf{S-CSI}}} \\ &\underbrace{-\mathbb{E}_{\boldsymbol{d},\boldsymbol{G}^{\mathsf{vir}}\sim q_{\boldsymbol{\lambda}}(\boldsymbol{d},\boldsymbol{G}^{\mathsf{vir}}|\tilde{\boldsymbol{Y}})}\left[\log p(\tilde{\boldsymbol{Y}}|\boldsymbol{d}, \boldsymbol{G}^{\mathsf{vir}})\right]}_{\mathcal{L}_3^{\mathsf{S-CSI}}}.\end{aligned} \tag{3.10}$$

Since the prior and auxiliary posterior of $\boldsymbol{G}^{\mathsf{vir}}$ align with the case addressed in channel estimations, the second loss, expressed as $\mathcal{L}_2^{\mathsf{S-CSI}} = \mathcal{L}_2^{\mathsf{I-CSI}}$, remains unchanged. However, the first loss, which involves the KL-divergence between an Exponential distribution and the Gamma distribution, can be expressed as follows:

$$\begin{aligned}\mathcal{L}_1^{\mathsf{S-CSI}}(\boldsymbol{\lambda}_1) &= \mathbb{E}_{\boldsymbol{d}\sim q_{\boldsymbol{\lambda}_1}(\boldsymbol{d}|\tilde{\boldsymbol{Y}})}\left[\log \frac{q_{\boldsymbol{\lambda}_1}(\boldsymbol{d}|\tilde{\boldsymbol{Y}})}{p(\boldsymbol{d})}\right] \\ &= \sum_{i=1}^{N} \mathbb{E}_{\boldsymbol{d}_i\sim q_{\boldsymbol{\lambda}_1}(\boldsymbol{d}_i|\tilde{\boldsymbol{Y}})}\left[\log q(\boldsymbol{d}_i|\tilde{\boldsymbol{Y}}) - \log p(\boldsymbol{d}_i)\right] \\ &= \sum_{i=1}^{N}(1-\boldsymbol{k}_i)\psi(1) - \log\Gamma(1.0) + \log\Gamma(\boldsymbol{k}_i),\end{aligned} \tag{3.11}$$

where $\Gamma(x)$ is the gamma function and $\psi(x)$ is the digamma function. The third loss, denoted as $\mathcal{L}_3^{\mathsf{S-CSI}}$, is defined as the log-likelihood of the received training signal and can be expressed as follows:

$$\begin{aligned}\mathcal{L}_3^{\mathsf{S-CSI}}(\boldsymbol{\lambda}_1, \boldsymbol{\lambda}_2) = \mathbb{E}_{\boldsymbol{d},\boldsymbol{G}^{\mathsf{vir}}\sim q_{\boldsymbol{\lambda}}(\boldsymbol{d},\boldsymbol{G}^{\mathsf{vir}}|\tilde{\boldsymbol{Y}})}&\left[\text{Tr}\left(\tilde{\boldsymbol{Y}}^H \boldsymbol{R}_{\tilde{\boldsymbol{Y}}}^{-1}\tilde{\boldsymbol{Y}}\right)\right. \\ &\left. + \log|\boldsymbol{R}_{\tilde{\boldsymbol{Y}}}|\right] + C_2,\end{aligned} \tag{3.12}$$

where C_2 is a constant. To compute the gradient with respect to the parameters of the auxiliary distribution of the RIS-BS channel link, $q_{\boldsymbol{\lambda}_2}(\boldsymbol{G}^{\mathsf{vir}}|\tilde{\boldsymbol{Y}})$, we employ the reparameterization trick. This technique involves generating Monte-Carlo samples where each sample is computed by $\widehat{\boldsymbol{G}^{\mathsf{vir}}}_{i,j} = \boldsymbol{M}_{i,j} + \boldsymbol{B}_{i,j} \times \mathcal{CL}(0,1)$ to maintain the differentiability and enabling efficient optimization through gradient-based methods. To address the complexity of directly applying the standard reparameterization trick to the Gamma distribution, an alternative technique known as the implicit reparameterization technique, as outlined by [54], is employed. This technique facilitates the generation of Monte-Carlo samples that remain differentiable with respect to the shape parameter vector $\boldsymbol{k}$.

After learning neural networks, denoted as *Encoder* $\mathcal{F}$ and *Encoder* $\mathcal{G}$, that predicts the distribution parameters of $q_{\boldsymbol{\lambda}_1}(\boldsymbol{d}|\tilde{\boldsymbol{Y}})$ and $q_{\boldsymbol{\lambda}_2}(\boldsymbol{G}^{\mathsf{vir}}|\tilde{\boldsymbol{Y}})$, respectively, the channels are estimated using the MAP method applied on the auxiliary distributions:

$$\widehat{\boldsymbol{d}} = \arg\max_{\boldsymbol{d}} \; q_{\boldsymbol{\lambda}_1}(\boldsymbol{d}|\tilde{\boldsymbol{Y}}) = \boldsymbol{k} - \boldsymbol{1}; \tag{3.13}$$

$$\widehat{\boldsymbol{G}^{\mathsf{vir}}} = \arg\max_{\boldsymbol{G}^{\mathsf{vir}}} \; q_{\boldsymbol{\lambda}_2}(\boldsymbol{G}^{\mathsf{vir}}|\tilde{\boldsymbol{Y}}) = \boldsymbol{M}. \tag{3.14}$$

3.4 Optimization of the phase-shifts based on RIS-BS I-CSI and UE-RIS S-CSI

In this section, we propose a closed-form expression of the phase-shifts that maximize the achievable rate of the UE-RIS-BS channel link based on the RIS-BS channel and the UE-RIS CCM. The problem is formulated as follows:

$$\begin{aligned} &\max_{\{\theta_i\}} \; \mathbb{E}_{\boldsymbol{h}}\Big[\log_2\left(1+\rho\,||\boldsymbol{G}\,\mathrm{diag}(\boldsymbol{v})\,\boldsymbol{h}||^2\right)\Big], \\ &\text{Subject to: } \boldsymbol{v}_i = e^{\theta_i}. \end{aligned} \tag{3.15}$$

The problem in Eq. 3.15 is challenging to solve due to the lack of an explicit expression for the expectation over the logarithm. To address this difficulty, we adopt a strategy of maximizing a reliable upper bound on this expression [30]:

$$\mathbb{E}_{\boldsymbol{h}}\left[\log_2\left(1+\rho||\boldsymbol{G}\text{diag}(\boldsymbol{v})\boldsymbol{h}||^2\right)\right] \leq \log_2\left(1+\rho\mathbb{E}_{\boldsymbol{h}}\left[||\boldsymbol{G}\text{diag}(\boldsymbol{v})\boldsymbol{h}||^2\right]\right). \tag{3.16}$$

It is important to acknowledge that the upper bound in Eq. 3.16 is highly accurate and serves as a reliable approximation of the original objective function, particularly for large values of ρ [30]. By maximizing this upper bound, we are able to formulate the subsequent optimization problem as follows:

$$\begin{aligned} &\max_{\{\theta_i\}} \ \mathbb{E}_{\boldsymbol{h}}\Big[||\boldsymbol{G}\ \text{diag}(\boldsymbol{v})\ \boldsymbol{h}||^2\Big], \\ &\text{Subject to: } \boldsymbol{v}_i = e^{\theta_i}. \end{aligned} \tag{3.17}$$

The objective can be further expressed as follows:

$$\mathbb{E}_{\boldsymbol{h}}\Big[||\boldsymbol{G}\text{diag}(\boldsymbol{v})\boldsymbol{h}||^2\Big] = \text{Tr}\Big(\boldsymbol{G}\text{diag}(\boldsymbol{v})\boldsymbol{R}_{\boldsymbol{h}}\text{diag}(\boldsymbol{v})^H\boldsymbol{G}^H\Big). \tag{3.18}$$

Given the SVD of $\boldsymbol{G} = \boldsymbol{U}\boldsymbol{S}\boldsymbol{V}^H$ and the eigenvalue decomposition of the covariance matrix $\boldsymbol{R}_{\boldsymbol{h}} = \boldsymbol{P}\boldsymbol{\Sigma}\boldsymbol{P}^H$, the objective function can be expressed as follows:

$$\mathbb{E}_{\boldsymbol{h}}\Big[||\boldsymbol{G}\text{diag}(\boldsymbol{v})\boldsymbol{h}||^2\Big] = \sum_{i=1}^{M}\sum_{j=1}^{N}\left|\boldsymbol{S}_{i,i}\boldsymbol{\Sigma}_{j,j}^{1/2}\sum_{k=1}^{N}\boldsymbol{V}_{k,i}^{*}\boldsymbol{P}_{k,j}e^{i\theta_k}\right|^2. \tag{3.19}$$

Therefore, we take the phases that align with the phases of the largest eigenvector of $\boldsymbol{G}$ and $\boldsymbol{R}_{\boldsymbol{h}}$, referred to as $\boldsymbol{\vartheta}^{\max}$ and $\boldsymbol{p}^{\max}$, respectively, to maximize the objective function and satisfy the constraints, which are given by:

$$\theta_k^* = -(\angle\boldsymbol{p}_k^{\max} - \angle\boldsymbol{\vartheta}_k^{\max}). \tag{3.20}$$

3.5 Simulations and results

We examine an RIS configuration comprising $N = 64$ passive elements, alongside a BS equipped with $M = 4$ antennas. In order to estimate the RIS-BS channel and the UE-

RIS CCM, we transmit $N_p = 4$ pilot symbols per UE-RIS coherence block considering $N_b = 200$ coherence blocks during an uplink SIMO setup. The resulting training signals are preprocessed as discussed in Section 3.3 and fed to the encoders *Encoder* $\mathcal{F}$ and *Encoder* $\mathcal{G}$. The encoders are modeled by fully connected neural networks with 2 hidden layers with 300 units utilizing Relu activation. The output layer of the *Encoder* $\mathcal{F}$ for auxiliary distribution $q_{\boldsymbol{\lambda}_1}(\boldsymbol{d}|\tilde{\boldsymbol{Y}})$ is composed of N unit with Sigmoid activation function. In addition, the output layer of the *Encoder* $\mathcal{G}$ for the auxiliary distribution $q_{\boldsymbol{\lambda}_2}(\boldsymbol{G}^{\mathsf{vir}}|\tilde{\boldsymbol{Y}})$ is composed of two heads: one head outputs the mean with $2 * N * M$ units with Tanh activation, and a second head outputs the scales with $N * M$ units with Softmax activation. For training these neural networks, the Adam optimizer [50] is employed with an initial learning rate of 0.001. The training process involves maximizing the ELBO function using an unlabeled dataset consisting of 10^4 samples. Expectations within the ELBO functions are computed based on 1000 samples. The performance assessment of the methods is conducted through 50 Monte-Carlo samples. Furthermore, the channels are generated according to the path-based channel model described in Section 2.5.4. The AOAs of the RIS in the UE-RIS channel ϕ_q and φ_q are generated uniformly from different clusters obtained by dividing the interval $[0, 2\pi)$ into 100 sub-intervals. This clustering results in a covariance matrix that exhibits sparsity in the angular domain. In addition to the baseline mentioned in Section 2.5.1, i.e perfect CSI, random phase-shifts and MO-EST method [49], we benchmark the proposed method with the capacity obtained where the phase-shifts are computed based on the true RIS-BS channel $\boldsymbol{G}$ and the true UE-RIS covariance matrix $\boldsymbol{R_h}$, referred to as **Perfect channel and perfect covariance (PC-PCov)**.

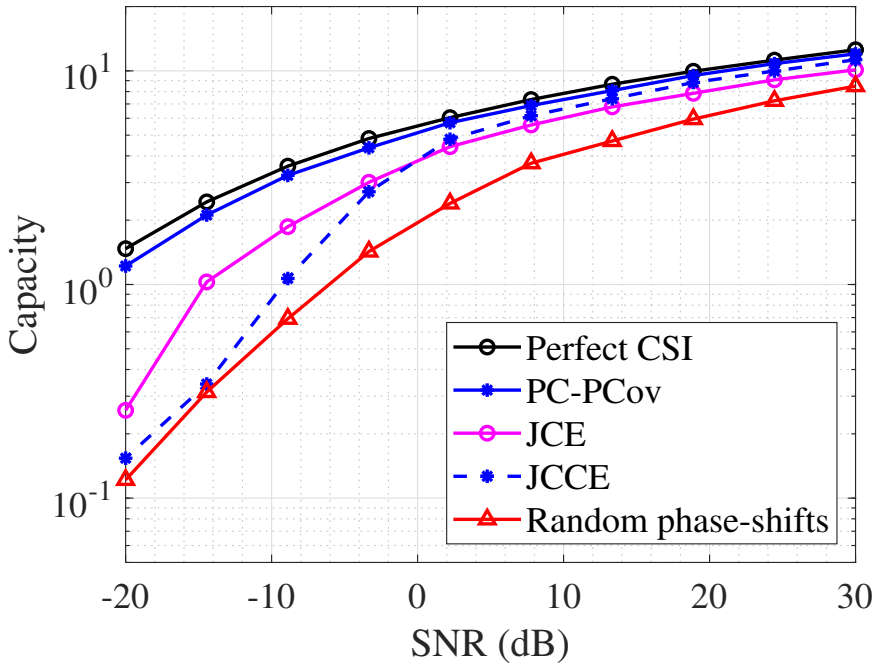


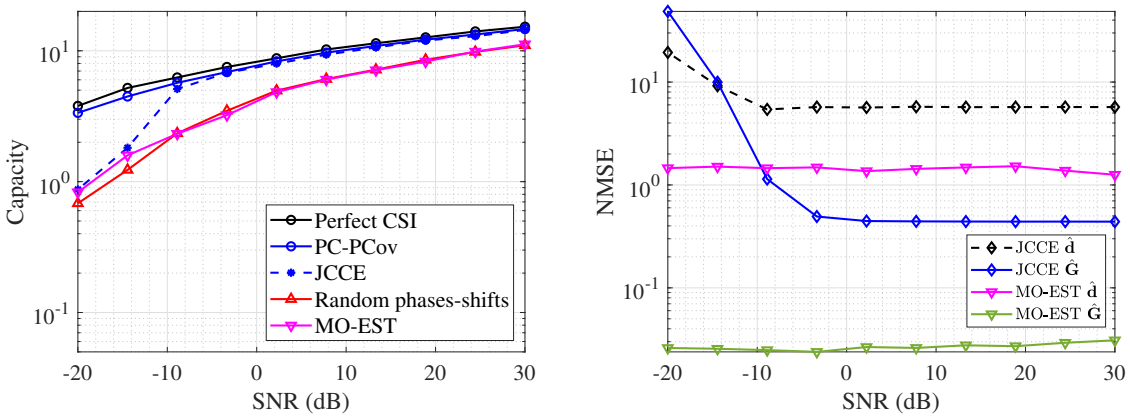
Figure 3.2: Performance of the proposed methods.

3.5.1 Comparison Between the Proposed Methods

To compare the JCE and JCCE methods, we evaluate the capacity taking into account the number of pilots used to get the training signals that is expressed as $C_p = (1 - \alpha)\log_2(1 + \rho||\boldsymbol{G}\text{diag}(\boldsymbol{v})\boldsymbol{h}||^2)$ where $\alpha = N_{\text{pilots used}}/N_{\text{Total transmissions}}$. The number of paths are set to $P = Q = 3$. In Fig. 3.2 shows that the JCE method exhibits superior performance over the JCCE approach at low SNR, while the JCCE method outperforms the JCE at high SNR when the estimates closely approach the PC-PCov. The observed performance improvement can be attributed to the inherent differences in their channel estimation approaches. With the JCE, the channels $\boldsymbol{G}$ and $\boldsymbol{h}$ need to be estimated at each coherence block of $\boldsymbol{h}$, which is relatively short compared to the quasi-static nature of channel $\boldsymbol{G}$ and the covariance $\boldsymbol{R_h}$, and thus it leads to higher values of α. In contrast, the JCCE method utilizes the estimates of the RIS-BS channel and the UE-

RIS covariance matrix, enabling the use of phase-shifts without the need to estimate the UE-RIS in subsequent coherence blocks. This leads to reduced training overhead and signaling complexity for configuring the RIS, resulting in lower values of α, which validates the efficiency of leveraging the estimates of the RIS-BS channel and the UE-RIS covariance matrix for obtaining the phase-shifts.

3.5.2 Performance of Joint Channel-Covariance Estimation



(a) Achieved capacity

(b) NMSE

Figure 3.3: Performance of the VI-based estimation of RIS-BS channel and UE-RIS channel covariance.

To evaluate the JCCE method, we compare it against the MO-EST estimation approach, where the channels are estimated at each coherence block and used to estimate the covariance matrix $\boldsymbol{R}_{\boldsymbol{h}}$. We set $P = 3$ and $Q = 1$ to represent the number of paths for the RIS-BS and UE-RIS channels, respectively.

Fig. 3.3(a) shows a degradation in performance by substituting the UE-RIS CCM (PC-PCov) compared to using the UE-RIS channel itself (Perfect CSI). However, by updating the RIS phase-shifts based on the UE-RIS CCM, we reduce the signaling

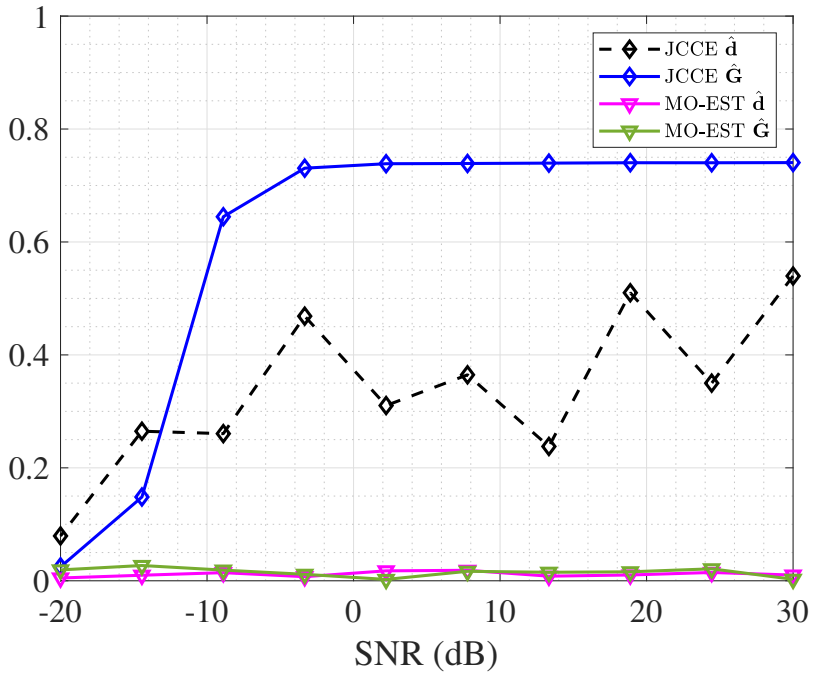


Figure 3.4: Inner Product of the estimated largest eigenvectors with ground truth.

overhead associated with the RIS configuration. This approach enables the RIS configuration to remain fixed for an extended period while ensuring an acceptable rate performance since the UE-RIS CCM and the RIS-BS channel are considered quasi-static for the subsequent coherence blocks of the UE-RIS channel. Moreover, the capacity using the phase-shifts derived from the estimated channel and CCM via JCCE gets closer with the increase of the SNR to the exact capacity which validates the proposed method. Furthermore, the proposed method demonstrates superior performance compared to the MO-EST method, which fails to capture the sparse structure of the channel and its covariance. Fig. 3.3(b) showcases the NMSE evaluation across different SNR values. Notably, the MO-EST method reaches lower values of NMSE compared to the proposed method for the RIS-BS channel $\boldsymbol{G}$ and the angular spectrum $\boldsymbol{d}$. For a further investigation, in Fig. 3.4, we evaluate the absolute value of the complex inner prod-

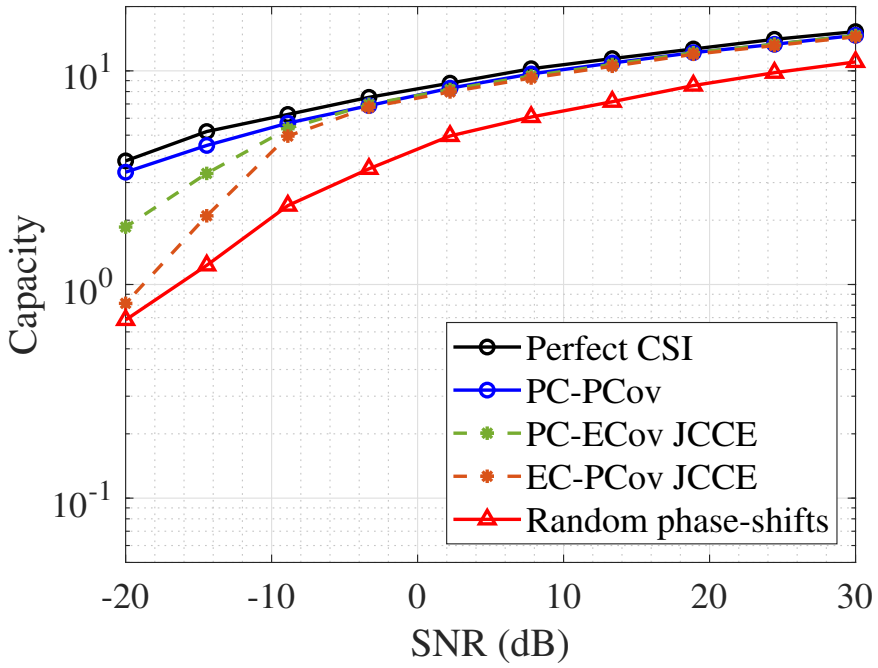


Figure 3.5: Performance of the estimates separated.

uct of the largest eigenvectors of the estimated RIS-BS channels $\widehat{\boldsymbol{G}}$ and the estimated CCM $\widehat{\boldsymbol{R}}_{\boldsymbol{h}}$, expressed as $\langle \widehat{\boldsymbol{\vartheta}^{\max}}, \boldsymbol{\vartheta}^{\max} \rangle = \widehat{\boldsymbol{\vartheta}^{\max}}^{H} \boldsymbol{\vartheta}^{\max}$, with the largest eigenvectors from the PC-PCov, as well as for the largest singular vectors for UE-RIS CCM expressed by $\langle \widehat{\boldsymbol{p}^{\max}}, \boldsymbol{p}^{\max} \rangle = \widehat{\boldsymbol{p}^{\max}}^{H} \boldsymbol{p}^{\max}$. We observe that the proposed method is able to effectively estimate the largest eigenvectors of the RIS-BS channel and the UE-RIS CCM, as the inner product gets closer to 1, with the increase of the SNR, which can be interpreted as an alignment of the estimated largest eigenvector to the largest eigenvector of the actual channel and CCM.

3.5.3 Effectiveness of Separate Channel Estimates

We examine the performance of each estimate aside from the baselines of the capacity with phase-shifts derived from the exact channels and the phase-shifts derived from the

exact RIS-BS channel and UE-RIS CCM. Fig. 3.5 shows the effectiveness of both neural networks *Encoder* $\mathcal{F}$ and *Encoder* $\mathcal{G}$ to estimate approximate posterior distributions to achieve desirable performance. We observe that at high SNR, both estimates, which are referred to as Perfect Channel - Estimated Covariance (**PC-ECov**) and Estimated Channel - Perfect Covariance (**EC-PCov**), can separately achieve the capacity with perfect RIS-BS channel and UE-RIS CCM. However, at low SNR, the CCM estimates surpass the capacity achieved using channel estimates. This superiority is attributed to the highly sparse structure present in $\boldsymbol{d}$, originating from the clusters of AoAs and AoDs contributing to the UE-RIS channel $\boldsymbol{h}$. Moreover, the vector $\boldsymbol{d}$ has a lower dimension of $\boldsymbol{d} \in \mathbb{C}^N$ compared to the RIS-BS channel $\boldsymbol{G} \in \mathbb{C}^{M \times N}$, which facilitates the estimation of the non-sparse values.

3.6 Summary

We have presented a novel approach of long-term CSI estimation in a fully passive single user SIMO RIS-assisted communication system by tackling the problem of estimating the RIS-BS channel and the UE-RIS CCM. We employed the VI-based estimation framework to infer over the channel and covariance matrices by approximating the intractable posteriors with a convenient distribution through maximizing the ELBO. Thanks to the low-rank structure of the CCM, our approach estimates a diagonal sparse matrix that models the spatial correlation rather than estimating a full matrix. The parameters of the auxiliary are obtained from a trained encoders given the preprocessed training signal as input. Furthermore, we propose a solution to the passive beamforming based on the long-term CSI considered. We have showcased that the joint channel-covariance estimation method reduces the training overhead compare to the joint channel estimation method since the former does not require estimating the UE-RIS I-CSI that has a short coherence time. The adopted approach leads to a sustainable and robust configuration

of the RIS by considering long-term CSIs that remain static for extended period, rather than relying on I-CSI that changes rapidly.

Chapter 4

Conclusion

4.1 Concluding Remarks

Precise channel estimation forms the cornerstone of effective phase shift optimization within Reconfigurable Intelligent Surfaces (RIS), directly impacting RIS performance and thereby contributing to the advancement of next-generation wireless communication systems. We considered a single user single-input multiple-ouput (SIMO) RIS-assisted communication network, and provided a framework for the CSI estimation for the I-CSI and S-CSI of the RIS related channels. In Chapter 2, we propose a Variational Inference (VI)-based framework to separately estimate the user equipment (UE)-RIS and RIS-base station (BS) channels using the uplink training signal bypassing the requirement of the foreknowledge of the number of paths among the channels by considering a Laplacian prior over the channels. Leveraging the power of neural network to capture the channels' structure, the auxiliary distribution are obtained from trained neural networks in unsupervised manner by maximizing the Evidence Lower-Bound (ELBO). Chapter 3 extended the VI-based framework to estimate the RIS-BS channel and the UE-RIS CCM taking into account the large coherence time of the RIS-BS channel and

the low-rank structure of the UE-RIS CCM due to the signal subspace of the user. From numerical results we conclude that *i)* the VI-based approach estimate effectively the UE-RIS channel/covariance matrix and the RIS-BS channel, *ii)* relying on the RIS-BS channel and the UE-RIS CCM for the RIS phase-shifts provide a near-optimal performance in terms of the capacity, and *iii)* optimizing the phase-shifts based on the long-term CSI, i.e. RIS-BS channel and UE-RIS CCM, rather than relying on the I-CSI reduces the training overhead and the signaling complexity of the RIS by maintaining the phase-shifts constant for the subsequent UE-RIS coherence blocks.

4.2 Future Directions

This work can be extended in a few directions to provide a comprehensive framework for the VI-based channel estimation for RIS-assisted wireless systems.

4.2.1 Physically Consistent Phase-Shift Models

The proposed VI-based CSI estimation methods assumes the approximative phase-shift model, however in real world the the elements of the RIS are mutually coupled which is neglected in the conventional model of the reflection. A physically-consistent RIS considers the coupling where the reflection matrix has a complex expression based on scattering matrices rather than a diagonal matrix [55, 56]. Hence, the proposed approaches can be extended to other reflection model since it is not limited to the system model introduced.

4.2.2 Multi-user Setting through clustering

The estimation of the UE-RIS covariance matrix relies on the signal subspace of the user where the directions of arrivals to the RIS are in a specific range of values which

can be referred to as the angular cluster. In a multi-user setting, the channel paths of the nearby users have a similar directions of arrivals, and thus the channels of different nearby users have similar covariance matrices. Therefore, the proposed joint channel-covariance estimation method can be extended to the multi-user setting where nearby users are served based on the same phase-shifts considering they share the RIS-BS channel and the UE-RIS CCM.

4.2.3 Variational User Tracking

A possible direction to use the VI framework is tracking the user mobility where the user mobility is modeled by a probability transition. In practical scenarios, the angles of arrivals change stochastically through time, hence a VI-based approach can be developed to predict the next angles of arrivals at the user given the received signal.

Appendix A

In this section, we give a detailed derivation of the losses under the distributions investigated.

To compute $\mathcal{L}_3^{\mathsf{I-CSI-Rayleigh}}$ (Eq. 2.16) for the Gaussian channels, we use the property $\mathbb{E}_{\boldsymbol{x}}[(\boldsymbol{a}-\boldsymbol{x})^H(\boldsymbol{a}-\boldsymbol{x})] = \mathrm{Tr}(\boldsymbol{\Lambda}) + (\boldsymbol{a}-\boldsymbol{m}_{\boldsymbol{x}})^H(\boldsymbol{a}-\boldsymbol{m}_{\boldsymbol{x}})$ where $\boldsymbol{m}_{\boldsymbol{x}}$ is the mean of $\boldsymbol{x}$ and $\boldsymbol{\Lambda}$ is the covariance matrix of $\boldsymbol{x}$ to compute the expectation over $\boldsymbol{h}$:

$$
\begin{aligned}
\mathcal{L}_3^{\mathsf{I-CSI-Rayleigh}} &= \mathbb{E}_{\boldsymbol{h},\boldsymbol{G}\sim q_{\boldsymbol{\lambda}}(\boldsymbol{h},\boldsymbol{G}|\boldsymbol{Y})}\Bigg[\frac{1}{\sigma_{\boldsymbol{w}}^2}\sum_{l=1}^{N_p}\big(\boldsymbol{y}_l - \sqrt{\rho}\boldsymbol{G}\mathrm{diag}(\boldsymbol{v}_l)\boldsymbol{h}x_l\big)^H \\
&\qquad \times \big(\boldsymbol{y}_l - \sqrt{\rho}\boldsymbol{G}\mathrm{diag}(\boldsymbol{v}_l)\boldsymbol{h}x_l\big)\Bigg] + C_1 \\
&= \frac{1}{\sigma_{\boldsymbol{w}}^2}\sum_{l=1}^{N_p}\mathbb{E}_{\boldsymbol{G}\sim q_{\boldsymbol{\lambda}_2}(\boldsymbol{G}|\boldsymbol{Y})}\Bigg[\rho|x_l|^2\mathrm{Tr}\big(\boldsymbol{G}\mathrm{diag}(\boldsymbol{v}_l)\mathrm{diag}(\boldsymbol{\gamma})\mathrm{diag}(\boldsymbol{v}_l)^H\boldsymbol{G}^H\big) \\
&\quad + \big(\boldsymbol{y}_l - \sqrt{\rho}\boldsymbol{G}\mathrm{diag}(\boldsymbol{v}_l)\boldsymbol{m}x_l\big)^H\big(\boldsymbol{y}_l - \sqrt{\rho}\boldsymbol{G}\mathrm{diag}(\boldsymbol{v}_l)\boldsymbol{m}x_l\big) + C_1 \\
&= \frac{1}{\sigma_{\boldsymbol{w}}^2}\sum_{l=1}^{N_p}\mathbb{E}_{\boldsymbol{G}\sim q_{\boldsymbol{\lambda}_2}(\boldsymbol{G}|\boldsymbol{Y})}\Bigg[\rho|x_l|^2\mathrm{Tr}\big(\boldsymbol{G}^H\boldsymbol{G}\ \mathrm{diag}(\boldsymbol{\gamma})\big) \\
&\quad + \big(\boldsymbol{y}_l - \sqrt{\rho}\boldsymbol{G}\mathrm{diag}(\boldsymbol{v}_l)\boldsymbol{m}x_l\big)^H\big(\boldsymbol{y}_l - \sqrt{\rho}\boldsymbol{G}\mathrm{diag}(\boldsymbol{v}_l)\boldsymbol{m}x_l\big) + C_1. \qquad \text{(A.1)}
\end{aligned}
$$

We note that the covariance of $\boldsymbol{h} \sim q_{\boldsymbol{\lambda}_1}(\boldsymbol{h}|\boldsymbol{Y})$ is defined by the diagonal matrix $\mathrm{diag}(\boldsymbol{\gamma})$ since the elements of $\boldsymbol{h}$ are assumed to be independent. Furthermore, we note that $\mathbb{E}_{\boldsymbol{G}}[\boldsymbol{G}^H\boldsymbol{G}] = \boldsymbol{R} + \boldsymbol{M}^H\boldsymbol{M}$ where $\boldsymbol{R} = \mathbb{E}_{\boldsymbol{G}}[(\boldsymbol{G}-\boldsymbol{M})^H(\boldsymbol{G}-\boldsymbol{M})]$ is the covariance matrix over the columns of $\boldsymbol{G}$. $\boldsymbol{R}$ is a diagonal matrix because the elements $\boldsymbol{G}_{i,j}$ are assumed

to be independent which makes the columns are independent as well and the elements on the main diagonal, denoted by $\boldsymbol{\tau}$, are given by $\boldsymbol{\tau}_i = \sum_{m=1}^{M} \boldsymbol{\Gamma}_{m,i}$. Therefore, $\mathcal{L}_3$ can be expressed as follows:

$$\begin{aligned}\mathcal{L}_3^{\mathsf{I-CSI-Rayleigh}} = \frac{1}{\sigma_{\boldsymbol{w}}^2}\Big[\rho||\boldsymbol{x}||^2\mathrm{Tr}\big(\mathrm{diag}(\boldsymbol{\tau})\ \mathrm{diag}(\boldsymbol{\gamma})\big)+\rho||\boldsymbol{x}||^2\boldsymbol{m}^H\mathrm{diag}(\boldsymbol{\tau})\boldsymbol{m} \\ + \rho||\boldsymbol{x}||^2\mathrm{Tr}\big(\boldsymbol{M}^H\boldsymbol{M}\mathrm{diag}(\boldsymbol{\gamma})\big)+\sum_{l=1}^{N_p}||\boldsymbol{y}_l - \sqrt{\rho}\boldsymbol{M}\mathrm{diag}(\boldsymbol{v}_l)\boldsymbol{m}x_l||^2\Big] + C_1.\end{aligned} \tag{A.2}$$

We derive the entropy of a complex Laplace random variable $z \sim \mathcal{CL}\big(m, b\big)$ with mean m and scale b:

$$\begin{aligned}H\big(q(z)\big) &= \int_{\mathbb{C}} -q(z)\log q(z)\ dz \\ &= \int_{\mathbb{C}} -\frac{1}{2\pi b^2}e^{-\frac{|z-m|}{b}}\log\frac{1}{2\pi b^2}e^{-\frac{|z-m|}{b}}dz \\ &= \log(2\pi b^2) + \int_{\mathbb{C}}\frac{|u|}{2\pi b^3}e^{-\frac{|u|}{b}}du \quad (u = z-m) \\ &= \log(2\pi b^2) + 2.\end{aligned} \tag{A.3}$$

Next, we derive the closed-form of $\mathcal{L}_3^{\mathsf{I-CSI}}$ (Eq. 2.28) with complex Laplace priors. In first step, we compute the expectation over $\boldsymbol{h}^{\mathsf{vir}}$ where we denote $\boldsymbol{A} = \frac{\sqrt{\rho}}{MN^2}\boldsymbol{F}_M^H\boldsymbol{G}^{\mathsf{vir}}\boldsymbol{F}_N^H\mathrm{diag}(\boldsymbol{v}_l)\boldsymbol{F}_N^H x_l$ which is a constant with respect to $\boldsymbol{h}^{\mathsf{vir}}$:

$$\begin{aligned}\mathcal{L}_3^{\mathsf{I-CSI}} &= \sum_{l=1}^{N_p}\mathbb{E}_{\boldsymbol{h}^{\mathsf{vir}},\boldsymbol{G}^{\mathsf{vir}}\sim q_{\boldsymbol{\lambda}}(\boldsymbol{h}^{\mathsf{vir}},\boldsymbol{G}^{\mathsf{vir}}|\boldsymbol{Y})}\Big[(\boldsymbol{y}_l - \boldsymbol{A}\boldsymbol{h}^{\mathsf{vir}})^H \times (\boldsymbol{y}_l - \boldsymbol{A}\boldsymbol{h}^{\mathsf{vir}})\Big] + C_1 \\ &= \sum_{l=1}^{N_p}\mathbb{E}_{\boldsymbol{G}^{\mathsf{vir}}\sim q_{\boldsymbol{\lambda}_2}(\boldsymbol{G}^{\mathsf{vir}}|\boldsymbol{Y})}\Big[\mathrm{Tr}\big(\boldsymbol{A}\boldsymbol{\Lambda}\boldsymbol{A}^H\big)+(\boldsymbol{y}_l - \boldsymbol{A}\boldsymbol{m})^H(\boldsymbol{y}_l - \boldsymbol{A}\boldsymbol{m})\Big] + C_1,\end{aligned} \tag{A.4}$$

where C_1 is a constant, $\boldsymbol{m}$ a vector of means of $\boldsymbol{h}^{\mathsf{vir}}$ following $q_{\boldsymbol{\lambda}_1}(\boldsymbol{h}^{\mathsf{vir}}|\boldsymbol{Y})$ distribution and $\boldsymbol{\Lambda} = \mathbb{E}_{\boldsymbol{h}^{\mathsf{vir}}\sim q_{\boldsymbol{\lambda}_1}(\boldsymbol{h}^{\mathsf{vir}}|\boldsymbol{Y})}[(\boldsymbol{h}^{\mathsf{vir}} - \boldsymbol{m})(\boldsymbol{h}^{\mathsf{vir}} - \boldsymbol{m})^H]$ is the covariance matrix of $\boldsymbol{h}^{\mathsf{vir}}$. The latter is a diagonal matrix with a main diagonal containing the variances of the elements.

The variance of a complex Laplace is defined as follows:

$$\begin{aligned}\mathrm{Var}(z) &= \int_{\mathbb{C}} \frac{|z-m|^2}{2\pi b^2} e^{-\frac{|z-m|}{b}} dz \\ &= \int_{\mathbb{C}} \frac{|u|^2}{2\pi b^2} e^{-\frac{|u|}{b}} du \quad \text{(Substitution } u = z - m\text{)} \\ &= \int_0^{2\pi} \int_0^{\infty} \frac{r^2}{2\pi b^2} e^{-\frac{r}{b}} r \; dr \; d\theta \quad \text{(polar coordinates)} \\ &= 6b^2. \end{aligned} \tag{A.5}$$

Hence, the covariance matrix $\boldsymbol{\Lambda}$ is expressed as follows:

$$\boldsymbol{\Lambda}_{i,j} = 6 \; \mathrm{diag}(\boldsymbol{b})^2. \tag{A.6}$$

To compute $\boldsymbol{G}^{\mathsf{vir}}$, we define a constant matrix $\boldsymbol{C} = \frac{\sqrt{\rho}}{MN^2} \boldsymbol{F}_N^H \mathrm{diag}(\boldsymbol{v}_l) \boldsymbol{F}_N^H x_l$, i.e, $\boldsymbol{A} = \boldsymbol{F}_M^H \boldsymbol{G}^{\mathsf{vir}} \boldsymbol{C}$. Hence, we get:

$$\begin{aligned}\mathcal{L}_3^{\mathsf{I-CSI}} &= \sum_{l=1}^{N_p} \mathbb{E}_{\boldsymbol{G}^{\mathsf{vir}} \sim q(\boldsymbol{G}^{\mathsf{vir}} | \boldsymbol{Y})} \Big[\mathrm{Tr}\big(\boldsymbol{A}^H \boldsymbol{A} \boldsymbol{\Lambda}\big) \\ &\quad + (\boldsymbol{y}_l - \boldsymbol{A}\boldsymbol{m})^H (\boldsymbol{y}_l - \boldsymbol{A}\boldsymbol{m}) \Big] + C_2 \\ &= \sum_{l=1}^{N_p} \mathbb{E}_{\boldsymbol{G}^{\mathsf{vir}} \sim q(\boldsymbol{G}^{\mathsf{vir}} | \boldsymbol{Y})} \Big[M \mathrm{Tr}\big(\boldsymbol{C}^H \boldsymbol{G}^{\mathsf{vir}^H} \boldsymbol{G}^{\mathsf{vir}} \boldsymbol{C} \boldsymbol{\Lambda}\big) \\ &\quad + (\boldsymbol{y}_l - \boldsymbol{F}_M^H \boldsymbol{G}^{\mathsf{vir}} \boldsymbol{C}\boldsymbol{m})^H (\boldsymbol{y}_l - \boldsymbol{F}_M^H \boldsymbol{G}^{\mathsf{vir}} \boldsymbol{C}\boldsymbol{m}) \Big] + C_2. \end{aligned} \tag{A.7}$$

Then we use the property $\mathbb{E}_{\boldsymbol{G}^{\mathsf{vir}}}[\boldsymbol{G}^{\mathsf{vir}^H} \boldsymbol{G}^{\mathsf{vir}}] = \boldsymbol{Q} + \boldsymbol{M}^H \boldsymbol{M}$ where $\boldsymbol{Q} = \mathbb{E}_{\boldsymbol{G}^{\mathsf{vir}}}[(\boldsymbol{G}^{\mathsf{vir}} - \boldsymbol{M})^H (\boldsymbol{G}^{\mathsf{vir}} - \boldsymbol{M})]$ is the covariance matrix over the columns of $\boldsymbol{G}^{\mathsf{vir}}$. $\boldsymbol{Q}$ is a diagonal matrix since the elements $\boldsymbol{G}_{i,j}^{\mathsf{vir}}$ are assumed to be independent which makes the columns are independent as well and the elements on the diagonal are given by:

$$\boldsymbol{Q}_{i,i} = \sum_{m=1}^{M} \mathrm{Var}(\boldsymbol{G}_{m,i}^{\mathsf{vir}}) = \sum_{m=1}^{M} 6\boldsymbol{B}_{m,i}^2. \tag{A.8}$$

Therefore, we have:

$$\mathcal{L}_3^{\mathsf{I-CSI}} = \sum_{l=1}^{N_p} \Big[M\mathrm{Tr}\big(\boldsymbol{C}^H\boldsymbol{Q}\boldsymbol{C}\boldsymbol{\Lambda}\big) + M\mathrm{Tr}\big(\boldsymbol{C}^H\boldsymbol{M}^H\boldsymbol{M}\boldsymbol{C}\boldsymbol{\Lambda}\big) + (\boldsymbol{y}_l - \boldsymbol{F}_M^H\boldsymbol{M}\boldsymbol{C}\boldsymbol{m})^H(\boldsymbol{y}_l - \boldsymbol{F}_M^H\boldsymbol{M}\boldsymbol{C}\boldsymbol{m}) + M\boldsymbol{m}^H\boldsymbol{C}^H\boldsymbol{Q}\boldsymbol{C}\boldsymbol{m} \Big] + C_1. \tag{A.9}$$

Bibliography

[1] B. Zheng, C. You, W. Mei, and R. Zhang, "A survey on channel estimation and practical passive beamforming design for intelligent reflecting surface aided wireless communications," *IEEE Communications Surveys & Tutorials*, vol. 24, no. 2, pp. 1035–1071, 2022.

[2] 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, 2019.

[3] S. Dang, O. Amin, B. Shihada, and M.-S. Alouini, "What should 6G be?," *Nature Electronics*, vol. 3, no. 1, pp. 20–29, 2020.

[4] Q.-U.-A. Nadeem, A. Kammoun, A. Chaaban, M. Debbah, and M.-S. Alouini, "Intelligent reflecting surface assisted wireless communication: Modeling and channel estimation," *arXiv preprint arXiv:1906.02360*, 2019.

[5] X. Shao, C. You, W. Ma, X. Chen, and R. Zhang, "Target sensing with intelligent reflecting surface: Architecture and performance," *IEEE Journal on Selected Areas in Communications*, vol. 40, no. 7, pp. 2070–2084, 2022.

[6] X. Pei, H. Yin, L. Tan, L. Cao, Z. Li, K. Wang, K. Zhang, and E. Björnson, "RIS-aided wireless communications: Prototyping, adaptive beamforming, and in-

door/outdoor field trials," *IEEE Transactions on Communications*, vol. 69, no. 12, pp. 8627–8640, 2021.

[7] Y. Liu, X. Liu, X. Mu, T. Hou, J. Xu, M. Di Renzo, and N. Al-Dhahir, "Reconfigurable intelligent surfaces: Principles and opportunities," *IEEE communications surveys & tutorials*, vol. 23, no. 3, pp. 1546–1577, 2021.

[8] L. You, J. Xiong, D. W. K. Ng, C. Yuen, W. Wang, and X. Gao, "Energy efficiency and spectral efficiency tradeoff in RIS-aided multiuser MIMO uplink transmission," *IEEE Transactions on Signal Processing*, vol. 69, pp. 1407–1421, 2020.

[9] P. Staat, H. Elders-Boll, M. Heinrichs, R. Kronberger, C. Zenger, and C. Paar, "Intelligent reflecting surface-assisted wireless key generation for low-entropy environments," in *2021 IEEE 32nd Annual International Symposium on Personal, Indoor and Mobile Radio Communications (PIMRC)*, pp. 745–751, IEEE, 2021.

[10] S. Buzzi, I. Chih-Lin, T. E. Klein, H. V. Poor, C. Yang, and A. Zappone, "A survey of energy-efficient techniques for 5G networks and challenges ahead," *IEEE Journal on selected areas in communications*, vol. 34, no. 4, pp. 697–709, 2016.

[11] Z. Lin, H. Niu, K. An, Y. Wang, G. Zheng, S. Chatzinotas, and Y. Hu, "Refracting RIS-aided hybrid satellite-terrestrial relay networks: Joint beamforming design and optimization," *IEEE Transactions on Aerospace and Electronic Systems*, vol. 58, no. 4, pp. 3717–3724, 2022.

[12] S. Kisseleff, W. A. Martins, H. Al-Hraishawi, S. Chatzinotas, and B. Ottersten, "Reconfigurable intelligent surfaces for smart cities: Research challenges and opportunities," *IEEE Open Journal of the Communications Society*, vol. 1, pp. 1781–1797, 2020.

[13] R. Chen, M. Liu, Y. Hui, N. Cheng, and J. Li, "Reconfigurable intelligent surfaces for 6G IoT wireless positioning: A contemporary survey," *IEEE Internet of Things Journal*, vol. 9, no. 23, pp. 23570–23582, 2022.

[14] S. F. Drampalou, N. I. Miridakis, H. C. Leligou, and P. A. Karkazis, "A survey on optimal channel estimation methods for RIS-aided communication systems," *Signals*, vol. 4, no. 1, pp. 208–234, 2023.

[15] Z. Peng, C. Pan, G. Zhou, and H. Ren, "Error propagation and overhead reduced channel estimation for RIS-aided multi-user mmwave systems," in *2022 International Symposium on Wireless Communication Systems (ISWCS)*, pp. 1–6, IEEE, 2022.

[16] Z. Peng, G. Zhou, C. Pan, H. Ren, A. L. Swindlehurst, P. Popovski, and G. Wu, "Channel estimation for RIS-aided multi-user mmwave systems with uniform planar arrays," *IEEE Transactions on Communications*, vol. 70, no. 12, pp. 8105–8122, 2022.

[17] Z.-Q. He and X. Yuan, "Cascaded channel estimation for large intelligent metasurface assisted massive MIMO," *IEEE Wireless Communications Letters*, vol. 9, no. 2, pp. 210–214, 2019.

[18] P. Wang, J. Fang, H. Duan, and H. Li, "Compressed channel estimation for intelligent reflecting surface-assisted millimeter wave systems," *IEEE signal processing letters*, vol. 27, pp. 905–909, 2020.

[19] B. Zheng, C. You, and R. Zhang, "Intelligent reflecting surface assisted multi-user OFDMA: Channel estimation and training design," *IEEE Transactions on Wireless Communications*, vol. 19, no. 12, pp. 8315–8329, 2020.

[20] K. Ardah, S. Gherekhloo, A. L. de Almeida, and M. Haardt, "TRICE: A channel estimation framework for RIS-aided millimeter-wave MIMO systems," *IEEE signal processing letters*, vol. 28, pp. 513–517, 2021.

[21] X. Hu, R. Zhang, and C. Zhong, "Semi-passive elements assisted channel estimation for intelligent reflecting surface-aided communications," *IEEE Transactions on Wireless Communications*, vol. 21, no. 2, pp. 1132–1142, 2021.

[22] I.-s. Kim, M. Bennis, J. Oh, J. Chung, and J. Choi, "Bayesian channel estimation for intelligent reflecting surface-aided mmwave massive MIMO systems with semi-passive elements," *IEEE Transactions on Wireless Communications*, 2023.

[23] G. T. de Araújo, A. L. De Almeida, and R. Boyer, "Channel estimation for intelligent reflecting surface assisted MIMO systems: A tensor modeling approach," *IEEE Journal of Selected Topics in Signal Processing*, vol. 15, no. 3, pp. 789–802, 2021.

[24] M. K. Samimi, G. R. MacCartney, S. Sun, and T. S. Rappaport, "28 GHz millimeter-wave ultrawideband small-scale fading models in wireless channels," in *2016 IEEE 83rd Vehicular Technology Conference (VTC Spring)*, pp. 1–6, IEEE, 2016.

[25] H. Zhang, S. Venkateswaran, and U. Madhow, "Channel modeling and MIMO capacity for outdoor millimeter wave links," in *2010 IEEE wireless communication and networking conference*, pp. 1–6, IEEE, 2010.

[26] S. Rangan, T. S. Rappaport, and E. Erkip, "Millimeter-wave cellular wireless networks: Potentials and challenges," *Proceedings of the IEEE*, vol. 102, no. 3, pp. 366–385, 2014.

[27] G. Zhou, C. Pan, H. Ren, P. Popovski, and A. L. Swindlehurst, "Channel estimation for RIS-aided multiuser millimeter-wave systems," *IEEE Transactions on Signal Processing*, vol. 70, pp. 1478–1492, 2022.

[28] M. R. Akdeniz, Y. Liu, M. K. Samimi, S. Sun, S. Rangan, T. S. Rappaport, and E. Erkip, "Millimeter wave channel modeling and cellular capacity evaluation," *IEEE journal on selected areas in communications*, vol. 32, no. 6, pp. 1164–1179, 2014.

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

[30] M.-M. Zhao, Q. Wu, M.-J. Zhao, and R. Zhang, "Intelligent reflecting surface enhanced wireless networks: Two-timescale beamforming optimization," *IEEE Transactions on Wireless Communications*, vol. 20, no. 1, pp. 2–17, 2020.

[31] F. Yang, J.-B. Wang, H. Zhang, C. Chang, and J. Cheng, "Intelligent reflecting surface-assisted mmwave communication exploiting statistical CSI," in *ICC 2020-2020 IEEE International Conference on Communications (ICC)*, pp. 1–6, IEEE, 2020.

[32] K. Zhi, C. Pan, H. Ren, and K. Wang, "Statistical CSI-based design for reconfigurable intelligent surface-aided massive MIMO systems with direct links," *IEEE Wireless Communications Letters*, vol. 10, no. 5, pp. 1128–1132, 2021.

[33] R. Schmidt, "Multiple emitter location and signal parameter estimation," *IEEE transactions on antennas and propagation*, vol. 34, no. 3, pp. 276–280, 1986.

[34] V. Molodtsov, A. Kureev, and E. Khorov, "Experimental study of smoothing modifications of the MUSIC algorithm for direction of arrival estimation in indoor environments," *IEEE Access*, vol. 9, pp. 153767–153774, 2021.

[35] A. L. Swindlehurst, G. Zhou, R. Liu, C. Pan, and M. Li, "Channel estimation with reconfigurable intelligent surfaces—a general framework," *Proceedings of the IEEE*, vol. 110, no. 9, pp. 1312–1338, 2022.

[36] H. Wang, J. Fang, H. Duan, and H. Li, "Spatial channel covariance estimation and two-timescale beamforming for IRS-assisted millimeter wave systems," *IEEE Transactions on Wireless Communications*, 2023.

[37] S. Haghighatshoar and G. Caire, "Massive MIMO channel subspace estimation from low-dimensional projections," *IEEE Transactions on Signal Processing*, vol. 65, no. 2, pp. 303–318, 2016.

[38] D. G. Tzikas, A. C. Likas, and N. P. Galatsanos, "The variational approximation for bayesian inference," *IEEE Signal Processing Magazine*, vol. 25, no. 6, pp. 131–146, 2008.

[39] D. M. Blei, A. Kucukelbir, and J. D. McAuliffe, "Variational inference: A review for statisticians," *Journal of the American statistical Association*, vol. 112, no. 518, pp. 859–877, 2017.

[40] M. D. Hoffman, D. M. Blei, C. Wang, and J. Paisley, "Stochastic variational inference," *Journal of Machine Learning Research*, 2013.

[41] C. Wang and D. M. Blei, "Variational inference in nonconjugate models," *Journal of Machine Learning Research*, 2013.

[42] D. D. Lin and T. J. Lim, "The variational inference approach to joint data detection and phase noise estimation in OFDM," *IEEE Transactions on Signal Processing*, vol. 55, no. 5, pp. 1862–1874, 2007.

[43] F. Li, S. Zhu, and M. Rong, "Detection for OFDM systems with channel estimation errors using variational inference," *IEEE Signal Processing Letters*, vol. 16, no. 5, pp. 434–437, 2009.

[44] V. Raj and S. Kalyani, "Design of communication systems using deep learning: A variational inference perspective," *IEEE Transactions on Cognitive Communications and Networking*, vol. 6, no. 4, pp. 1320–1334, 2020.

[45] Y. Miao, L. Yu, and P. Blunsom, "Neural variational inference for text processing," in *International conference on machine learning*, pp. 1727–1736, PMLR, 2016.

[46] X. Li, J. Fang, H. Li, and P. Wang, "Millimeter wave channel estimation via exploiting joint sparse and low-rank structures," *IEEE Transactions on Wireless Communications*, vol. 17, no. 2, pp. 1123–1133, 2017.

[47] M. Xiao, S. Mumtaz, Y. Huang, L. Dai, Y. Li, M. Matthaiou, G. K. Karagiannidis, E. Björnson, K. Yang, I. Chih-Lin, *et al.*, "Millimeter wave communications for future mobile networks," *IEEE Journal on Selected Areas in Communications*, vol. 35, no. 9, pp. 1909–1935, 2017.

[48] I. A. Hemadeh, K. Satyanarayana, M. El-Hajjar, and L. Hanzo, "Millimeter-wave communications: Physical channel models, design considerations, antenna constructions, and link-budget," *IEEE Communications Surveys & Tutorials*, vol. 20, no. 2, pp. 870–913, 2017.

[49] T. Lin, X. Yu, Y. Zhu, and R. Schober, "Channel estimation for IRS-assisted millimeter-wave MIMO systems: Sparsity-inspired approaches," *IEEE Transactions on Communications*, vol. 70, no. 6, pp. 4078–4092, 2022.

[50] D. P. Kingma and J. Ba, "Adam: A method for stochastic optimization," *arXiv preprint arXiv:1412.6980*, 2014.

[51] T. Lin, X. Yu, Y. Zhu, and R. Schober, "Channel estimation for intelligent reflecting surface-assisted millimeter wave MIMO systems," in *GLOBECOM 2020-2020 IEEE Global Communications Conference*, pp. 1–6, IEEE, 2020.

[52] J. Chen, Y.-C. Liang, H. V. Cheng, and W. Yu, "Channel estimation for reconfigurable intelligent surface aided multi-user mmwave MIMO systems," *IEEE Transactions on Wireless Communications*, 2023.

[53] V. D. P. Souto, R. D. Souza, B. F. Uchoa-Filho, A. Li, and Y. Li, "Beamforming optimization for intelligent reflecting surfaces without CSI," *IEEE Wireless Communications Letters*, vol. 9, no. 9, pp. 1476–1480, 2020.

[54] M. Figurnov, S. Mohamed, and A. Mnih, "Implicit reparameterization gradients," *Advances in neural information processing systems*, vol. 31, 2018.

[55] A. Mezghani, F. Bellili, and E. Hossain, "Reconfigurable intelligent surfaces for quasi-passive mmwave and thz networks: Should they be reflective or redirective?," in *2022 56th Asilomar Conference on Signals, Systems, and Computers*, pp. 1076–1080, IEEE, 2022.

[56] D. Wijekoon, A. Mezghani, and E. Hossain, "Beamforming optimization in RIS-aided MIMO systems under multiple-reflection effects," in *ICASSP 2023-2023 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pp. 1–5, IEEE, 2023.