Message-Passing Based Channel Estimation for Reconfigurable Intelligent Surface Assisted MIMO

Hang Liu*, Xiaojun Yuan† and Ying-Jun Angela Zhang*
*Department of Information Engineering, The Chinese University of Hong Kong, Shatin, N.T., Hong Kong SAR
†Center for Intelligent Networking and Communications,
University of Electronic Science and Technology of China, Chengdu, China
Email: lh117@ie.cuhk.edu.hk, xjyuan@uestc.edu.cn, yjzhang@ie.cuhk.edu.hk

***Abstract*—In this paper, we study the channel acquisition problem in a reconfigurable intelligent surface (RIS) assisted multiuser multiple-input multiple-output (MIMO) system, where an RIS with fully passive phase-shift elements is deployed to assist the MIMO communication. The state-of-the-art channel acquisition approach in such a system estimates the cascaded transmitter-to-RIS and RIS-to-receiver channels by adopting excessively long training sequences. To estimate the cascaded channels with an affordable training overhead, we formulate the channel estimation problem as a matrix-calibration based matrix factorization task. By exploiting the information on the slow-varying channel components and the hidden channel sparsity, we propose a novel message-passing based algorithm to factorize the cascaded channels.**

I. Introduction

Reconfigurable intelligent surface (RIS) assisted multiple-input multiple-output (MIMO) [1]–[4] is deemed very promising to realize similar or even higher array gains with significant cost reduction compared with massive MIMO. However, most of the existing work on the RIS design assumes perfect channel state information (CSI) to optimize the RIS parameters without considering its acquisition difficulty. In fact, channel estimation in an RIS-assisted wireless system is much more challenging than in a conventional system. This is because the passive RIS elements are incapable of sensing and estimating channel information. Therefore, we shall rely on the receiver to estimate both the transmitter-to-RIS and RIS-to-receiver channels by observing only a noisy cascade of the two channels.

To address the channel estimation challenge in RIS-assisted communication systems, some pioneering work has recently emerged. For example, Ref. [5] assumes that a portion of RIS elements are actively connected to signal processing units to perform channel estimation, so that the channels of the remaining passive RIS elements can be inferred via a compressed-sensing based approach. Compared with active RIS elements, purely passive RIS elements are undoubtedly more appealing due to their extremely low hardware and deployment costs. Ref. [6] shows that channel estimation in passive-RIS-assisted systems can be converted into a sequence of conventional MIMO channel estimation problems by turning on one RIS element at a time. However, the training overhead of this method is proportional to the size of the RIS and may be prohibitively large as the RIS typically comprises a large number of elements. To relieve the overwhelming training burden, more advanced channel estimation approaches are proposed in [7]–[10]. In [7], the authors develop a cascaded channel estimation algorithm for an RIS-assisted *single-user* MIMO system. In [8], the authors sequentially estimate the cascaded channels for users. Since users share the same RIS-to-receiver channel, by exploiting such channel correlations among users, the required training overhead is largely reduced. In [9], the authors exploit the sparsity of the transmitter-to-RIS-to-receiver channels and estimate the cascaded channels based on compressed sensing. In [10], the authors employ parallel factor decomposition to alternatively estimate the cascaded channels.

In this paper, we consider the cascaded channel estimation problem for an RIS-assisted *multiuser* MIMO system, where a fully passive RIS is used to assist the communication. In practice, the RIS can be coated onto a wall, a ceiling, or a furniture; or mounted on a building facade, an advertising panel, or a highway fence. In a typical application scenario, both the BS and the RIS rarely move after deployment. As a result, the channel between the BS and the RIS can be modelled as a quasi-static end-to-end MIMO channel, in which most of the channel components evolve much more slowly compared with conventional mobile communication channels. Meanwhile, a small portion of the channel components of the BS-to-RIS channel may experience sudden changes. For example, an opening/closing of a door in an indoor scenario or a moving car in an outdoor scenario may change the scattering geometry. By modeling both the fast and slow varying channel components, we formulate the CSI acquisition problem as a *matrix-calibration-based matrix factorization* task. Then, we propose a novel message-passing based algorithm to effectively estimate the two cascaded channels. Furthermore, we introduce additional approximations to the messages based on the approximate message passing (AMP) framework [11]. The proposed algorithm only needs to update the means and variances of the messages and hence avoids high-dimensional

This work was supported in part by General Research Funding (Project number 14208017) from the Research Grants Council of Hong Kong, in part by the National Key R&D Program of China under Grant 2018YFB1801105, and in part by the 111 Project of China under Grant B20064.

978-1-7281-6432-8/20/$31.00 ©2020 IEEE

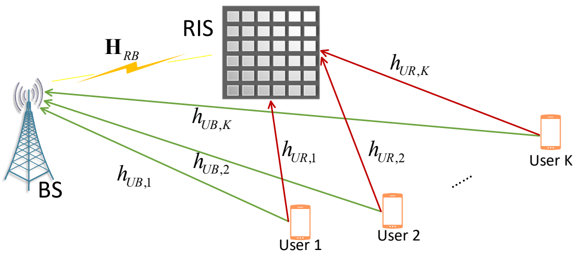

Fig. 1. An RIS-assisted multiuser system.

integrations in the canonical message passing algorithm.

II. System Model and Problem Formulation

A. RIS-Assisted Multiuser MIMO

Consider a single-cell RIS-assisted multiuser MIMO system depicted in Fig. 1. Assume that K single-antenna users simultaneously communicate with an M-antenna BS. An RIS comprising L phase-shift elements is deployed to assist the communication between the users and the BS. A uniform linear array (ULA) is adopted at the BS, and the passive reflecting elements in the RIS are arranged in the form of an $L_1 \times L_2$ uniform rectangular array (URA) with $L_1 L_2 = L$. Denote by $\mathbf{h}_{UB,k} \in \mathbb{C}^{M\times 1}$, $\mathbf{h}_{UR,k} \in \mathbb{C}^{L\times 1}$, and $\mathbf{H}_{RB} \in \mathbb{C}^{M\times L}$ the coefficient vectors/matrix of the k-th-user-to-BS, the k-th-user-to-RIS, and the RIS-to-BS channels, respectively. We assume that the RIS elements induce independent phase shifts on the incident signals. Denote the RIS phase-shift vector at time t by $\boldsymbol{\psi}(t) \triangleq [e^{j\psi_1(t)}, \cdots, e^{j\psi_L(t)}]^T$, where $\psi_l(t) \in [0, 2\pi)$ represents the phase shift of the l-th RIS element.

We assume a quasi-static flat fading channel model, where the channel coefficients remain invariant within the coherence time. We model $\mathbf{H}_{RB}$ by the MIMO Rician fading model as [12]

$$\mathbf{H}_{RB} = \sqrt{\frac{\kappa}{\kappa+1}}\bar{\mathbf{H}}_{RB} + \sqrt{\frac{1}{\kappa+1}}\widetilde{\mathbf{H}}_{RB}, \tag{1}$$

where $\bar{\mathbf{H}}_{RB}$ (or $\widetilde{\mathbf{H}}_{RB}$) represents the slow-varying (or fast-varying) channel component matrix; and κ is the Rician factor denoting the power ratio between the two components.

Moreover, the slow-varying component matrix $\bar{\mathbf{H}}_{RB}$ sums up the paths from all the deterministic scattering clusters and can be modeled as

$$\bar{\mathbf{H}}_{RB} = \sqrt{\beta_0}\sum_{p=1}^{\bar{P}_{RB}} \alpha_p \mathbf{a}_B(\theta_p)\mathbf{a}_R^H(\phi_p, \sigma_p), \tag{2}$$

where β_0 is the large-scale path gain; $\bar{P}_{RB}$ is the number of the slow-varying paths; α_p is the corresponding channel coefficient; θ_p is the corresponding azimuth angle-of-arrival (AoA) at the BS; ϕ_p (or σ_p) is the corresponding azimuth (or elevation) angle-of-departure (AoD) at the RIS; and $\mathbf{a}_B$ (or $\mathbf{a}_R$) is the steering vector associated with the BS (or RIS) antenna geometry. In specific,

$$\mathbf{a}_B(\theta) = \mathbf{f}_M(\sin\theta), \tag{3}$$

$$\mathbf{a}_R(\phi,\sigma) = \mathbf{f}_{L_2}(-\cos\sigma\cos\phi) \otimes \mathbf{f}_{L_1}(\cos\sigma\sin\phi), \tag{4}$$

where the transform vector is defined as $\mathbf{f}_M(x) \triangleq [1, \cdots, e^{-j\frac{2\pi}{\varrho}d(M-1)x}]^T/\sqrt{M}$; ϱ denotes the carrier wavelength; d denotes the distance between any two adjacent antennas; and $\otimes$ denotes the Kronecker product.

Meanwhile, we represent $\widetilde{\mathbf{H}}_{RB}$ and $\mathbf{h}_{UR,k}$ as

$$\widetilde{\mathbf{H}}_{RB} = \sqrt{\beta_0}\sum_{p=1}^{\widetilde{P}_{RB}} \alpha_p \mathbf{a}_B(\theta_p)\mathbf{a}_R^H(\phi_p, \sigma_p), \tag{5}$$

$$\mathbf{h}_{UR,k} = \sqrt{\beta_k}\sum_{p=1}^{P_k} \alpha_p \mathbf{a}_R(\phi_p, \sigma_p), \tag{6}$$

where $\widetilde{P}_{RB}$ is the number of the fast-varying paths; β_k is the large-scale path gain for the channel between the k-th user and the RIS; and P_k is the number of paths between the k-th user and the RIS.

We assume that the slow-varying component matrix $\bar{\mathbf{H}}_{RB}$ in (2) keeps static over a time interval much larger than the coherence block length. As a consequence, $\bar{\mathbf{H}}_{RB}$ can be accurately estimated by long-term channel averaging prior to the RIS channel estimation procedure. Without loss of generality, we assume that $\mathbb{E}[\|\bar{\mathbf{H}}_{RB}\|_F^2] = \mathbb{E}[\|\widetilde{\mathbf{H}}_{RB}\|_F^2] = \beta_0 ML$ and $\mathbb{E}[\|\mathbf{h}_{UR,k}\|_2^2] = \beta_k L, \forall k$.

Following [13], we employ a pre-discretized sampling grid $\boldsymbol{\vartheta}$ with length M' $(\geq M)$ to discretize $\{\sin(\theta_p) : 1 \leq p \leq \widetilde{P}_{RB}\}$ over $[0, 1]$. Similarly, we employ two sampling grids $\boldsymbol{\varphi}$ with length L_1' $(\geq L_1)$ and $\boldsymbol{\varsigma}$ with length L_2' $(\geq L_2)$ to discretize $\{\cos(\sigma_p)\sin(\phi_p)\}$ and $\{-\cos(\sigma_p)\cos(\phi_p)\}$, respectively. Then, we represent $\widetilde{\mathbf{H}}_{RB}$ and $\{\mathbf{h}_{UR,k}\}$ under the angular bases as [13]

$$\sqrt{\frac{1}{\kappa+1}}\widetilde{\mathbf{H}}_{RB} = \mathbf{A}_B\mathbf{S}\mathbf{A}_R^H, \tag{7}$$

$$\mathbf{h}_{UR,k} = \mathbf{A}_R\mathbf{g}_k, \tag{8}$$

where $\mathbf{A}_B \triangleq [\mathbf{f}_M(\vartheta_1), \cdots, \mathbf{f}_M(\vartheta_{M'})]$ is an over-complete ULA array response; $\mathbf{A}_R \triangleq [\mathbf{f}_{L_2}(\varsigma_1), \cdots, \mathbf{f}_{L_2}(\varsigma_{L_2'})] \otimes [\mathbf{f}_{L_1}(\varphi_1), \cdots, \mathbf{f}_{L_1}(\varphi_{L_1'})]$ is an over-complete URA array response; and $\mathbf{S}$ (or $\{\mathbf{g}_k\}$) is the corresponding channel coefficient matrix (or vectors) in the angular domain with $L' = L_1'L_2'$.

We assume that the fast-varying component matrix $\widetilde{\mathbf{H}}_{RB}$ contains a limited number of paths, i.e., $\widetilde{P}_{RB}$ in (5) is small. As a result, only a few entries of $\mathbf{S}$ are nonzero with each corresponding to a channel path. That is, $\mathbf{S}$ is a sparse matrix. Moreover, experimental studies have shown that the propagation channel often exhibits limited scattering geometry [14]. As a consequence, $\{\mathbf{g}_k\}$ are sparse vectors as well. We note that the sparsity of $\mathbf{S}$ and $\{\mathbf{g}_k\}$ plays an important role in our channel estimation design.

B. Cascaded Channel Estimation

To facilitate RIS channel estimation, users simultaneously transmit training sequences with length T to the BS. Denote by $\mathbf{x}_k = [x_{k1}, \cdots, x_{kT}]^T$ the training sequence of user k, where x_{kt} is the training symbol of user k in time slot t. We assume that the users transmit at constant power τ_X, i.e., $\mathbb{E}[|x_{kt}|^2] = \tau_X, \forall k, t$. Over the time duration T, all the RIS elements are turned on and are set to have the same phase shift. Without loss of generality, we assume that $\boldsymbol{\psi}(t) = \mathbf{1}, 1 \leq t \leq T$. The received signal at the BS in time slot t is given by

$$\mathbf{y}_0(t) = \sum_{k=1}^{K} (\mathbf{h}_{UB,k} + \mathbf{H}_{RB}\mathbf{h}_{UR,k}) x_{kt} + \mathbf{n}(t), \tag{9}$$

where $\mathbf{n}(t)$ is an additive white Gaussian noise (AWGN) vector following the distribution of $\mathcal{CN}(\mathbf{0}, \tau_N \mathbf{I})$. We assume that all the direct channels $\{\mathbf{h}_{UB,k}\}$ are accurately estimated before the RIS channel estimation procedure.[1] Collecting all the received signals in time duration T and canceling the direct channels, we rewrite (9) in a matrix form as

$$\mathbf{Y} = \mathbf{H}_{RB}\mathbf{H}_{UR}\mathbf{X} + \mathbf{N}, \tag{10}$$

where $\mathbf{X} = [\mathbf{x}_1, \cdots, \mathbf{x}_K]^T$; $\mathbf{Y} = [\mathbf{y}_0(1), \cdots, \mathbf{y}_0(T)] - [\mathbf{h}_{UB,1}, \cdots, \mathbf{h}_{UB,K}]\mathbf{X}$; $\mathbf{H}_{UR} = [\mathbf{h}_{UR,1}, \cdots, \mathbf{h}_{UR,K}]$; and $\mathbf{N} = [\mathbf{n}(1), \cdots, \mathbf{n}(T)]$. From (1), (7), and (8), we obtain the system model as

$$\begin{aligned}\mathbf{Y} &= \left(\sqrt{\frac{\kappa}{\kappa+1}} \bar{\mathbf{H}}_{RB} + \mathbf{A}_B \mathbf{S} \mathbf{A}_R^H \right) \mathbf{A}_R \mathbf{G} \mathbf{X} + \mathbf{N} \\ &= (\mathbf{H}_0 + \mathbf{A}_B \mathbf{S} \mathbf{R}) \mathbf{G} \mathbf{X} + \mathbf{N}, \end{aligned} \tag{11}$$

where $\mathbf{G} \triangleq [\mathbf{g}_1, \cdots, \mathbf{g}_K] \in \mathbb{C}^{L' \times K}$; $\mathbf{H}_0 \triangleq \sqrt{\kappa/(\kappa+1)} \bar{\mathbf{H}}_{RB} \mathbf{A}_R \in \mathbb{C}^{M \times L'}$; and $\mathbf{R} \triangleq \mathbf{A}_R^H \mathbf{A}_R \in \mathbb{C}^{L' \times L'}$.

Upon the reception of $\mathbf{Y}$ in (11), the BS aims to factorize the channel matrices $\mathbf{S}$ and $\mathbf{G}$ with the knowledge of the training signal matrix $\mathbf{X}$. Once the angular bases $\mathbf{A}_B$ and $\mathbf{A}_R$ are predetermined and the slow-varying component matrix $\bar{\mathbf{H}}_{RB}$ is given, the sensing matrices $\mathbf{A}_B$, $\mathbf{H}_0$, and $\mathbf{R}$ are also known to the BS. We refer to the above problem as *matrix-calibration* based cascaded channel estimation.

III. Matrix-Calibration Based Cascaded Channel Estimation Algorithm

A. Bayesian Inference

Define $\mathbf{W} \triangleq \mathbf{H}_0 + \mathbf{A}_B \mathbf{S} \mathbf{R}$, $\mathbf{Z} \triangleq \mathbf{W}\mathbf{G}$, and $\mathbf{Q} \triangleq \mathbf{Z}\mathbf{X}$. Under the assumption of AWGN, we have

$$p(\mathbf{Y}|\mathbf{Q}) = \prod_{m=1}^{M} \prod_{t=1}^{T} \mathcal{CN}\left(y_{mt}; q_{mt}, \tau_N\right). \tag{12}$$

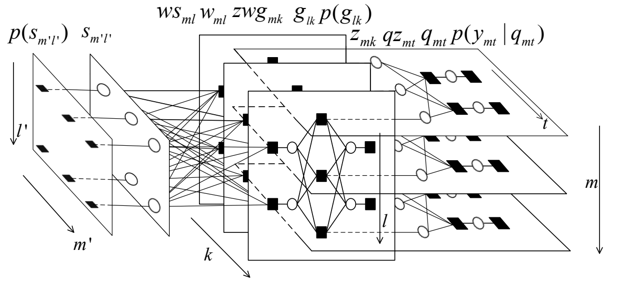
ws_{ml} w_{ml} zwg_{mk} g_{lk} $p(g_{lk})$
$p(s_{m'l'})$ $s_{m'l'}$
z_{mk} qz_{mt} q_{mt} $p(y_{mt} | q_{mt})$
l'
m'
k
l
t
m

Fig. 2. An illustration of the factor graph representation for $M = M' = K = 3$ and $T = L' = 2$, where blank circles and black squares represent variable nodes and factor nodes, respectively.

Motivated by the sparsity of $\mathbf{S}$ and $\mathbf{G}$, we employ Bernoulli-Gaussian distributions to model their prior distributions as

$$p(\mathbf{S}) = \prod_{m'=1}^{M'} \prod_{l'=1}^{L'} (1-\lambda_S)\delta(s_{m'l'}) + \lambda_S \mathcal{CN}\left(s_{m'l'}; 0, \tau_S\right), \tag{13}$$

$$p(\mathbf{G}) = \prod_{l=1}^{L'} \prod_{k=1}^{K} (1-\lambda_G)\delta(g_{lk}) + \lambda_G \mathcal{CN}\left(g_{lk}; 0, \tau_G\right), \tag{14}$$

where λ_S (or λ_G) is the corresponding Bernoulli parameter of $\mathbf{S}$ (or $\mathbf{G}$); and τ_S (or τ_G) is the variance of the nonzero entries of $\mathbf{S}$ (or $\mathbf{G}$). From Bayes' theorem, the posterior distribution $p(\mathbf{S}, \mathbf{G}|\mathbf{Y})$ is given by

$$p(\mathbf{S}, \mathbf{G}|\mathbf{Y}) = \frac{1}{p(\mathbf{Y})} p(\mathbf{Y}|\mathbf{S}, \mathbf{G}) p(\mathbf{S}) p(\mathbf{G}). \tag{15}$$

With (15), the posterior mean estimators of $\mathbf{S}$ and $\mathbf{G}$ are given by $\hat{\mathbf{S}} = [\hat{s}_{m'l'}]$ and $\hat{\mathbf{G}} = [\hat{g}_{lk}]$, where

$$\hat{s}_{m'l'} = \int s_{m'l'} p(s_{m'l'}|\mathbf{Y}) \mathrm{d}s_{m'l'}, \hat{g}_{lk} = \int g_{lk} p(g_{lk}|\mathbf{Y}) \mathrm{d}g_{lk}. \tag{16}$$

In the above, $p(s_{m'l'}|\mathbf{Y}) = \int\int p(\mathbf{S}, \mathbf{G}|\mathbf{Y}) \mathrm{d}\mathbf{G} \mathrm{d}(\mathbf{S} \setminus s_{m'l'})$ and $p(g_{lk}|\mathbf{Y}) = \int\int p(\mathbf{S}, \mathbf{G}|\mathbf{Y}) \mathrm{d}\mathbf{S} \mathrm{d}(\mathbf{G} \setminus g_{lk})$ are the marginal distributions with respect to $s_{m'l'}$ and g_{lk}, respectively, where $\mathbf{X} \setminus x_{ij}$ means the collection of all the elements of matrix $\mathbf{X}$ except for the (i,j)-th one. The posterior mean estimators in (16) achieve the minimum mean square errors (MMSEs) defined as $\mathrm{MMSE}_{\mathbf{S}} = \frac{1}{M'L'} \mathbb{E}\left[\|\mathbf{S} - \hat{\mathbf{S}}\|_F^2 \right]$ and $\mathrm{MMSE}_{\mathbf{G}} = \frac{1}{L'K} \mathbb{E}\left[\|\mathbf{G} - \hat{\mathbf{G}}\|_F^2 \right]$, where the expectations are taken over the joint distribution of $\mathbf{S}$, $\mathbf{G}$, and $\mathbf{Y}$.

Exact evaluation of $\hat{\mathbf{S}}$ and $\hat{\mathbf{G}}$ are generally intractable due to the high-dimensional integrations involved in the marginalization. In the following, we provide an approximate solution by following the message passing principle.

B. Message Passing for Marginal Posterior Computation

Plugging (12)–(14) into (15), we obtain

$$\begin{aligned} p(\mathbf{S}, \mathbf{G}|\mathbf{Y}) \propto\, & p(\mathbf{Y}|\mathbf{Q}) \delta\left(\mathbf{Q} - \mathbf{Z}\mathbf{X}\right) \delta\left(\mathbf{Z} - \mathbf{W}\mathbf{G}\right) \\ & \delta\left(\mathbf{W} - \mathbf{H}_0 - \mathbf{A}_B \mathbf{S} \mathbf{R}\right) p(\mathbf{S}) p(\mathbf{G}), \end{aligned} \tag{17}$$

[1] By turning off the RIS reflecting elements, the estimation of $\{\mathbf{h}_{UB,k}\}$ can be done by using conventional channel estimation methods for multiuser MIMO systems.

TABLE I
NOTATION OF FACTOR NODES

Factor	Distribution
$p(s_{m'l'})$	$(1-\lambda_S)\delta(s_{m'l'})+\lambda_S\mathcal{CN}\left(s_{m'l'};0,\tau_S\right)$
$p(g_{lk})$	$(1-\lambda_G)\delta(g_{lk})+\lambda_G\mathcal{CN}\left(g_{lk};0,\tau_G\right)$
ws_{ml}	$\delta(w_{ml}-h_{0,ml}-\sum_{m',l'}a_{B,mm'}s_{m'l'}r_{l'l})$
zwg_{mk}	$\delta(z_{mk}-\sum_{l=1}^{L'}w_{ml}g_{lk})$
qz_{mt}	$\delta(q_{mt}-\sum_{k=1}^{K}z_{mk}x_{kt})$
$p(y_{mt}\|q_{mt})$	$\mathcal{CN}(y_{mt};q_{mt},\tau_N)$

where $\delta(\cdot)$ denotes the Dirac delta function applied to the argument in a component-wise manner. We construct a factor graph to represent (17) and apply the canonical message passing algorithm to approximately compute the estimators in (16). The factor graph is depicted in Fig. 2. The variables $\mathbf{S}$, $\mathbf{G}$, $\mathbf{W}$, $\mathbf{Z}$, and $\mathbf{Q}$ are represented by the variable nodes $\{s_{m'l'}\}$, $\{g_{lk}\}$, $\{w_{ml}\}$, $\{z_{mk}\}$, and $\{q_{mt}\}$, respectively. The factorizable functions in (17), represented by factor nodes $\{p(s_{m'l'})\}$, $\{p(g_{lk})\}$, $\{ws_{ml}\}$, $\{zwg_{mk}\}$, $\{qz_{mt}\}$, and $\{p(y_{mt}|q_{mt})\}$, are connected to their associated arguments. We summarize the notation of the factor nodes in Table I. Denote by $\Delta^i_{a\to b}(\cdot)$ the message from node a to b in iteration i, and by $\Delta^i_c(\cdot)$ the marginal message computed at variable node c in iteration i. Applying the sum-product rule, we obtain the following messages:

1) Messages between $\{qz_{mt}\}$ and $\{z_{mk}\}$:

$$\Delta^i_{qz_{mt}\to z_{mk}}(z_{mk})\propto\int\prod_{j\neq k}\left(\Delta^i_{z_{mj}\to qz_{mt}}(z_{mj})\mathrm{d}z_{mj}\right)p(y_{mt}|q_{mt}), \tag{18}$$

$$\Delta^{i+1}_{z_{mk}\to qz_{mt}}(z_{mk})\propto\mathcal{P}^i_{z_{mk}}(z_{mk})\prod_{j\neq t}\Delta^i_{qz_{mj}\to z_{mk}}(z_{mk}), \tag{19}$$

where $\mathcal{P}^i_{z_{mk}}(z_{mk})$ is defined as

$$\mathcal{P}^i_{z_{mk}}(z_{mk})\propto\int zwg_{mk}\prod_{l=1}^{L'}\left(\Delta^i_{w_{ml}\to zwg_{mk}}(w_{ml})\Delta^i_{g_{lk}\to zwg_{mk}}(g_{lk})\mathrm{d}w_{ml}\mathrm{d}g_{lk}\right). \tag{20}$$

2) Messages between $\{g_{lk}\}$ and $\{zwg_{mk}\}$:

$$\Delta^i_{zwg_{mk}\to g_{lk}}(g_{lk})\propto\int\prod_{t=1}^{T}\Delta^i_{qz_{mt}\to z_{mk}}(z_{mk})zwg_{mk}\mathrm{d}z_{mk}\prod_{j\neq l}\left(\Delta^i_{g_{jk}\to zwg_{mk}}(g_{jk})\mathrm{d}g_{jk}\right)\prod_{l=1}^{L'}\left(\Delta^i_{w_{ml}\to zwg_{mk}}(w_{ml})\mathrm{d}w_{ml}\right), \tag{21}$$

$$\Delta^{i+1}_{g_{lk}\to zwg_{mk}}(g_{lk})\propto p(g_{lk})\prod_{j\neq m}\Delta^i_{zwg_{jk}\to g_{lk}}(g_{lk}). \tag{22}$$

3) Messages between $\{w_{ml}\}$ and $\{zwg_{mk}\}$:

$$\Delta^i_{zwg_{mk}\to w_{ml}}(w_{ml})\propto\int\prod_{t=1}^{T}\Delta^i_{qz_{mt}\to z_{mk}}(z_{mk})zwg_{mk}\mathrm{d}z_{mk}\prod_{l=1}^{L'}\left(\Delta^i_{g_{lk}\to zwg_{mk}}(g_{lk})\mathrm{d}g_{lk}\right)\prod_{j\neq l}\left(\Delta^i_{w_{mj}\to zwg_{mk}}(w_{mj})\mathrm{d}w_{mj}\right), \tag{23}$$

$$\Delta^{i+1}_{w_{ml}\to zwg_{mk}}(w_{ml})\propto\mathcal{P}^i_{w_{ml}}(w_{ml})\prod_{j\neq k}\Delta^i_{zwg_{mj}\to w_{ml}}(w_{ml}), \tag{24}$$

where

$$\mathcal{P}^i_{w_{ml}}(w_{ml})\propto\int ws_{ml}\prod_{m'=1}^{M'}\prod_{l'=1}^{L'}\left(\Delta^i_{s_{m'l'}\to ws_{ml}}(s_{m'l'})\mathrm{d}s_{m'l'}\right). \tag{25}$$

4) Messages between $\{s_{m'l'}\}$ and $\{ws_{ml}\}$:

$$\Delta^i_{ws_{ml}\to s_{m'l'}}(s_{m'l'})\propto\int\prod_{k=1}^{K}\Delta^i_{zwg_{mk}\to w_{ml}}(w_{ml})ws_{ml}\mathrm{d}w_{ml}\prod_{(j,n)\neq(m',l')}\left(\Delta^i_{s_{jn}\to ws_{ml}}(s_{jn})\mathrm{d}s_{jn}\right), \tag{26}$$

$$\Delta^{i+1}_{s_{m'l'}\to ws_{ml}}(s_{m'l'})\propto p(s_{m'l'})\prod_{(j,n)\neq(m,l)}\Delta^i_{ws_{jn}\to s_{m'l'}}(s_{m'l'}). \tag{27}$$

5) Marginal messages at variable nodes:

$$\Delta^{i+1}_{z_{mk}}(z_{mk})\propto\mathcal{P}^i_{z_{mk}}(z_{mk})\prod_{t=1}^{T}\Delta^i_{qz_{mt}\to z_{mk}}(z_{mk}), \tag{28}$$

$$\Delta^{i+1}_{w_{ml}}(w_{ml})\propto\mathcal{P}^i_{w_{ml}}(w_{ml})\prod_{k=1}^{K}\Delta^i_{zwg_{mk}\to w_{ml}}(w_{ml}), \tag{29}$$

$$\Delta^{i+1}_{g_{lk}}(g_{lk})\propto p(g_{lk})\prod_{m=1}^{M}\Delta^i_{zwg_{mk}\to g_{lk}}(g_{lk}), \tag{30}$$

$$\Delta^{i+1}_{s_{m'l'}}(s_{m'l'})\propto p(s_{m'l'})\prod_{m=1}^{M}\prod_{l=1}^{L'}\Delta^i_{ws_{ml}\to s_{m'l'}}(s_{m'l'}). \tag{31}$$

C. Approximations for Message Passing

The messages and marginals in (18)–(31) are computationally intractable in general due to the high-dimensional integrations and normalizations therein. To tackle this, we simplify the calculation of (18)–(31) by following the idea of AMP [11] in the large-system limit, i.e., $M,M',K,L,L',T,\tau_N\to\infty$ with the ratios M/K, M'/K, L/K, L'/K, T/K, and τ_N/K^2 fixed. The detailed derivations are omitted here due to space limitation and can be found in the extended version of this work from ArXiv [15, Sec. IV-C].

IV. Numerical Results

In this section, we conduct simulations to investigate the performance of the proposed algorithm. We generate $\bar{\mathbf{H}}_{RB}$ by (2) with 20 clusters of paths and 10 subpaths per cluster.

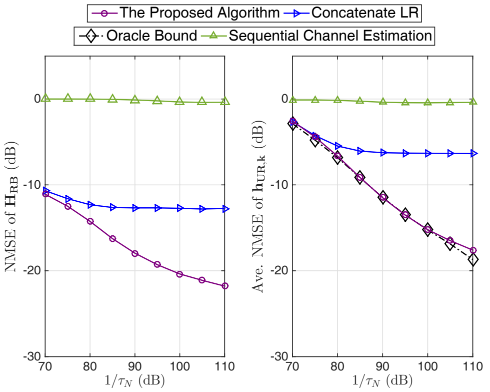

Fig. 3. The NMSE performance versus noise power with $\eta = 2$.

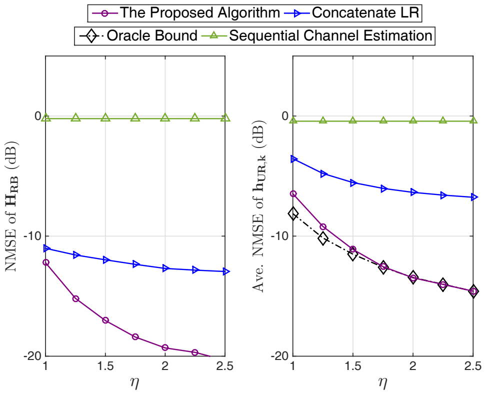

Fig. 4. The NMSE performance versus sampling resolution η with $\tau_N = -95$ dB.

We uniformly draw the central azimuth AoAs at the BS and the central azimuth (or elevation) AoDs at the RIS of each cluster, and draw each subpath with a 10° angular spread. The channels $\widetilde{\mathbf{H}}_{RB}$ and $\mathbf{h}_{UR,k}$ are generated by (5)–(6) in a similar way both with a cluster of 10 subpaths. We set $K = 20$, $M = 60$, $T = 35$, $L_1 = L_2 = 4$, $\tau_X = 1$, and $\kappa = 9$. The large-scale fading component is given by $\beta_i = \beta_{\text{ref}} \cdot d_i^{-\alpha_i}$, where $0 \leq i \leq K$; β_{ref} is the reference path loss at the distance 1 meter (m); d_i is the corresponding link distance; and α_i is the corresponding pass loss exponent. We set $\beta_{\text{ref}} = -20$ dB; $\alpha_0 = 2$; $\alpha_k = 2.6, 1 \leq k \leq K$; $d_0 = 50$ m; and d_k uniformly drawn from $[10 \text{ m}, 12 \text{ m}]$.

We define the evaluation metrics as the normalized MSE (NMSE) of $\mathbf{H}_{RB}$: $\|\hat{\mathbf{H}}_{RB} - \mathbf{H}_{RB}\|_F^2 / \|\mathbf{H}_{RB}\|_F^2$, and the NMSE of $\mathbf{h}_{UR,k}$: $\frac{1}{K}\sum_{k=1}^{K} \|\hat{\mathbf{h}}_{UR,k} - \mathbf{h}_{UR,k}\|_2^2 / \|\mathbf{h}_{UR,k}\|_2^2$. For the proposed algorithm, we set $\boldsymbol{\vartheta}$, $\boldsymbol{\varphi}$ and $\boldsymbol{\varsigma}$ to be uniform sampling grids covering $[-1, 1]$ with lengths $M'/M = L_1'/L_1 = L_2'/L_2 = \eta$ with η specified later. Apart from the proposed algorithm, the following baselines are involved for comparisons:

- Concatenate linear regression (LR): By setting aside $\widetilde{\mathbf{H}}_{RB}$, we first set the estimate of $\mathbf{H}_{RB}$ as $\hat{\mathbf{H}}_{RB} = \sqrt{\kappa/(\kappa+1)}\bar{\mathbf{H}}_{RB}$. We then infer $\hat{\mathbf{G}}$ by employing generalized AMP (GAMP) [16]. Finally, we employ GAMP to estimate $\widetilde{\mathbf{H}}_{RB}$ with the estimated $\hat{\mathbf{G}}$.
- Oracle bound with $\mathbf{H}_{RB}$ known: Assume that an oracle gives the accurate value of $\mathbf{H}_{RB}$. We employ GAMP to estimate $\hat{\mathbf{G}}$ with $\hat{\mathbf{H}}_{RB} = \mathbf{H}_{RB}$.
- Sequential channel estimation [6]: From time slots $(l-1)K + 1$ to lK, $1 \leq l \leq L$, we turn off all the RIS elements but the l-th one. With orthogonal training symbols from the users, the BS computes the LMMSE estimators of the channel coefficients associated with the l-th RIS element.

In Fig. 3, we plot the NMSEs of the proposed channel estimation algorithms as τ_N varies with $\eta = 2$. It can be seen that 1) the proposed algorithm achieves an NMSE of $\mathbf{h}_{UR,k}$ that is very close to the oracle bound, where the NMSE of $\mathbf{H}_{RB}$ is assumed to be zero. Moreover, the proposed algorithm outperforms the other baselines in all noise power levels; 2) the sequential channel estimation method has relatively large estimation errors, since it neither exploits the information on the slow-varying channel components nor exploits the hidden channel sparsity; 3) The NMSEs of concatenate LR does not decrease when $\tau_N \leq -90$ dB. The reason is that the effective noise in this case becomes a combination of the original AWGN noise and the error resulted from the model mismatch. Essentially, the latter term strongly correlates with $\mathbf{G}$. When the model mismatch error dominates the effective noise in the low noise power regime, the correlation issue compromises the convergence of GAMP. On the contrary, the proposed method avoids this problem since the estimate of $\widetilde{\mathbf{H}}_{RB}$, or equivalently $\mathbf{S}$, are updated iteratively during the message passing iteration.

Next, we study the effect of the grid lengths M', L_1', and L_2'. Fig. 4 plots the NMSEs of the channel estimation algorithms under various η, where τ_N is fixed to -95 dB. We have the following observations: 1) the NMSEs of the proposed algorithm, concatenate LR and the oracle estimator decrease as η increases, since increasing the sampling grid length leads to a higher angle resolution and hence sparser $\mathbf{S}$ and $\mathbf{G}$; 2) the baseline [6] does not exploit the channel sparsity in the angular domain and its performance is invariant to η; 3) the proposed method substantially improves the estimation performance compared to the other algorithms and closely approaches the oracle bound.

V. Conclusions

In this paper, we studied the channel estimation problem in the RIS-assisted multiuser MIMO system. We formulated the cascaded channel estimation task as a matrix-calibration based sparse matrix factorization problem by exploiting the knowledge of the slow-varying channel components and the hidden channel sparsity in the angular domain. Then, we proposed a novel message-passing based algorithm to infer the cascaded BS-to-RIS and RIS-to-user channels. Finally, we used numerical results to confirm the efficiency of the proposed algorithm.

References

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

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

[3] Q. Nadeem, A. Kammoun, A. Chaaban, M. Debbah, and M. Alouini, "Asymptotic max-min SINR analysis of reconfigurable intelligent surface assisted MISO systems," *IEEE Trans. Wireless Commun.*, pp. 1–1, 2020.

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

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

[6] Q.-U.-A. Nadeem, A. Kammoun, A. Chaaban, M. Debbah, and M.-S. Alouini, "Intelligent reflecting surface assisted multi-user MISO communication," *arXiv preprint arXiv:1906.02360*, 2019.

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

[8] Z. Wang, L. Liu, and S. Cui, "Channel estimation for intelligent reflecting surface assisted multiuser communications," *arXiv preprint arXiv:1911.03084*, 2019.

[9] J. Chen, Y.-C. Liang, H. V. Cheng, and W. Yu, "Channel estimation for reconfigurable intelligent surface aided multi-user MIMO systems," *arXiv preprint arXiv:1912.03619*, 2019.

[10] L. Wei, C. Huang, G. C. Alexandropoulos, and C. Yuen, "Parallel factor decomposition channel estimation in RIS-assisted multi-user MISO communication," *arXiv preprint arXiv:2001.09413*, 2020.

[11] D. L. Donoho, A. Maleki, and A. Montanari, "Message passing algorithms for compressed sensing," *Proc. Natl. Acad. Sci.*, vol. 106, no. 45, pp. 18 914–18 919, 2009.

[12] D. Tse and P. Viswanath, *Fundamentals of Wireless Communication*. New York, NY, USA: Cambridge University Press, 2005.

[13] X. Li, J. Fang, H. Li, and P. Wang, "Millimeter wave channel estimation via exploiting joint sparse and low-rank structures," *IEEE Trans. Wireless Commun.*, vol. 17, no. 2, pp. 1123–1133, Feb. 2018.

[14] A. F. Molisch, A. Kuchar, J. Laurila, K. Hugl, and R. Schmalenberger, "Geometry-based directional model for mobile radio channels–principles and implementation," *European Trans. Telecommun.*, vol. 14, no. 4, pp. 351–359, 2003.

[15] H. Liu, X. Yuan, and Y.-J. A. Zhang, "Matrix-calibration-based cascaded channel estimation for reconfigurable intelligent surface assisted multiuser MIMO," *arXiv preprint arXiv:1912.09025*, 2019.

[16] S. Rangan, P. Schniter, E. Riegler, A. K. Fletcher, and V. Cevher, "Fixed points of generalized approximate message passing with arbitrary matrices," *IEEE Trans. Inf. Theory*, vol. 62, no. 12, pp. 7464–7474, 2016.