

# Variational Inference-Based Channel Estimation for Reconfigurable Intelligent Surface-Aided Wireless Systems

Firas Fredj, Amal Feriani, Amine Mezghani, and Ekram Hossain, *Fellow, IEEE*

**Abstract**—We propose a variational inference-based channel estimation method in fully passive reconfigurable intelligent surface (RIS)-aided mmWave single-user single-input multiple-output (SIMO) communication systems. The main goal is to jointly 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. Specifically, by using a variational inference framework, we approximate the posterior of the channels 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 a 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 UE-RIS and RIS-BS channels. We evaluate the proposed channel estimation solution 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.

**Index Terms**—Reconfigurable Intelligent Surface (RIS), channel estimation, Variational Inference (VI), mmWave communications

## I. INTRODUCTION

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 [1]. Although mmWave communications offer 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 losses. 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 [2] [3]. An 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 shift pattern [4] [5]. To achieve the desired performance (e.g. high spectral

The authors are with the Department of Electrical and Computer Engineering at the University of Manitoba, Winnipeg, MB, Canada. (emails: {fredjf1,feriania }@umanitoba.ca, {Amine.Mezghani,Ekram.Hossain}@umanitoba.ca). This work was supported by the Discovery Grants Program of the Natural Sciences and Engineering Research Council of Canada (NSERC).

efficiency, energy efficiency) through passive and active beam-forming, it is crucial to have the channel state information (CSI) between the base station (BS) and the RIS and between the RIS and the receiver [6], [7].

Channel estimation in RIS-aided systems has been extensively investigated in the literature along with the reflection optimization. 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 [8]. Based on the low-rank structure of the mmWave communication of the RIS-BS and UE-RIS channels that comes from the large number of elements in the RIS, a non-iterative channel estimation framework can be adopted through the estimation of the directions of departure (DODs) of the BS-RIS paths and the directions of arrival (DOAs) of the UEs-RIS paths in a first stage, then the cascaded BS-RIS-UE channel can be directly computed using the estimated DODs and DOAs [9]. To reduce the training overhead triggered by the large large dimension of the channels, 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. And using the property of static channel for the RIS-BS channels (since the RIS and BS are in fixed positions), only the UE-RIS channel is estimated in the training time of the subsequent coherence blocks [10]. 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 where they require the downlink and uplink signals for channel estimation [11]. Furthermore, the decomposition of the cascaded UE-RIS-BS channel into two separate channels, i.e. UE-RIS and RIS-BS channels, has been studied in RIS-aided systems with fully passive RIS. For instance, it was shown that the received signal follows the parallel factor tensor model [12] which is used to develop channel estimation methods based on Khatri-Rao factorization of the cascaded channel and iterative alternating estimation scheme.

In this paper, we propose a joint channel estimation algorithm based on VI [13], [14] for RIS-aided single-user single-input multiple-output (SIMO) systems with fully passive elements to separately estimate the UE-RIS and RIS-BS channels. More specifically, we estimate the channels based on the uplink training signals using maximum a posteriori (MAP) estimation where we approximate the true posterior

of the channels based on convenient prior distributions. We consider two families of channel distribution: (i) complex Gaussian priors for Rayleigh fading channels, and (ii) complex Laplace priors to exploit the sparse structure in the angular domain in mmWave communication. Using the mean-field approximation [13], the VI-based framework enables the joint estimation of the RIS-BS and UE-RIS channels. **In contrast to the aforementioned works (e.g. [11], [12]), we use VI-based unsupervised learning to solve the joint estimation of the UE-RIS and RIS-BS channels using only the uplink training signals in a fully passive RIS setup.** Our solution is flexible and takes into account the sparsity of mmWave channels. *The proposed algorithm can also be extended to other types of channels.* The simulation results show the effectiveness of the proposed approach in terms of the capacity of the UE-RIS-BS link and the performance improvement with sparse channels compared to channels with Gaussian priors.

**Notations:** Scalars, vectors and matrices are denoted  $x$ ,  $\mathbf{x}$  and  $\mathbf{X}$ , respectively.  $\mathbf{X}^*$  and  $\mathbf{X}^H$  denote the complex conjugate and conjugate transpose of  $\mathbf{X}$ . The  $i$ -th element of a vector  $\mathbf{a}$  is  $[\mathbf{a}]_i$ , while the  $(i, j)$ -th element of a matrix  $\mathbf{A}$  is  $[\mathbf{A}]_{i,j}$ . The  $n \times n$  identity matrix is written as  $\mathbf{I}_n$ . The  $\text{diag}(\mathbf{a})$  is the diagonal matrix with the elements of the vector  $\mathbf{a}$  on the main diagonal. The element-wise product of  $\mathbf{X}$  and  $\mathbf{Y}$  is written as  $\mathbf{X} \odot \mathbf{Y}$ .  $\text{Tr}(\mathbf{X})$  and  $|\mathbf{X}|$  represent the trace and determinant of the matrix  $\mathbf{X}$ , respectively, and  $|x|$  represent the absolute value of a complex number  $x$ . The notation  $\text{vec}(\mathbf{A})$  denotes the vectorization of matrix  $\mathbf{A}$ . The complex Gaussian random vector is denoted  $\mathbf{x} \sim \mathcal{CN}(\mathbf{m}, \mathbf{\Sigma})$  with mean  $\mathbf{m}$  and covariance matrix  $\mathbf{\Sigma}$ , whereas a complex Laplace random variable  $x$  is denoted  $x \sim \mathcal{CL}(m, b)$  with mean  $m$ , scale  $b$  and probability density function (PDF)  $p(x) = (1/(2\pi b^2)) \exp(-|x - m|/b)$ .

## II. SYSTEM MODEL AND PROBLEM FORMULATION

We consider an RIS-assisted single-user communication system with  $M$  antennas at the BS,  $N$  passive reflecting elements at the RIS and a single-antenna user. Considering uplink transmission, the UE-RIS and RIS-BS channels are denoted by  $\mathbf{h} \in \mathbb{C}^N$  and  $\mathbf{G} \in \mathbb{C}^{M \times N}$ , respectively. We ignore the direct UE-BS link considering that it can be estimated using the conventional SIMO channel estimation methods by turning off the RIS. Furthermore, we adopt a block-fading channel model with a coherence time  $T$ . Hence, the received signal at the BS at the  $t$ -th time slot can be expressed as follows:

$$\mathbf{y}[t] = \sqrt{\rho} \mathbf{G} \text{diag}(\mathbf{v}[t]) \mathbf{h} \mathbf{x}[t] + \mathbf{w}[t], \quad t = 1, \dots, T, \quad (1)$$

where  $\rho$ ,  $\mathbf{x} \in \mathbb{C}$  and  $\mathbf{w} \in \mathbb{C}^M$  are, respectively, the transmission power, the transmitted signal and the additive white noise with  $\sigma_w^2$  being the noise power, i.e.  $\mathbf{w}[t] \sim \mathcal{CN}(\mathbf{0}, \sigma_w^2 \mathbf{I}_M)$ . The phase shifts contributed by the RIS are represented by the diagonal matrix  $\text{diag}(\mathbf{v})$ , where  $\mathbf{v} = [e^{j\theta_1}, \dots, e^{j\theta_N}]$  with  $\theta_n \in [0, 2\pi)$  is the phase shift of the  $n$ -th element in the RIS.

In order to jointly estimate the UE-RIS channel  $\mathbf{h}$  and the RIS-BS channel  $\mathbf{G}$ , a pilot sequence is sent by the user to

the BS through the UE-RIS-BS link. Different phase shifts are considered for each pilot signal. By sending  $L < T$  pilot signals denoted  $\mathbf{x} \in \mathbb{C}^L$ , the received pilot signal is given by:

$$\mathbf{Y} = \sqrt{\rho} \mathbf{G} (\mathbf{V} \odot (\mathbf{h} \mathbf{x}^T)) + \mathbf{W}, \quad (2)$$

where  $\mathbf{Y} = [\mathbf{y}[1], \dots, \mathbf{y}[L]] \in \mathbb{C}^{M \times L}$  is the concatenation of the received signals,  $\mathbf{x} = [x[1], \dots, x[L]]^T$  the pilot signal,  $\mathbf{V} = [v[1], \dots, v[L]]$  is composed of the phase shift vectors with  $v[l]$  is assigned to the  $l$ -th pilot signal  $x[l]$ , and  $\mathbf{W} = [w[1], \dots, w[L]]$ .

The channels  $\mathbf{h}$  and  $\mathbf{G}$  can be computed using the MAP estimation by maximizing the posterior distribution conditioned on the observations  $\mathbf{Y}$ :

$$\hat{\mathbf{h}}, \hat{\mathbf{G}} = \arg \max_{\mathbf{h}, \mathbf{G}} p(\mathbf{h}, \mathbf{G} | \mathbf{Y}). \quad (3)$$

However, the true posterior distribution of the channels  $p(\mathbf{G}, \mathbf{h} | \mathbf{Y})$  is intractable. Therefore, we aim to approximate the true posterior distribution  $p$  using tractable distributions based on the VI framework. In the next section, we will introduce the necessary background on VI and detail the proposed joint channel estimation algorithm.

## III. JOINT CHANNEL ESTIMATION VIA VARIATIONAL INFERENCE

### A. Variational Inference

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 [13] [15].

Applied to the joint channel estimation problem described in Section II, our aim is to approximate the intractable posterior  $p(\mathbf{h}, \mathbf{G} | \mathbf{Y})$  by a tractable distribution  $q(\mathbf{h}, \mathbf{G} | \mathbf{Y})$  where the channels  $\mathbf{G}$  and  $\mathbf{h}$  determine the received pilot signals  $\mathbf{Y}$ , i.e. the observations. To do so, we start by expressing the log-likelihood function of  $\mathbf{Y}$ :

$$\begin{aligned} \log p(\mathbf{Y}) &= \int_{\mathbf{h}, \mathbf{G}} q(\mathbf{h}, \mathbf{G} | \mathbf{Y}) \cdot \log p(\mathbf{Y}) \, d\mathbf{h} \, d\mathbf{G} \\ &= \int_{\mathbf{h}, \mathbf{G}} q(\mathbf{h}, \mathbf{G} | \mathbf{Y}) \log \left[ \frac{p(\mathbf{h}, \mathbf{G}, \mathbf{Y})}{q(\mathbf{h}, \mathbf{G} | \mathbf{Y})} \cdot \frac{q(\mathbf{h}, \mathbf{G} | \mathbf{Y})}{p(\mathbf{h}, \mathbf{G} | \mathbf{Y})} \right] \, d\mathbf{h} \, d\mathbf{G} \\ &= \mathbb{E}_{\mathbf{h}, \mathbf{G} \sim q(\mathbf{h}, \mathbf{G} | \mathbf{Y})} \left[ \log \frac{p(\mathbf{h}, \mathbf{G}, \mathbf{Y})}{q(\mathbf{h}, \mathbf{G} | \mathbf{Y})} \right] \\ &\quad + \underbrace{\mathbb{E}_{\mathbf{h}, \mathbf{G} \sim q(\mathbf{h}, \mathbf{G} | \mathbf{Y})} \left[ \log \frac{q(\mathbf{h}, \mathbf{G} | \mathbf{Y})}{p(\mathbf{h}, \mathbf{G} | \mathbf{Y})} \right]}_{\text{Kullback-Leibler divergence}}. \end{aligned} \quad (4)$$

The second term in Eq. 4 is the Kullback-Leibler (KL) divergence  $D_{KL}(q(\mathbf{h}, \mathbf{G} | \mathbf{Y}) || p(\mathbf{h}, \mathbf{G} | \mathbf{Y}))$  which is a statistical distance in the distribution space that measures how close the approximated distribution  $q(\mathbf{h}, \mathbf{G} | \mathbf{Y})$  to the exact posterior  $p(\mathbf{h}, \mathbf{G} | \mathbf{Y})$ . The minimization of the KL divergence yields a better approximation of the exact posterior, thus we can derive the estimated channels using the approximated distribution  $q(\mathbf{h}, \mathbf{G} | \mathbf{Y})$ . However, the KL-divergence cannot be derived directly since the PDF of the posterior is intractable. Thus, a lower bound, also named *Evidence Lower Bound* (ELBO), on the

log-likelihood of the observations can be derived given that the KL divergence is a non-negative term:

$$\begin{aligned} \log p(\mathbf{Y}) &\geq \mathbb{E}_{\mathbf{h}, \mathbf{G} \sim q(\mathbf{h}, \mathbf{G} | \mathbf{Y})} \left[ \log \frac{p(\mathbf{h}, \mathbf{G}, \mathbf{Y})}{q(\mathbf{h}, \mathbf{G} | \mathbf{Y})} \right] \\ &\triangleq -\mathcal{L}(\mathbf{Y}; q). \end{aligned} \quad (5)$$

Given that  $\log p(\mathbf{Y})$  in Eq. 4 is an unknown constant, maximizing the ELBO  $-\mathcal{L}(\mathbf{Y}; q)$  is equivalent to minimizing the KL divergence, which solves the approximation problem of the posterior  $p(\mathbf{h}, \mathbf{G} | \mathbf{Y})$ . Assuming that  $q(\mathbf{h}, \mathbf{G} | \mathbf{Y})$  belongs to a family of tractable distributions defined by a set of parameters  $\lambda$ , the VI approach consists in optimizing the parameters  $\lambda$  of the approximated distribution  $q(\mathbf{h}, \mathbf{G} | \mathbf{Y})$  that minimizes the objective function  $\mathcal{L}(\mathbf{Y}; \lambda)$ .

We further assume that the approximated distribution can be factorized as  $q(\mathbf{h}, \mathbf{G} | \mathbf{Y}) = q(\mathbf{h} | \mathbf{Y}) \cdot q(\mathbf{G} | \mathbf{Y})$  and we optimize the independent distributions by minimizing  $\mathcal{L}(\mathbf{Y}; \lambda_G, \lambda_h)$ , which is written as  $\mathcal{L}(q)$  for ease of notation. This independence assumption is referred to as the *mean-field approximation* [15]. It is equivalent to assuming a low correlation between the channels  $\mathbf{G}$  and  $\mathbf{h}$  conditioned on  $\mathbf{Y}$ . Hence, the objective function is simplified to a general form given by:

$$\begin{aligned} \mathcal{L}(q) &= \mathbb{E}_{\mathbf{h}, \mathbf{G} \sim q(\mathbf{h}, \mathbf{G} | \mathbf{Y})} \left[ \log \frac{q(\mathbf{h}, \mathbf{G} | \mathbf{Y})}{p(\mathbf{h}, \mathbf{G}, \mathbf{Y})} \right] \\ &= \underbrace{\mathbb{E}_{\mathbf{h} \sim q(\mathbf{h} | \mathbf{Y})} \left[ \log \frac{q(\mathbf{h} | \mathbf{Y})}{p(\mathbf{h})} \right]}_{\mathcal{L}_1} + \underbrace{\mathbb{E}_{\mathbf{G} \sim q(\mathbf{G} | \mathbf{Y})} \left[ \log \frac{q(\mathbf{G} | \mathbf{Y})}{p(\mathbf{G})} \right]}_{\mathcal{L}_2} \\ &\quad - \underbrace{\mathbb{E}_{\mathbf{h}, \mathbf{G} \sim q(\mathbf{h}, \mathbf{G} | \mathbf{Y})} [\log p(\mathbf{Y} | \mathbf{h}, \mathbf{G})]}_{\mathcal{L}_3}. \end{aligned} \quad (6)$$

We obtain the parameters of the two variational distributions  $q(\mathbf{G} | \mathbf{Y})$  and  $q(\mathbf{h} | \mathbf{Y})$  by two trainable neural networks denoted by *Encoder 1* and *Encoder 2*, as shown in Fig. 1. In particular, given the received pilot signal  $\mathbf{Y}$  as an input, the neural networks predict the statistical parameters of the distributions  $q(\mathbf{G} | \mathbf{Y})$  and  $q(\mathbf{h} | \mathbf{Y})$ .

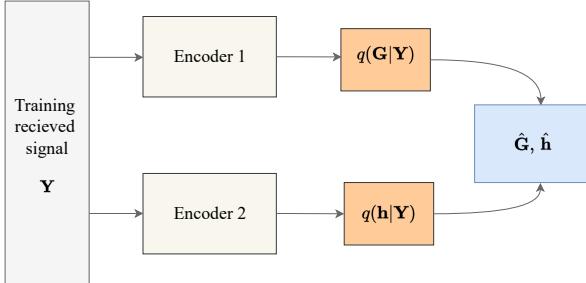


Fig. 1: Joint channel estimation framework structure.

Note that  $\mathcal{L}_1$  and  $\mathcal{L}_2$  in Eq. 6 represent the KL divergence between the variational distributions  $q(\mathbf{h} | \mathbf{Y})$  and  $q(\mathbf{G} | \mathbf{Y})$  returned by the neural networks and their actual priors  $p(\mathbf{h})$  and  $p(\mathbf{G})$ , respectively. Regarding  $\mathcal{L}_3$ , it corresponds to the reconstruction error of the estimated pilot signal  $\hat{\mathbf{Y}}$  with the variational distributions  $q(\mathbf{G} | \mathbf{Y})$  and  $q(\mathbf{h} | \mathbf{Y})$ . Hence, minimizing the objective function  $\mathcal{L}(q) = \mathcal{L}_1 + \mathcal{L}_2 + \mathcal{L}_3$  ensures that the generated posterior distributions are close to the prior distributions and the reconstructed signal  $\hat{\mathbf{Y}}$  is similar to the received signal.

Deriving the ELBO is the first step in a VI-based algorithm design. The next step focuses on the choice of the approximating distribution family. In what follows, we study two cases: the first one uses complex Gaussian priors for Rayleigh fading channels; and the second one assumes a complex Laplace distribution in the angular domain.

## B. Channel Estimation With Complex Gaussian Priors

In order to derive an explicit expression of the objective function in Eq. 6, we consider the Rayleigh fading channel for  $\mathbf{G}$  and  $\mathbf{h}$  where the elements of the channels are independent and identically distributed (i.i.d.) with variances  $\sigma_G^2$  and  $\sigma_h^2$ , respectively, i.e.  $p([\mathbf{G}]_{i,j}) \sim \mathcal{CN}(0, \sigma_G^2)$  and  $p([\mathbf{h}]_i) \sim \mathcal{CN}(0, \sigma_h^2)$ .

Additionally, the approximated posterior distributions  $q(\mathbf{G} | \mathbf{Y})$  and  $q(\mathbf{h} | \mathbf{Y})$  are assumed to be Gaussian distributions parameterized with means  $\mathbf{M}_G$  and  $\mathbf{m}_h$  and variances  $\Gamma_G$  and  $\gamma_h$ , respectively. Assuming that the channel elements are independent in the variational distributions, we have  $q([\mathbf{G}]_{i,j} | \mathbf{Y}) \sim \mathcal{CN}([\mathbf{M}_G]_{i,j}, [\Gamma_G]_{i,j})$  and  $q([\mathbf{h}]_i | \mathbf{Y}) \sim \mathcal{CN}([\mathbf{m}_h]_i, [\gamma_h]_i)$  where the means  $\mathbf{M}_G$  and  $\mathbf{m}_h$  and the variances  $\Gamma_G$  and  $\gamma_h$  are predicted by the neural networks *Encoder 1* and *Encoder 2*, respectively, given the pilot signals  $\mathbf{Y}$  (Fig. 1).

We recall that  $\mathcal{L}_1$  has a closed-form expression since it is the KL divergence between two Gaussian distributions  $p(\mathbf{h})$  and  $q(\mathbf{h} | \mathbf{Y})$ :

$$\begin{aligned} \mathcal{L}_1 &= \mathbb{E}_{\mathbf{h} \sim q(\mathbf{h} | \mathbf{Y})} \left[ \log \frac{q(\mathbf{h} | \mathbf{Y})}{p(\mathbf{h})} \right] \\ &= \frac{1}{\sigma_h^2} \left( \text{tr}(\text{diag}(\gamma_h)) + \mathbf{m}_h^H \mathbf{m}_h \right) - \log |\text{diag}(\gamma_h)| \\ &\quad + N \log \sigma_h^2 - N \\ &= \frac{1}{\sigma_h^2} \sum_{i=1}^N [\gamma_h]_i + \frac{\mathbf{m}_h^H \mathbf{m}_h}{\sigma_h^2} - \sum_{i=1}^N \log [\gamma_h]_i + N \log \sigma_h^2 - N. \end{aligned} \quad (7)$$

Similarly, we derive the closed-form expression of  $\mathcal{L}_2$  as follows:

$$\begin{aligned} \mathcal{L}_2 &= \mathbb{E}_{\mathbf{G} \sim q(\mathbf{G} | \mathbf{Y})} \left[ \log \frac{q(\mathbf{G} | \mathbf{Y})}{p(\mathbf{G})} \right] \\ &= \frac{1}{\sigma_G^2} \sum_{i=1}^M \sum_{j=1}^N [\Gamma_G]_{i,j} + \frac{\text{vec}(\mathbf{M}_G)^H \text{vec}(\mathbf{M}_G)}{\sigma_G^2} \\ &\quad - \sum_{i=1}^M \sum_{j=1}^N \log [\Gamma_G]_{i,j} + NM \log \sigma_G^2 - NM. \end{aligned} \quad (8)$$

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

$$\begin{aligned} \mathcal{L}_3 &= -\mathbb{E}_{\mathbf{h}, \mathbf{G} \sim q(\mathbf{h}, \mathbf{G} | \mathbf{Y})} [\log p(\mathbf{Y} | \mathbf{h}, \mathbf{G})] \\ &= \frac{1}{\sigma_w^2} \left[ \rho \|\mathbf{x}\|^2 \text{tr}(\text{diag}(\boldsymbol{\tau}) \text{diag}(\gamma_h)) + \rho \|\mathbf{x}\|^2 \mathbf{m}_h^H \text{diag}(\boldsymbol{\tau}) \mathbf{m}_h \right. \\ &\quad \left. + \rho \|\mathbf{x}\|^2 \text{tr}(\mathbf{M}_G^H \mathbf{M}_G \text{diag}(\gamma_h)) \right. \\ &\quad \left. + \sum_{l=1}^L \|\mathbf{y}_l - \sqrt{\rho} \mathbf{G} \text{diag}(\mathbf{v}_l) \mathbf{m}_h \mathbf{x}_l\|^2 \right] + C_1, \end{aligned} \quad (9)$$

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

## C. Channel Estimation With Complex Laplace Prior in the Angular Domain

With a large number of elements in the RIS, the Rayleigh fading channel model may not be appropriate as the channels are more sparse in the angular domain [2]. Therefore, we model the sparse channels as

channels that follow the Laplace distributions in the angular domain. In particular, by applying Discrete Fourier Transform (DFT) to the channels  $\mathbf{G}$  and  $\mathbf{h}$ , the obtained channels are sparse and modeled as a random complex Laplace matrix and vector, respectively:

$$[\tilde{\mathbf{G}}]_{i,j} = [\mathbf{F}_M \mathbf{G} \mathbf{F}_N]_{i,j} \sim \mathcal{CL}(0, \alpha_{\tilde{\mathbf{G}}}); \quad (10)$$

$$[\tilde{\mathbf{h}}]_i = [\mathbf{F}_N \mathbf{h}]_i \sim \mathcal{CL}(0, \alpha_{\tilde{\mathbf{h}}}), \quad (11)$$

where  $\mathbf{F}_N$  and  $\mathbf{F}_M$  are the DFT matrices of size  $N \times N$  and  $M \times M$ , respectively. Then,  $\tilde{\mathbf{G}}$  and  $\tilde{\mathbf{h}}$  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  $\alpha_{\tilde{\mathbf{G}}}$  and  $\alpha_{\tilde{\mathbf{h}}}$ , respectively. Given that  $\mathbf{F}_N^{-1} = \frac{1}{N} \mathbf{F}_N^H$  for any DFT matrix of size  $N \times N$ , the received signal is expressed as follows:

$$\mathbf{y} = \frac{\sqrt{\rho}}{MN^2} \mathbf{F}_M^H \tilde{\mathbf{G}} \mathbf{F}_N^H \text{diag}(\mathbf{v}) \mathbf{F}_N^H \tilde{\mathbf{h}} x + \mathbf{w}. \quad (12)$$

Rather than approximating the posterior distribution  $p(\mathbf{G}, \mathbf{h} | \mathbf{Y})$  of the actual channels  $\mathbf{G}$  and  $\mathbf{h}$ , we attempt to estimate the posterior distribution  $p(\tilde{\mathbf{G}}, \tilde{\mathbf{h}} | \mathbf{Y})$  of the channels in the angular domain. Similar to the methodology adopted for Gaussian priors, we use the mean-field approximation to factorize the approximated distribution  $q(\tilde{\mathbf{G}}, \tilde{\mathbf{h}} | \mathbf{Y}) = q(\tilde{\mathbf{G}} | \mathbf{Y}) \cdot q(\tilde{\mathbf{h}} | \mathbf{Y})$ . Moreover, we assume that the approximated distributions follow complex Laplace distributions with independent elements where  $q([\tilde{\mathbf{G}}]_{i,j} | \mathbf{Y}) \sim \mathcal{CL}([\mathbf{M}_{\tilde{\mathbf{G}}}]_{i,j}, [\mathbf{B}_{\tilde{\mathbf{G}}}]_{i,j})$  and  $q([\tilde{\mathbf{h}}]_i | \mathbf{Y}) \sim \mathcal{CL}([\mathbf{m}_{\tilde{\mathbf{h}}}]_i, [\mathbf{b}_{\tilde{\mathbf{h}}}]_i)$ . The parameters of the approximated posterior distributions  $\mathbf{M}_{\tilde{\mathbf{G}}}$ ,  $\mathbf{m}_{\tilde{\mathbf{h}}}$ ,  $\mathbf{B}_{\tilde{\mathbf{G}}}$  and  $\mathbf{b}_{\tilde{\mathbf{h}}}$  are returned by the neural networks *Encoder 1* and *Encoder 2*. Therefore, the expression of  $\mathcal{L}_1$  is given by:

$$\begin{aligned} \mathcal{L}_1 &= \mathbb{E}_{\tilde{\mathbf{h}} \sim q(\tilde{\mathbf{h}} | \mathbf{Y})} [\log q(\tilde{\mathbf{h}} | \mathbf{Y})] - \mathbb{E}_{\tilde{\mathbf{h}} \sim q(\tilde{\mathbf{h}} | \mathbf{Y})} [\log p(\tilde{\mathbf{h}})] \\ &= \sum_{i=1}^N \mathbb{E}_{[\tilde{\mathbf{h}}]_i \sim q([\tilde{\mathbf{h}}]_i | \mathbf{Y})} [\log q([\tilde{\mathbf{h}}]_i | \mathbf{Y})] \\ &\quad - \mathbb{E}_{[\tilde{\mathbf{h}}]_i \sim q([\tilde{\mathbf{h}}]_i | \mathbf{Y})} [\log p([\tilde{\mathbf{h}}]_i)] \\ &= \sum_{i=1}^N H(q([\tilde{\mathbf{h}}]_i | \mathbf{Y}), p([\tilde{\mathbf{h}}]_i)) - H(q([\tilde{\mathbf{h}}]_i | \mathbf{Y})), \end{aligned} \quad (14)$$

where  $H(q([\tilde{\mathbf{h}}]_i | \mathbf{Y}))$  is the entropy of  $q([\tilde{\mathbf{h}}]_i | \mathbf{Y})$  and  $H(q([\tilde{\mathbf{h}}]_i | \mathbf{Y}), p([\tilde{\mathbf{h}}]_i))$  is the cross entropy between  $q([\tilde{\mathbf{h}}]_i | \mathbf{Y})$  and  $p([\tilde{\mathbf{h}}]_i)$ . The entropy of the complex Laplace distribution is:

$$H(q([\tilde{\mathbf{h}}]_i | \mathbf{Y})) = \log(2\pi[\mathbf{b}]_i^2) + 2. \quad (15)$$

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  $\tilde{\mathbf{h}}$ , therefore, it is given by:

$$H(q([\tilde{\mathbf{h}}]_i | \mathbf{Y}), p([\tilde{\mathbf{h}}]_i)) \approx \log(2\pi\alpha_{\tilde{\mathbf{h}}}^2) + \frac{1}{D} \sum_{d=1}^D \frac{|\tilde{[\mathbf{h}]}_i^{(d)}|}{\alpha_{\tilde{\mathbf{h}}}}, \quad (16)$$

where the  $d$ -th sample is computed as  $\tilde{[\mathbf{h}]}_i^{(d)} = [\mathbf{m}_{\tilde{\mathbf{h}}}]_i + [\mathbf{b}_{\tilde{\mathbf{h}}}]_i \times \mathcal{CL}(0, 1)$ . Hence, the closed-form of  $\mathcal{L}_1$  is expressed as:

$$\mathcal{L}_1 = \frac{1}{D} \sum_{i=1}^N \sum_{d=1}^D \frac{|\tilde{[\mathbf{h}]}_i^{(d)}|}{\alpha_{\tilde{\mathbf{h}}}} - \sum_{i=1}^N \log(2\pi[\mathbf{b}]_i^2) + N \log(2\pi\alpha_{\tilde{\mathbf{h}}}^2) - 2N. \quad (17)$$

Similarly, we derive  $\mathcal{L}_2$  using the independence property of the elements  $[\tilde{\mathbf{G}}]_{i,j}$  for the prior and the posterior:

$$\begin{aligned} \mathcal{L}_2 &= \mathbb{E}_{\tilde{\mathbf{G}} \sim q(\tilde{\mathbf{G}} | \mathbf{Y})} [\log q(\tilde{\mathbf{G}} | \mathbf{Y})] - \mathbb{E}_{\tilde{\mathbf{G}} \sim q(\tilde{\mathbf{G}} | \mathbf{Y})} [\log p(\tilde{\mathbf{G}})] \\ &= \frac{1}{D} \sum_{i=1}^M \sum_{j=1}^N \sum_{d=1}^D \frac{|\tilde{[\mathbf{G}]}_{i,j}^{(d)}|}{\alpha_{\tilde{\mathbf{G}}}} - \sum_{i=1}^M \sum_{j=1}^N \log(2\pi[\mathbf{B}]_{i,j}^2) \\ &\quad + NM \log(2\pi\alpha_{\tilde{\mathbf{G}}}^2) - 2NM, \end{aligned} \quad (18)$$

where the Monte-Carlo samples are computed as  $\tilde{[\mathbf{G}]}_{i,j}^{(d)} = [\mathbf{M}_{\tilde{\mathbf{G}}}]_{i,j} + [\mathbf{B}_{\tilde{\mathbf{G}}}]_{i,j} \times \mathcal{CL}(0, 1)$ .

To compute  $\mathcal{L}_3$ , we perform the same steps as for the Gaussian channels (Eq. 9) with the signal model defined in Eq. 12. The covariance matrix of  $\tilde{\mathbf{h}}$  and the covariance matrix of the columns of  $\tilde{\mathbf{G}}$  are diagonal matrices with main diagonals denoted  $\boldsymbol{\lambda}$  and  $\boldsymbol{\tau}$ , respectively. These quantities are expressed as follows:

$$[\boldsymbol{\lambda}]_i = 6[\mathbf{b}_{\tilde{\mathbf{h}}}]_i^2; \quad [\boldsymbol{\tau}]_i = \sum_{m=1}^M 6[\mathbf{B}_{\tilde{\mathbf{G}}}]_{m,i}^2. \quad (19)$$

Hence,  $\mathcal{L}_3$  can be expressed as in Eq. 13 where  $C_2$  is a constant. A detailed derivation can be found in the Appendix.

#### IV. NUMERICAL RESULTS

In this section, we evaluate the performance of the proposed joint channel estimation method based on VI in RIS-aided SIMO system. We consider an RIS composed of  $N = 64$  elements and a BS equipped with  $M = 4$  antennas. We aim to estimate the channels by sending  $L = 50$  pilot symbols received at the BS. *Encoder 1* and *Encoder 2* are fully connected neural networks where each neural network consists of an input layer, two 300 unit hidden layers with Tanh 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 variance. Adam optimizer [16] is used to train the neural networks with 0.001 as the initial learning rate.

The primary evaluation metric is the capacity of the RIS-assisted network. Using the estimated channels that maximize the auxiliary posteriors obtained from the neural networks, we compute the phase shifts as  $\boldsymbol{\theta}^* = \angle(\mathbf{v} \odot \mathbf{h})$  where  $\mathbf{v}$  is the eigenvector of the highest eigenvalue of  $\mathbf{G}$ , and then compute the achieved capacity. We consider the following two baselines to compare our method with:

- **Exact capacity:** this is an upper bound where the capacity is obtained using the optimal phase shifts using the true channels  $\mathbf{G}$  and  $\mathbf{h}$ ;
- **Capacity with random phase shifts:** this represents a lower bound for our method.

Figures 2a and 2b depict the channel estimation performance with Rayleigh channel model with  $\sigma_{\mathbf{G}}^2 = \sigma_{\mathbf{h}}^2 = 1$ .

In Fig. 2a, we illustrate the capacity as a 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. 2b, we evaluate the normalized mean square error (NMSE) that is defined by  $\text{NMSE} = \|\hat{\mathbf{H}} - \mathbf{H}\|_F^2 / \|\mathbf{H}\|_F^2$  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. 3a, 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.  $\alpha_{\tilde{\mathbf{G}}} = 1$  and  $\alpha_{\tilde{\mathbf{h}}} = 1$ . 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. 3b shows that the NMSE decreases with the SNR. The performance with sparse channels is better than the performance we 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.

Furthermore, we benchmark our method with the Least Square (LS) estimator [17] in SISO setting, i.e.  $M = 1$ , with the complex Laplace channels. The LS method estimates the cascaded channel

$$\begin{aligned}
\mathcal{L}_3 = & \frac{1}{\sigma_w^2} \sum_{l=1}^L \left[ \left\| \mathbf{y}_l - \frac{\sqrt{\rho}}{MN^2} \mathbf{F}_M^H \mathbf{M}_{\bar{G}} \mathbf{F}_N^H \text{diag}(\mathbf{v}_l) \mathbf{F}_N^H \mathbf{m}_{\bar{h}} x_l \right\|^2 + \frac{\rho |x_l|^2}{MN^4} \cdot \text{tr}(\text{diag}(\boldsymbol{\tau}) \mathbf{F}_N^H \text{diag}(\mathbf{v}_l) \mathbf{F}_N^H \text{diag}(\boldsymbol{\lambda}) \mathbf{F}_N \text{diag}(\mathbf{v}_l)^H \mathbf{F}_N) \right. \\
& + \frac{\rho |x_l|^2}{MN^4} \text{tr}(\mathbf{M}_{\bar{G}}^H \mathbf{M}_{\bar{G}} \mathbf{F}_N^H \text{diag}(\mathbf{v}_l) \mathbf{F}_N^H \text{diag}(\boldsymbol{\lambda}) \mathbf{F}_N \text{diag}(\mathbf{v}_l)^H \mathbf{F}_N) + \frac{\rho |x_l|^2}{MN^4} \mathbf{m}_{\bar{h}}^H \mathbf{F}_N \text{diag}(\mathbf{v}_l)^H \mathbf{F}_N \text{diag}(\boldsymbol{\tau}) \mathbf{F}_N^H \text{diag}(\mathbf{v}_l) \mathbf{F}_N^H \mathbf{m}_{\bar{h}} \right] \\
& + C_2.
\end{aligned} \tag{13}$$

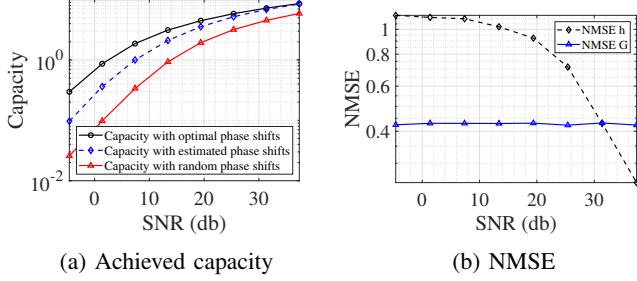


Fig. 2: Performance of the VI-based RIS joint channel estimation using complex Gaussian channels.

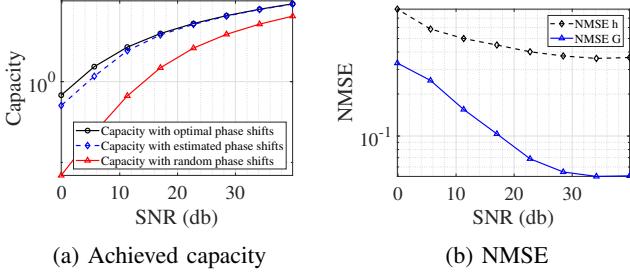


Fig. 3: Performance of the VI-based RIS joint channel estimation using complex Laplace channels.

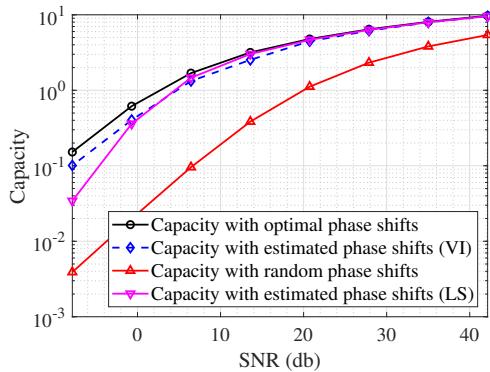


Fig. 4: Achieved capacity with complex Laplace channels and  $M = 1$ .

$\mathbf{H} = \mathbf{G} \cdot \text{diag}(\mathbf{h})$  by minimizing the mean squared error  $\|\mathbf{Y} - (\mathbf{x} \otimes \mathbf{I}_M) \mathbf{H} \mathbf{V}\|_F^2$  where  $\mathbf{V} \in \mathbb{C}^{N \times L}$  is the RIS configuration matrix with the  $l$ -th column corresponding to the phase-shifts of the RIS at the  $l$ -th pilot signal. The expression of LS estimator is given by [17]:

$$\hat{\mathbf{H}}_{\text{LS}} = \sqrt{\rho} \mathbf{Y} (\mathbf{V}^H \mathbf{V})^{-1} \mathbf{V}^H. \tag{20}$$

The optimal phase shifts can be obtained directly from the estimated

cascaded channel as  $\theta_k^* = -\angle g_k - \angle h_k$  where  $g_k$  is the channel link between the BS and the  $k$ -th element in the RIS and  $h_k$  is the channel link between the  $k$ -th RIS and the user. Fig. 4 shows that VI-based method outperforms the LS estimator in low SNR regime due to taking into consideration the sparse structure of the channels which improves the estimation of the channels. In high SNR, the LS estimator slightly outperforms the proposed method where both methods gets closer to the upper bound capacity.

## V. CONCLUSION

We have proposed a joint channel estimation framework for RIS-aided mmWave systems with fully passive elements. Unlike other works on channel estimation for RIS, our work only requires an uplink training pilot and considers the sparsity of the mmWave channels. Based on the variational inference framework, the joint channel estimation problem is solved by approximating the true posteriors. We have showcased that sampling from the approximated posteriors yields a similar capacity to the one achieved with the true posteriors. Several future directions can be adopted based on this work. For instance, a more physically-consistent RIS modeling, where the elements of the RIS suffer from mutual coupling which leads to a non-diagonal reflection matrix, can be studied. Furthermore, the robustness of the variational inference-based method on different channel priors and channel models should be investigated using real-world data.

## REFERENCES

- [1] 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.
- [2] 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.
- [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] Q. Wu and R. Zhang, "Intelligent reflecting surface enhanced wireless network via joint active and passive beamforming," *IEEE Transactions on Wireless Communications*, vol. 18, no. 11, pp. 5394–5409, 2019.
- [7] C. Huang, A. Zappone, G. C. Alexandropoulos, M. Debbah, and C. Yuen, "Reconfigurable intelligent surfaces for energy efficiency in wireless communication," *IEEE Transactions on Wireless Communications*, vol. 18, no. 8, pp. 4157–4170, 2019.
- [8] 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.
- [9] K. Ardash, S. Ghorekhloo, 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.
- [10] 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.

- [11] 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," *arXiv preprint arXiv:2206.06605*, 2022.
- [12] 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.
- [13] 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.
- [14] M. J. Wainwright, M. I. Jordan *et al.*, "Graphical models, exponential families, and variational inference," *Foundations and Trends® in Machine Learning*, vol. 1, no. 1–2, pp. 1–305, 2008.
- [15] 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.
- [16] D. P. Kingma and J. Ba, "Adam: A method for stochastic optimization," *arXiv preprint arXiv:1412.6980*, 2014.
- [17] M. Joham, H. Gao, and W. Utschick, "Estimation of channels in systems with intelligent reflecting surfaces," in *ICASSP 2022-2022 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. IEEE, 2022, pp. 5368–5372.

## APPENDIX

In this section, we give detailed derivation of the losses under the Gaussian and Laplace channels investigated. To compute  $\mathcal{L}_3$  (Eq. 9) for the Gaussian channels, we use the property  $\mathbb{E}_{\mathbf{x}}[(\mathbf{a} - \mathbf{x})^H(\mathbf{a} - \mathbf{x})] = \text{tr}(\mathbf{\Lambda}) + (\mathbf{a} - \mathbf{m}_x)^H(\mathbf{a} - \mathbf{m}_x)$  where  $\mathbf{m}_x$  is the mean of  $\mathbf{x}$  and  $\mathbf{\Lambda}$  is the covariance matrix of  $\mathbf{x}$  to compute the expectation over  $\mathbf{h}$ :

$$\begin{aligned} \mathcal{L}_3 &= \mathbb{E}_{\mathbf{h}, \mathbf{G} \sim q(\mathbf{h}, \mathbf{G} | \mathbf{Y})} \left[ \frac{1}{\sigma_w^2} \sum_{l=1}^L (\mathbf{y}_l - \sqrt{\rho} \mathbf{G} \text{diag}(\mathbf{v}_l) \mathbf{h} x_l)^H \right. \\ &\quad \left. (\mathbf{y}_l - \sqrt{\rho} \mathbf{G} \text{diag}(\mathbf{v}_l) \mathbf{h} x_l) \right] + C_1 \\ &= \frac{1}{\sigma_w^2} \sum_{l=1}^L \mathbb{E}_{\mathbf{G} \sim q(\mathbf{G} | \mathbf{Y})} \left[ \rho |x_l|^2 \text{tr}(\mathbf{G} \text{diag}(\mathbf{v}_l) \text{diag}(\gamma_h) \text{diag}(\mathbf{v}_l)^H \mathbf{G}^H) \right. \\ &\quad \left. + (\mathbf{y}_l - \sqrt{\rho} \mathbf{G} \text{diag}(\mathbf{v}_l) \mathbf{m}_h x_l)^H (\mathbf{y}_l - \sqrt{\rho} \mathbf{G} \text{diag}(\mathbf{v}_l) \mathbf{m}_h x_l) \right] + C_1 \\ &= \frac{1}{\sigma_w^2} \sum_{l=1}^L \mathbb{E}_{\mathbf{G} \sim q(\mathbf{G} | \mathbf{Y})} \left[ \rho |x_l|^2 \text{tr}(\mathbf{G}^H \mathbf{G} \text{diag}(\gamma_h)) \right. \\ &\quad \left. + (\mathbf{y}_l - \sqrt{\rho} \mathbf{G} \text{diag}(\mathbf{v}_l) \mathbf{m}_h x_l)^H (\mathbf{y}_l - \sqrt{\rho} \mathbf{G} \text{diag}(\mathbf{v}_l) \mathbf{m}_h x_l) \right] + C_1. \end{aligned} \quad (21)$$

We note that the covariance of  $\mathbf{h} \sim q(\mathbf{h} | \mathbf{Y})$  is defined by the diagonal matrix  $\text{diag}(\gamma_h)$  since the elements of  $\mathbf{h}$  are assumed to be independent. Furthermore, we note that  $\mathbb{E}_{\mathbf{G}}[\mathbf{G}^H \mathbf{G}] = \mathbf{V} + \mathbf{M}_G^H \mathbf{M}_G$  where  $\mathbf{V} = \mathbb{E}_{\mathbf{G}}[(\mathbf{G} - \mathbf{M}_G)^H(\mathbf{G} - \mathbf{M}_G)]$  is the covariance matrix over the columns of  $\mathbf{G}$ .  $\mathbf{V}$  is a diagonal matrix because the elements  $[\mathbf{G}]_{i,j}$  are assumed to be independent which makes the columns are independent as well and the elements on the main diagonal, denoted by  $\tau$ , are given by  $\tau_i = \sum_{m=1}^M [\Gamma_{\mathbf{G}}]_{m,i}$ . Therefore,  $\mathcal{L}_3$  can be expressed as follows:

$$\begin{aligned} \mathcal{L}_3 &= \frac{1}{\sigma_w^2} \left[ \rho ||\mathbf{x}||^2 \text{tr}(\text{diag}(\tau) \text{diag}(\gamma_h)) + \rho ||\mathbf{x}||^2 \mathbf{m}_h^H \text{diag}(\tau) \mathbf{m}_h \right. \\ &\quad \left. + \rho ||\mathbf{x}||^2 \text{tr}(\mathbf{M}_G^H \mathbf{M}_G \text{diag}(\gamma_h)) \right. \\ &\quad \left. + \sum_{l=1}^L ||\mathbf{y}_l - \sqrt{\rho} \mathbf{M}_G \text{diag}(\mathbf{v}_l) \mathbf{m}_h x_l||^2 \right] + C_1. \end{aligned} \quad (22)$$

Moreover, we derive the entropy  $H(q(z))$  in Eq. 23 of a complex Laplace random variable where  $q(z) \sim \mathcal{CL}(m, b)$  with mean  $m$  and scale  $b$ :

$$\begin{aligned} H(q(z)) &= \int_{\mathbb{C}} -q(z) \log q(z) dz \\ &= \int_{\mathbb{C}} -\frac{1}{2\pi b^2} e^{-\frac{|z-m|}{b}} \log \frac{1}{2\pi b^2} e^{-\frac{|z-m|}{b}} dz \\ &= \log(2\pi b^2) + \int_{\mathbb{C}} \frac{|u|}{2\pi b^3} e^{-\frac{|u|}{b}} du \quad (u = z - m) \\ &= \log(2\pi b^2) + 2. \end{aligned} \quad (23)$$

Finally, we derive the closed-form of  $\mathcal{L}_3$  (Eq. 13) with complex Laplace priors. We use the same steps as for Gaussian channels in Eq. 21 and Eq. 22 using the signal model defined in Eq. 12. In first step, we compute the expectation over  $\bar{\mathbf{h}}$  where we denote  $\mathbf{A} = \frac{\sqrt{\rho}}{MN^2} \mathbf{F}_M^H \bar{\mathbf{G}} \mathbf{F}_N^H \text{diag}(\mathbf{v}_l) \mathbf{F}_N^H x_l$  which is a constant with respect to  $\bar{\mathbf{h}}$ :

$$\begin{aligned} \mathcal{L}_3 &= \frac{1}{\sigma_w^2} \sum_{l=1}^L \mathbb{E}_{\bar{\mathbf{h}}, \bar{\mathbf{G}} \sim q(\bar{\mathbf{h}}, \bar{\mathbf{G}} | \mathbf{Y})} \left[ (\mathbf{y}_l - \mathbf{A} \bar{\mathbf{h}})^H (\mathbf{y}_l - \mathbf{A} \bar{\mathbf{h}}) \right] \\ &= \frac{1}{\sigma_w^2} \sum_{l=1}^L \mathbb{E}_{\bar{\mathbf{G}} \sim q(\bar{\mathbf{G}} | \mathbf{Y})} \left[ \text{tr}(\mathbf{A} \text{diag}(\lambda) \mathbf{A}^H) \right. \\ &\quad \left. + (\mathbf{y}_l - \mathbf{A} \mathbf{m}_{\bar{\mathbf{h}}})^H (\mathbf{y}_l - \mathbf{A} \mathbf{m}_{\bar{\mathbf{h}}}) \right] + C_2, \end{aligned} \quad (24)$$

where  $C_2$  is a constant and  $\lambda = [\text{Var}([\bar{\mathbf{h}}]_1), \dots, \text{Var}([\bar{\mathbf{h}}]_N)]$  is the covariance matrix of  $\bar{\mathbf{h}}$ . To compute  $\bar{\mathbf{G}}$ , we define a constant matrix  $\mathbf{C} = \frac{\sqrt{\rho}}{MN^2} \mathbf{F}_N^H \text{diag}(\mathbf{v}_l) \mathbf{F}_N^H x_l$ , i.e.  $\mathbf{A} = \mathbf{F}_M^H \bar{\mathbf{G}} \mathbf{C}$ . Hence, we get:

$$\begin{aligned} \mathcal{L}_3 &= \frac{1}{\sigma_w^2} \sum_{l=1}^L \mathbb{E}_{\bar{\mathbf{G}} \sim q(\bar{\mathbf{G}} | \mathbf{Y})} \left[ \text{tr}(\mathbf{A}^H \mathbf{A} \text{diag}(\lambda)) \right. \\ &\quad \left. + (\mathbf{y}_l - \mathbf{A} \mathbf{m}_{\bar{\mathbf{h}}})^H (\mathbf{y}_l - \mathbf{A} \mathbf{m}_{\bar{\mathbf{h}}}) \right] + C_2 \\ &= \frac{1}{\sigma_w^2} \sum_{l=1}^L \mathbb{E}_{\bar{\mathbf{G}} \sim q(\bar{\mathbf{G}} | \mathbf{Y})} \left[ \text{Mtr}(\mathbf{C}^H \bar{\mathbf{G}}^H \bar{\mathbf{G}} \mathbf{C} \text{diag}(\lambda)) \right. \\ &\quad \left. + (\mathbf{y}_l - \mathbf{F}_M^H \bar{\mathbf{G}} \mathbf{C} \mathbf{m}_{\bar{\mathbf{h}}})^H (\mathbf{y}_l - \mathbf{F}_M^H \bar{\mathbf{G}} \mathbf{C} \mathbf{m}_{\bar{\mathbf{h}}}) \right] + C_2. \end{aligned} \quad (25)$$

Then we use the property  $\mathbb{E}_{\mathbf{G}}[\mathbf{G}^H \mathbf{G}] = \mathbf{V} + \mathbf{M}_G^H \mathbf{M}_G$  where  $\mathbf{V} = \mathbb{E}_{\mathbf{G}}[(\mathbf{G} - \mathbf{M}_G)^H(\mathbf{G} - \mathbf{M}_G)]$  to compute the closed-form of the expectation over  $\bar{\mathbf{G}}$ .  $\mathbf{V}$  is a diagonal matrix with  $\tau$  as the main diagonal defined similarly as expressed for Gaussian case. Therefore, we have:

$$\begin{aligned} \mathcal{L}_3 &= \frac{1}{\sigma_w^2} \sum_{l=1}^L \left[ \text{Mtr}(\mathbf{C}^H \text{diag}(\tau) \mathbf{C} \text{diag}(\lambda)) \right. \\ &\quad \left. + \text{Mtr}(\mathbf{C}^H \mathbf{M}_{\bar{\mathbf{G}}}^H \mathbf{M}_{\bar{\mathbf{G}}} \mathbf{C} \text{diag}(\lambda)) \right. \\ &\quad \left. + \mathbf{M} \mathbf{m}_{\bar{\mathbf{h}}}^H \mathbf{C}^H \text{diag}(\tau) \mathbf{C} \mathbf{m}_{\bar{\mathbf{h}}} \right. \\ &\quad \left. + (\mathbf{y}_l - \mathbf{F}_M^H \mathbf{M}_{\bar{\mathbf{G}}} \mathbf{C} \mathbf{m}_{\bar{\mathbf{h}}})^H (\mathbf{y}_l - \mathbf{F}_M^H \mathbf{M}_{\bar{\mathbf{G}}} \mathbf{C} \mathbf{m}_{\bar{\mathbf{h}}}) \right]. \end{aligned} \quad (26)$$

Hence, we obtain the closed-form expression of  $\mathcal{L}_3$  defined in Eq. 13 with  $\text{diag}(\tau)$  the covariance matrix of the columns of  $\mathbf{G}$  and  $\text{diag}(\lambda)$  the covariance matrix of  $\mathbf{h}$ . The variance of a complex Laplace random variable  $z \sim \mathcal{CL}(m, b)$  is  $\text{Var}(z) = 6b^2$ .