Kalman Filter Based Channel Tracking for RIS-Assisted Multi-User Networks

Danyang Yu, Gan Zheng, *Fellow, IEEE*, Arman Shojaeifard, *Member, IEEE*, Sangarapillai Lambotharan, *Senior Member, IEEE*, and Yi Liu, *Senior Member, IEEE*

***Abstract*—In this paper, we investigate channel estimation in a reconfigurable intelligent surface (RIS) assisted multi-user network while considering the mobility of users. Based on a time-varying channel model, we utilize the Kalman filter (KF) that is able to exploit temporal correlation to track cascaded channels. In order to maintain a relatively low pilot overhead, we present a multiple sub-phases based transmission protocol where the number of pilot sequences in each sub-phase is less than the number of users, i.e., pilot contamination exists. For the sake of practicality, we directly utilize the discrete Fourier transform matrix as the RIS phase shift matrix during the training process. We analyze normalized mean square error and provide some asymptotic results. A more practical scenario with hardware impairments (HWI) at the transceiver and the RIS is considered. Since HWI is also part of the measurement matrix and is unknown to the base station, we propose a joint estimation of the channel and HWI. Under this joint estimation framework, the underlying state space model becomes nonlinear. We develop an extended KF (EKF) algorithm to tackle the nonlinearity through which the model can be linearized. Numerical results show that the proposed algorithms outperform benchmarks under various scenarios.**

***Index Terms*—Reconfigurable intelligent surface, channel prediction, pilot contamination, Kalman filter, extended Kalman filter, hardware impairments.**

I. Introduction

The rapid development of the fifth generation (5G) wireless network was underpinned by a variety of new techniques, including massive multiple-input multiple-output (MIMO) [1] and millimeter wave (mmWave) communication [2], with the aim of providing ultra-reliable low latency and high data rate communications to massive number of users [3]. However, due to high frequency bands, the transmission range of mmWave is limited due to the short wavelength and high path loss [4]. In addition, in massive MIMO systems, the pilot contamination problem arises due to the limited orthogonal pilot sequences [5]. Moreover, manufacturing small-size mmWave equipments and large-scale antenna arrays in massive MIMO systems incurs high costs.

Recent works have reported that reconfigurable intelligent surface (RIS) is able to reconfigure wireless propagation environment by careful selection of the phase shifts of a large number of passive reflecting elements with low cost and power consumption [6], [7]. Specifically, RIS is controlled by a smart controller which is able to tune the phase shifts of the incident waves before reflecting them out. Since RIS can strengthen the signal power of intended users and suppress the interference from unintended users, it achieves the reconfiguration of wireless channels. Moreover, different from conventional full-duplex (FD) relays, RIS can achieve passive beamforming in FD mode without processing delay, self-interference, and noise amplification [8]. With all the advantages mentioned above, RIS has attracted significant interest as a means of enhancing network capacity with lower implementation.

Achieving reflecting beamforming gain of RIS highly relies on obtaining accurate channel state information (CSI) [8]. Otherwise, the estimated channel would be outdated for the passive beamformer design. Generally, reflecting elements are passive and there is no active radio-frequency (RF) transceiver at the RIS, which makes channel estimation very challenging. Furthermore, with a large number of passive elements, the pilot overhead can be extremely high. Recently, numerous works have been devoted to the study of channel estimation in RIS systems [9]–[12]. Some prior works propose to equip the passive RIS with active antennas or RF chains to enable active channel estimation [9], but the potential advantages of RIS would be impaired. Since users can send pilot sequences, the CSI can also be obtained at the base station (BS) by analyzing the received pilot signal that is reflected by RIS. In this way, the cascaded BS-RIS-user channel can be obtained with the fully passive RIS, instead of estimating the BS-RIS and RIS-user channels separately. A simple ON/OFF method is developed to sequentially estimate the cascaded channel related to each element by switching on only one element of RIS and turning off all the other elements at a time [10]. However, the full RIS aperture gain is not exploited to improve estimation accuracy. In order to fully exploit the RIS aperture gain, the authors of [11], [12] designed the phase shift matrix based on the discrete Fourier transform (DFT) matrix. With the DFT-based reflection pattern, minimum estimation error can be achieved. On the other hand, there are some other works aiming to reduce pilot overhead [13], [14].

This work was supported in part by the National Key R&D Program of China (Grant No. 2021YFA0716500), in part by Project 111 of China under Grant B08038, and in part by the UK EPSRC under grant numbers EP/N007840/1, EPSRC EP/X012301/1, and EP/X04047X/1.

Danyang Yu and Yi Liu are with the State Key Laboratory of Integrated Service Network, Xidian University, Xi'an, 710071, China (email: dyyu_1@stu.xidian.edu.cn; yliu@xidian.edu.cn).

Gan Zheng is with the School of Engineering, University of Warwick, Coventry, CV4 7AL, U.K. (e-mail: gan.zheng@warwick.ac.uk).

Arman Shojaeifard is with the InterDigital, EC2A 3QR London, U.K. (e-mail: Arman.Shojaeifard@interdigital.com).

Sangarapillai Lambotharan is with the Institute for Digital Technologies, Loughborough University, London Queen Elizabeth Olympic Park, London E20 3BS. (e-mail: s.lambotharan@lboro.ac.uk).

Nonetheless, none of the works above have considered the mobility of users. When the users are mobile, the channel between RIS and users is time-varying. It is more challenging to accurately track the cascaded channel. Kalman filter (KF) is a useful tool to track unknown states based on a state space model. In a wireless communication system, the prior information of channels and observed measurements can be exploited by KF to improve estimation accuracy [15]. Several recent works have investigated channel estimation in RIS networks by utilizing KF [16]–[19]. However, all the works in [16]–[19] consider estimation in the case of a single user. When considering multiple users and pilot contamination, the estimation method derived from the single-user case cannot be directly applied. Moreover, the authors of [16], [17] assume Rayleigh fading for the BS-RIS and RIS-user channels, however, Rician fading is more relevant due to the ability of RIS to provide line-of-sight (LoS) links.

In addition, the aforementioned works are under the assumption of perfect hardware. While in practical systems, the inherent hardware impairments, e.g., phase noise, distortion noise, and phase drift at various kinds of devices, are not negligible, and they could significantly degrade the system performance [20], [21]. The authors of [22]–[24] have studied the channel estimation in RIS networks, taking hardware impairments into consideration. In [22]–[24], the impairments at transceiver and RIS are modeled as distortion noise and phase error, respectively. The authors of [22] adopted the linear minimum mean square error (LMMSE) method to estimate the overall channel which is the sum of the direct channel and the aggregate channel embedded with phase shift. Similarly, the authors of [23] proposed a deep-learning-based estimator to estimate the overall channel. However, the overall channel that is integrated with the fixed phase shift matrix cannot be utilized when the phase shift is changed. An ON/OFF protocol is adopted in [24] and the LMMSE estimation method is utilized to estimate the direct and the cascaded channels.

For non-RIS systems with hardware impairments (HWI), some works have investigated joint estimation of channel and HWI to improve estimation accuracy. For example, an earlier work proposed an approximate maximum-likelihood estimator for frequency offset and channel parameters [25], [26]. A joint estimation of frequency offset and channel is proposed in [27], where the frequency offset and channel are stacked into one vector to be estimated. A KF-based algorithm is proposed to jointly estimate carrier frequency offset (CFO) and time-varying channel in a MIMO system without RIS [15]. Since the CFO and channel are jointly estimated, the measurement equation is nonlinear, and thus the extended KF (EKF) that deals with the nonlinear case is adopted. The first-order Taylor approximation needs to be applied to linearize the nonlinear function and then the standard KF can be adopted. However, the joint estimation of channel and HWI has never been studied in RIS systems.

In order to fill the gaps mentioned above, in this paper, we investigate channel estimation in an RIS-assisted multi-user network where pilot contamination and HWI are taken into account. Specifically, considering the mobility of users, we use KF to track the time-varying cascaded channel. We consider a pilot contamination situation where the number of orthogonal pilot sequences is less than the number of users. Furthermore, we assume Rician fading for BS-RIS and RIS-user channels. We adopt the DFT matrix as the phase shift and provide some asymptotic results of normalized mean square error (NMSE). We then extend the work to the case where various kinds of HWI at transceiver and RIS exist. We adopt EKF to jointly estimate the HWI and the cascaded channel. The main contributions of this paper are summarized as follows.

- In the RIS system with mobile users, it is extremely challenging to acquire accurate channel state information of the cascaded channel. We propose a KF-based algorithm to track the time-varying cascaded channel for its simplicity and efficiency. To cater for the practical situation, we assume Rician fading for the BS-RIS and RIS-user channels in the considered RIS network. To reduce the pilot overhead, we present a transmission protocol through which less frequent pilots are used to complete the estimation of cascaded channels.
- In order to improve the estimation accuracy, the phase shift should be well designed. However, the mean square error (MSE) minimization problem is highly nonconvex due to the unit-modulus constraint. To avoid high computational complexity, we adopt the DFT matrix as the phase shift matrix. We analyze the NMSE and provide some asymptotic results under the DFT phase shift. We also prove that the optimal phase shift under Rayleigh fading and orthogonal pilot allocation can be reduced to any unitary matrix.
- We extend the channel estimation to the case of HWI. To the best of our knowledge, this is the first time that additive distortion noise and phase noise as well as multiplicative phase drift are taken into consideration simultaneously in an RIS system. The phase drift is part of the measurement matrix which is unknown to the BS, making the tracking process difficult. Fortunately, the phase drift can be modeled as a discrete-time Wiener process. Therefore, we propose to consider the phase drift as part of the state vector and jointly estimate the phase drift and cascaded channel. The measurement equation is nonlinear due to the HWI. We propose EKF to solve this problem by applying the first-order Taylor approximation to linearize it and then applying the standard KF.
- Finally, we validate the effectiveness of the proposed KF and EKF algorithms. The NMSE achieved by the proposed algorithms is superior to the benchmark schemes under various scenarios. Numerical results demonstrate that when there is no HWI, the NMSE performance can be improved by creating a stronger LoS link, improving the transmit power of pilot sequences, and increasing the number of RIS elements. When the HWI at transceiver and RIS is considered, the estimation accuracy is degraded particularly as the number of BS antennas and RIS elements increases.

The rest of this paper is organized as follows. In Section II, we present a system model. In Section III, a KF-based

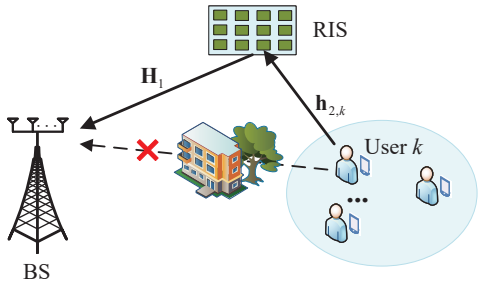

Fig. 1. System model of an RIS-assisted multi-user network.

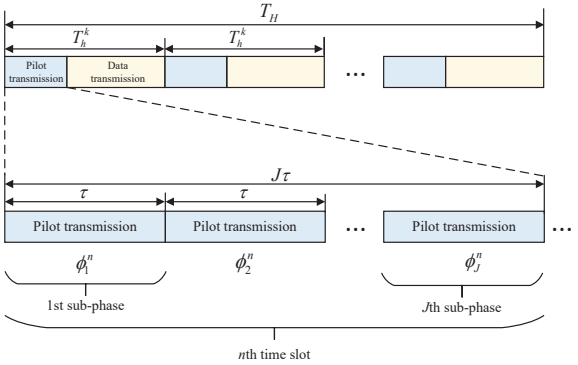

Fig. 2. Coherence time illustration and pilot transmission protocol.

algorithm is proposed to track the cascaded channel. The DFT matrix is adopted as the phase shift matrix during the estimation process and the NMSE analysis is presented. We extend the channel estimation to the case of various kinds of HWI and present an EKF-based algorithm in Section IV. Numerical results are shown in Section V followed by conclusions in Section VI.

Notations: $\mathbb{C}^{x\times y}$ denotes the space of $x \times y$ complex-valued matrices. For a complex-valued vector $\mathbf{x}$, $\mathrm{diag}(\mathbf{x})$ denotes a diagonal matrix with each diagonal element being the corresponding element in $\mathbf{x}$. For a matrix $\mathbf{M}$, $[\mathbf{M}]_{i,j}$ denotes the (i,j)th element and $\mathrm{tr}\{\mathbf{M}\}$ is the trace operator. The superscripts $(\cdot)^T$ and $(\cdot)^H$ stand for transpose and conjugate transpose, respectively. The distribution of a circularly symmetric complex Gaussian (CSCG) random variable with mean x and covariance σ is denoted by $\mathcal{CN}(x,\sigma)$, and $\sim$ stands for distributed as. $\mathcal{U}(\cdot)$ stands for uniform distribution. 's.t.' denotes 'subject to'. $\odot$ denotes Hadamard product. $\otimes$ denotes the Kronecker product. $\Re\{\cdot\}$ stands for taking the real part. $\mathbb{E}\{\cdot\}$ and $\mathbb{V}\{\cdot\}$ denote the expectation and variance of a matrix, respectively.

II. SYSTEM MODEL

In this paper, we investigate an RIS-assisted multi-user network, where K single-antenna users communicate with a BS with the aid of an RIS, as shown in Fig. 1. The BS is equipped with M antennas and the RIS consists of S passive reflecting elements. The RIS is connected to a smart controller that dynamically configures the phase shift of RIS by the BS. We assume that the direct link between the BS and users is blocked by obstacles, and the users move with low to medium speed.

A. Channel Model

We denote $\mathbf{H}_1^n \in \mathbb{C}^{M\times S}$ and $\mathbf{h}_{2,k}^n \in \mathbb{C}^{S\times 1}$ as the channel from the RIS to the BS as well as the channel from the kth user to the RIS at the nth time slot, respectively. We consider a time-varying channel model that remains constant within its coherence time. In Fig. 2, T_H and T_h^k denote the coherence time of $\mathbf{H}_1^n$ and $\mathbf{h}_{2,k}^n$, respectively. Since the BS and RIS have fixed locations, we assume that the BS-RIS channel changes slower than the RIS-user channel, i.e., $T_H > T_h^k$. Specifically, we assume $T_H = NT_h^k$, where N is an integer. For simplicity, we assume $\mathbf{H}_1^1 = \mathbf{H}_1^2 = \cdots \mathbf{H}_1^N = \mathbf{H}_1$.

We assume Rician fading for the BS-RIS and RIS-user channels. Within the coherence time T_H, the BS-RIS and RIS-user channels can be modeled as

$$\mathbf{H}_1 = \sqrt{\beta_{BR}}\sqrt{\frac{\kappa}{\kappa+1}}\bar{\mathbf{H}}_1 + \sqrt{\beta_{BR}}\sqrt{\frac{1}{\kappa+1}}\hat{\mathbf{H}}_1, \tag{1}$$

$$\mathbf{h}_{2,k}^n = \sqrt{\beta_{Ru}(k)}\sqrt{\frac{\kappa}{\kappa+1}}\bar{\mathbf{h}}_{2,k} + \sqrt{\beta_{Ru}(k)}\sqrt{\frac{1}{\kappa+1}}\hat{\mathbf{h}}_{2,k}^n, \tag{2}$$

where κ denotes the Rician factor. β_{BR} and $\beta_{Ru}(k)$ are the large-scale fading coefficients of the BS-RIS and RIS-user channels. $\hat{\mathbf{H}}_1$ and $\hat{\mathbf{h}}_{2,k}^n$ represent the normalized non-LoS (NLoS) components, with elements independently and identically distributed according to $\mathcal{CN}(0,1)$. $\bar{\mathbf{H}}_1$ and $\bar{\mathbf{h}}_{2,k}$ represent the normalized LoS components. Here, we drop the time index n of $\mathbf{H}_1^n$ since $\mathbf{H}_1$ does not change within its coherence time. Moreover, we assume the LoS component $\bar{\mathbf{h}}_{2,k}^n$ does not change within the coherence time T_H and drop the time index n, because it is more reasonable to assume a stable LoS component[1]. The LoS components are given by $\bar{\mathbf{H}}_1 = \mathbf{a}_r\left(\omega^{AoA}\right)\mathbf{a}_t^H\left(\vartheta^{AoD,a},\vartheta^{AoD,e}\right)$ and $\bar{\mathbf{h}}_{2,k} = \mathbf{a}_r\left(\vartheta_k^{AoA,a},\vartheta_k^{AoA,e}\right)$, respectively. We assume a uniform linear array and a uniform planar array (UPA) for the BS and the RIS, respectively. The steering vectors $\mathbf{a}_r\left(\omega^{AoA}\right)$ and $\mathbf{a}_t\left(\vartheta^{AoD,a},\vartheta^{AoD,e}\right)$ are given by

$$\mathbf{a}_r\left(\omega^{AoA}\right) = \left[1, e^{j\frac{2\pi d}{\lambda}\sin\omega^{AoA}}, \cdots, e^{j\frac{2\pi d}{\lambda}(M-1)\sin\omega^{AoA}}\right]^T, \tag{3}$$

$$\mathbf{a}_t\left(\vartheta^{AoD,a},\vartheta^{AoD,e}\right) = \left[1, e^{j\frac{2\pi d}{\lambda}\sin\vartheta^{AoD,a}\sin\vartheta^{AoD,e}}, \cdots, e^{j\frac{2\pi d(S_H-1)}{\lambda}\sin\vartheta^{AoD,a}\sin\vartheta^{AoD,e}}\right]^T \otimes \left[1, e^{j\frac{2\pi d}{\lambda}\cos\vartheta^{AoD,e}}, \cdots, e^{j\frac{2\pi d(S_V-1)}{\lambda}\cos\vartheta^{AoD,e}}\right]^T, \tag{4}$$

where d denotes the antenna or element separation distance and λ denotes the wavelength. For simplicity, we assume $d = \lambda/2$. S_H and S_V denote the number of elements in the

[1]The NLoS component $\hat{\mathbf{h}}_{2,k}^n$ changes between different time slots. Therefore, the channel $\mathbf{h}_{2,k}^n$ consisting of a LoS component and a NLoS component changes with n.

horizontal and vertical directions, respectively. The number of RIS elements is equal to $S = S_H \times S_V$. ω^{AoA} denotes the angle of arrival from the RIS to the BS. $\vartheta^{AoD,a}$ and $\vartheta^{AoD,e}$ are the azimuth and elevation angles of departure from the RIS to the BS. $\vartheta_k^{AoA,a}$ and $\vartheta_k^{AoA,e}$ are the azimuth and elevation angles of arrival from the kth user to the RIS. We can use (4) to calculate $\mathbf{a}_r\left(\vartheta_k^{AoA,a}, \vartheta_k^{AoA,e}\right)$ by replacing $(\vartheta^{AoD,a}, \vartheta^{AoD,e})$ with $(\vartheta_k^{AoA,a}, \vartheta_k^{AoA,e})$.

Considering possible different velocities of different users, the channel from the RIS to the kth user $\mathbf{h}_{2,k}^n$ can be modeled by the first-order auto-regressive (AR) model as follows [17], [28]:

$$\mathbf{h}_{2,k}^n = \mathbf{A}_k^n \mathbf{h}_{2,k}^{n-1} + \mathbf{u}_k^n, \tag{5}$$

where $\mathbf{A}_k^n = \mathrm{diag}(a_{k,1}^n, a_{k,2}^n, \cdots, a_{k,S}^n) \in \mathbb{C}^{S\times S}$ represents the AR parameter matrix of the kth user and $\mathbf{u}_k^n$ denotes the innovation process. Due to that $\mathbf{h}_{2,k}^n$ is Rician distributed, the elements of $\mathbf{u}_k^n$ are complex Gaussian distributed with mean $(1-a_k^n)\mathbb{E}\{\mathbf{h}_{2,k}^n\}$ and variance $(1-(a_k^n)^2)\mathbb{V}\{\mathbf{h}_{2,k}^n\}$.

The element in $\mathbf{A}_k^n$, i.e., $a_{k,s}^n$, denotes the time-correlation coefficient. For simplicity, we assume that different elements in the same RIS have similar time-correlation characteristics [16], [17], [29], i.e., $a_{k,s}^n = a_k^n$. Using the standard Jakes' model [17], [30], the time-correlation coefficient is determined by the symbol interval between the adjacent time slots, i.e., $J_0\left(2\pi f_D^k |n-(n-1)|T_h^k\right)$, where $J_0(\cdot)$ denotes the zero-order Bessel function and f_D^k denotes the maximum Doppler shift in frequency of the kth user. Therefore, the time-correlation coefficient a_k can be determined by $a_k = J_0\left(2\pi f_D^k T_h^k\right)$. $\mathbf{A}_k^n$ can be simplified as $\mathbf{A}_k = \mathrm{diag}(a_k, a_k, \cdots, a_k)$.

B. Signal Model

In the RIS-assisted network, we estimate the BS-RIS-user cascaded channel. The estimated CSI can be obtained by allowing users to send pilots to BS. To begin with, we should figure out how many training symbols are needed to complete the estimation of the cascaded channel in a multi-user network. Assume that there is only one element in the RIS and each user needs to send a pilot with length τ. If $\tau = K$ and all the pilots are orthogonal to each other, there is no pilot contamination. If $\tau < K$, the pilot contamination exists. Since each RIS has S elements, in order to estimate the channel accurately, we need to either consecutively estimate the cascaded channel of each element by the ON/OFF method, or turn on all the elements but extend the training period to obtain more observations [31]. In this case, the total length of the pilot should be $\tau \times S$. From [8], we know that the benefit brought by RIS could be deteriorated by using the first method while the full RIS aperture can be fully utilized by using the second method. As a result, we present a pilot transmission protocol as depicted in Fig. 2, where the pilot transmission period of each time slot is further divided into multiple sub-phases. Specifically, each time slot is separated into $J = S$ sub-phases and each sub-phase consists of τ training symbols[2].

[2]According to [32], the least square (LS) estimator requires $J \geq S$. We adopt $J = S$ to maintain a relatively low training overhead.

Considering the large number of elements in an RIS, a large amount of pilots is needed to estimate the cascaded channel. In order to reduce pilot overhead, we assume that in each sub-phase the length of pilots τ is shorter than the number of users, i.e., pilot contamination exists. In this case, it is possible that multiple users transmit the same pilot sequence to the BS. Define $\mathbf{s}_{k,j}^n = [s_{k,j,1}^n, s_{k,j,2}^n, \cdots, s_{k,j,\tau}^n] \in \mathbb{C}^{1\times\tau}$ as the pilot sequence used by the kth user at the jth sub-phase. In each sub-phase, $\mathbf{s}_{k,j}^n$ is randomly selected from the set of orthogonal pilots. We define the set $\mathcal{C}_{k,j} = \{i | \mathbf{s}_{i,j}^n = \mathbf{s}_{k,j}^n, \forall i \neq k\}$ as the set of indices of the users that have the same pilot sequence as the kth user. Therefore, we have

$$\mathbf{s}_{k',j}^n (\mathbf{s}_{k,j}^n)^H = \begin{cases} 1, & \text{if } k' \in \mathcal{C}_{k,j}, \\ 0, & \text{if } k' \notin \mathcal{C}_{k,j}. \end{cases} \tag{6}$$

When the kth user sends the pilot sequence $\mathbf{s}_{k,j}^n$ to the BS, the received signal at the BS at the jth sub-phase of the nth time slot is given by

$$\begin{aligned} \mathbf{y}_j^n &= \sqrt{p}\sum_{k=1}^K \mathbf{H}_1 \mathbf{\Phi}_j^n \mathbf{h}_{2,k}^n \mathbf{s}_{k,j}^n + \mathbf{w}_j^n \\ &= \sqrt{p}\sum_{k=1}^K \mathbf{H}_1 \,\mathrm{diag}(\mathbf{h}_{2,k}^n) \boldsymbol{\phi}_j^n \mathbf{s}_{k,j}^n + \mathbf{w}_j^n, \end{aligned} \tag{7}$$

where $\mathbf{\Phi}_j^n = \mathrm{diag}(\left[\phi_{j,1}^n, \phi_{j,2}^n, \cdots, \phi_{j,S}^n\right])$ represents the phase shift matrix of the RIS at the jth sub-phase in the nth slot. $\phi_{j,s}^n$ denotes the phase shift of the sth element at the jth sub-phase in the nth slot. p represents the transmit power of training sequences. $\mathbf{w}_j^n \in \mathbb{C}^{M\times\tau}$ denotes the Gaussian noise with zero mean and variance σ^2. Define $\boldsymbol{\phi}_j^n \triangleq \left[\phi_{j,1}^n, \phi_{j,2}^n, \cdots, \phi_{j,S}^n\right]^T \in \mathbb{C}^{S\times 1}$.

By defining $\mathbf{G}_k^n \triangleq \mathbf{H}_1 \,\mathrm{diag}(\mathbf{h}_{2,k}^n) \in \mathbb{C}^{M\times S}$ as the cascaded channel, (7) can be rewritten as

$$\mathbf{y}_j^n = \sqrt{p}\sum_{k=1}^K \mathbf{G}_k^n \boldsymbol{\phi}_j^n \mathbf{s}_{k,j}^n + \mathbf{w}_j^n. \tag{8}$$

In order to estimate the cascaded channel of the kth user, the LS estimation is performed and we have

$$\frac{\mathbf{y}_j^n \left(\mathbf{s}_{k,j}^n\right)^H}{\sqrt{p}} = \mathbf{G}_k^n \boldsymbol{\phi}_j^n + \sum_{i\in\mathcal{C}_{k,j}} \mathbf{G}_i^n \boldsymbol{\phi}_j^n + \frac{\mathbf{w}_j^n \left(\mathbf{s}_{k,j}^n\right)^H}{\sqrt{p}}, \tag{9}$$

where the second term on the right side is the contaminated signal of the kth user.

By defining $\mathbf{v}_{k,j}^n \triangleq \sum_{i\in\mathcal{C}_{k,j}} \mathbf{G}_i^n \boldsymbol{\phi}_j^n + \mathbf{w}_j^n (\mathbf{s}_{k,j}^n)^H/\sqrt{p}$ and applying transpose to (9), we have

$$\left[\frac{\mathbf{y}_j^n \left(\mathbf{s}_{k,j}^n\right)^H}{\sqrt{p}}\right]^T = (\boldsymbol{\phi}_j^n)^T (\mathbf{G}_k^n)^T + (\mathbf{v}_{k,j}^n)^T. \tag{10}$$

By stacking the received signal in (10) for all sub-phases, we have

$$\mathbf{Y}_k^n = \mathbf{F}^n (\mathbf{G}_k^n)^T + \mathbf{V}_k^n, \tag{11}$$

where

$$\mathbf{Y}_k^n = \left[\frac{\mathbf{y}_1^n \left(\mathbf{s}_{k,1}^n\right)^H}{\sqrt{p}}, \frac{\mathbf{y}_2^n \left(\mathbf{s}_{k,2}^n\right)^H}{\sqrt{p}}, \cdots, \frac{\mathbf{y}_J^n \left(\mathbf{s}_{k,J}^n\right)^H}{\sqrt{p}} \right]^T, \tag{12}$$

$$\mathbf{F}^n = [\boldsymbol{\phi}_1^n, \boldsymbol{\phi}_2^n, \cdots, \boldsymbol{\phi}_J^n]^T, \tag{13}$$

$$\mathbf{V}_k^n = \left[\mathbf{v}_{k,1}^n, \mathbf{v}_{k,2}^n, \cdots, \mathbf{v}_{k,J}^n\right]^T. \tag{14}$$

Finally, by stacking the received signal in (11) for all the users, we have

$$\mathbf{Y}^n = \bar{\mathbf{F}}^n \mathbf{G}^n + \mathbf{V}^n, \tag{15}$$

where $\mathbf{Y}^n = [(\mathbf{Y}_1^n)^T, (\mathbf{Y}_2^n)^T, \cdots, (\mathbf{Y}_K^n)^T]^T \in \mathbb{C}^{KJ\times M}$, $\bar{\mathbf{F}}^n = \mathrm{diag}(\mathbf{F}^n, \mathbf{F}^n, \cdots, \mathbf{F}^n) \in \mathbb{C}^{KJ\times KS}$, $\mathbf{G}^n = [\mathbf{G}_1^n, \mathbf{G}_2^n, \cdots, \mathbf{G}_K^n]^T \in \mathbb{C}^{KS\times M}$, and $\mathbf{V}^n = [(\mathbf{V}_1^n)^T, (\mathbf{V}_2^n)^T, \cdots, (\mathbf{V}_K^n)^T]^T \in \mathbb{C}^{KJ\times M}$.

III. Kalman Filter based Channel Prediction

A. Kalman Filter Based Tracking

Considering the mobility of users and the dynamic characteristic of channels, the required time for the channel estimation might exceed the coherence time of channels and thus the obtained CSI might be outdated. This will hinder the beamforming design that is based on the estimated CSI. One simple method to mitigate this issue is to transmit pilot sequences more frequently. However, the pilot overhead could be overwhelming since RIS has a large number of passive elements. Therefore, performing channel estimation only without consideration of its evolution over time is inefficient. Kalman filter is the Bayesian solution to the problem of sequentially estimating the states of a dynamical system. It can provide predictions of unknown states. Meanwhile, the estimation accuracy can be improved given the measurements observed over time. Therefore, it is appropriate for us to adopt the Kalman filter to track the cascaded channel. In the previous section, we have already obtained the dynamic model of the channel between RIS and users. Next, we need to obtain the dynamic model of the cascaded channel.

By applying diagonalization to both sides of (5) and multiplying it with $\mathbf{H}_1$, we have[3]

$$\begin{aligned}\mathbf{H}_1 \,\mathrm{diag}(\mathbf{h}_{2,k}^n) &= \mathbf{H}_1 \mathbf{A}_k \,\mathrm{diag}(\mathbf{h}_{2,k}^{n-1}) + \mathbf{H}_1 \,\mathrm{diag}(\mathbf{u}_k^n) \\ &= a_k \mathbf{H}_1 \,\mathrm{diag}(\mathbf{h}_{2,k}^{n-1}) + \mathbf{H}_1 \,\mathrm{diag}(\mathbf{u}_k^n). \end{aligned} \tag{16}$$

According to the definition of the cascaded channel, (16) can be reexpressed as

$$(\mathbf{G}_k^n)^T = a_k (\mathbf{G}_k^{n-1})^T + \mathbf{U}_k^n, \tag{17}$$

where $\mathbf{U}_k^n = [\mathrm{diag}(\mathbf{u}_k^n)]^T \mathbf{H}_1^T$.

By stacking the state equation (17) for all the users, we have

$$\mathbf{G}^n = \mathbf{A}\mathbf{G}^{n-1} + \mathbf{U}^n, \tag{18}$$

[3]If $\mathbf{H}_1$ changes when one coherence time T_H ends and a new time slot in the next coherence time begins, (16) is not satisfied. Therefore, we assume $\mathbf{H}_1$ does not change during the estimation process for simplicity.

where $\mathbf{A} = \mathrm{diag}(a_1\mathbf{I}_S, a_2\mathbf{I}_S, \cdots, a_K\mathbf{I}_S) \in \mathbb{C}^{KS\times KS}$. $\mathbf{U}^n = [\mathbf{H}_1 \mathrm{diag}(\mathbf{u}_1^n), \mathbf{H}_1 \mathrm{diag}(\mathbf{u}_2^n), \cdots, \mathbf{H}_1 \mathrm{diag}(\mathbf{u}_K^n)]^T \in \mathbb{C}^{KS\times M}$.

Based on the measurement equation (15) and the state equation (18), we obtain the overall state space model as follows:

$$\begin{cases} \mathbf{Y}^n = \bar{\mathbf{F}}^n \mathbf{G}^n + \mathbf{V}^n, & \text{(19a)} \\ \mathbf{G}^n = \mathbf{A}\mathbf{G}^{n-1} + \mathbf{U}^n. & \text{(19b)} \end{cases}$$

Based on the state space model, we show the procedures of predicting the cascaded channel in Algorithm 1, where $\mathbf{P}^{n|n-1}$ and $\mathbf{P}^{n|n}$ denote the priori and posteriori covariance matrices, respectively. The superscript notation '$i|j$' is used to indicate the association with the probability density function of the state vector $\mathbf{G}$ computed at the ith time slot using the measurements up to the jth time slot. $\mathbf{Q}^n$ represents the state noise covariance matrix $\mathbf{U}^n$. $\mathbf{R}^n$ is the measurement noise covariance matrix $\mathbf{V}^n$. The expressions of $\mathbf{Q}^n$ and $\mathbf{R}^n$ are given in Lemma 1.

Algorithm 1 Channel Prediction Based on Kalman Filter

1: **Initialization**: $\mathbf{G}^0 = \mathbf{0}_{KS\times M}$, $\mathbf{P}^0 = M\mathbf{I}_{KS}$, $n = 1$, and the total number of time slots N.
2: **repeat**
3: **Prediction**: $\mathbf{G}^{n|n-1} = \mathbf{A}\mathbf{G}^{n-1|n-1}$ and $\mathbf{P}^{n|n-1} = \mathbf{A}\mathbf{P}^{n-1|n-1}\mathbf{A}^H + \mathbf{Q}^n$.
4: **Kalman Gain**: $\mathbf{S}^n = \bar{\mathbf{F}}^n \mathbf{P}^{n|n-1} (\bar{\mathbf{F}}^n)^H + \mathbf{R}^n$, $\mathbf{K}^n = \mathbf{P}^{n|n-1} (\bar{\mathbf{F}}^n)^H (\mathbf{S}^n)^{-1}$.
5: **Correction**: $\mathbf{G}^{n|n} = \mathbf{G}^{n|n-1} + \mathbf{K}^n (\mathbf{Y}^n - \bar{\mathbf{F}}^n \mathbf{G}^{n|n-1})$ and $\mathbf{P}^{n|n} = \mathbf{P}^{n|n-1} - \mathbf{K}^n \bar{\mathbf{F}}^n \mathbf{P}^{n|n-1}$.
6: Set $n = n + 1$.
7: **until** $n = N$

Specifically, the overall prediction process consists of four major steps. At the beginning, the estimates of the channel matrix and the MSE matrix are initialized. Then, the prediction results are derived according to step 3. Next, the Kalman gain is calculated by step 4. Kalman gain matrix is used to determine how much of a new measurement should be used for the next update of the cascaded channel. Finally, after receiving the new measurements, the estimated channel and MSE matrices can be corrected by utilizing Kalman gain. Steps 3-5 are repeated until the end of transmission.

Lemma 1. Under Rician fading and pilot contamination, the expression of $\mathbf{Q}_k^n$ is given as

$$\begin{aligned}\mathbf{Q}_k = \mathbf{X}_k^{\mathrm{cov}} &- (1-a_k)^2 \left(\frac{\kappa}{\kappa+1}\right)^2 \beta_{Ru}(k)\beta_{BR} \\ &\times \mathrm{diag}(\bar{\mathbf{h}}_{2,k})(\bar{\mathbf{H}}_1)^T (\bar{\mathbf{H}}_1)^* \,\mathrm{diag}^H(\bar{\mathbf{h}}_{2,k}), \end{aligned} \tag{20}$$

where $\mathbf{X}_k^{\mathrm{cov}}$ is defined in (48). The diagonal elements of the diagonal matrix $\mathbf{R}_k^n$ is given as

$$[\mathbf{R}_k^n]_{j,j} = \sum_{i\in C_{k,j}} \left(\boldsymbol{\phi}_j^n\right)^T \mathbf{G}_i^{\mathrm{cov}} \left(\boldsymbol{\phi}_j^n\right)^* + M\sigma^2/p$$

$$- \left(\sum_{i\in C_{k,j}} \left(\boldsymbol{\phi}_j^n\right)^T \frac{\kappa}{\kappa+1} \sqrt{\beta_{Ru}(i)} \sqrt{\beta_{BR}} \,\mathrm{diag}(\bar{\mathbf{h}}_{2,i})(\bar{\mathbf{H}}_1)^T \right)$$

$$\times \left(\sum_{i \in C_{k,j}} \left(\boldsymbol{\phi}_j^n\right)^T \frac{\kappa}{\kappa+1} \sqrt{\beta_{Ru}(i)} \sqrt{\beta_{BR}}\, \mathrm{diag}(\bar{\mathbf{h}}_{2,i}) (\bar{\mathbf{H}}_1)^T \right)^H, \tag{21}$$

where the definition of $\mathbf{G}_i^{\mathrm{cov}}$ is given in Appendix A.

Proof: See in Appendix A. ∎

B. Complexity Analysis

Algorithm 1 is an iterative algorithm and the computational complexity of each iteration is determined by the prediction, update of Kalman gain, and the correction steps. The prediction step involves three matrix multiplications, and the number of real-valued multiplications is $8K^3S^3 + 2K^2S^2M$ [33]. The update of Kalman gain involves four matrix multiplications and one matrix inverse, and thus the number of real-valued multiplications is $17K^3S^3$. The correction step involves three matrix multiplications, and thus the number of real-valued multiplications is $8K^3S^3 + 4K^2S^2M$. In step 5, the computational complexity of the LS method is $O(KSM\tau)$. Overall, the total computational complexity of the Algorithm 1 is $O(33K^3S^3 + 6K^2S^2M + KSM\tau)$.

We also analyze the complexity of the existing methods introduced in section I. Since [17] adopts the same LS method and KF algorithm, the complexity can be obtained by setting $K = 1$ as they assume a single user case. In [13], [14], the LMMSE estimator is adopted. In [13], the main complexity is the calculation of the second moment of the aggregated channel and is given by $O(12KMS^3)$. The LMMSE estimator for each user and each time slot in [14] involves seven matrix multiplications and seven matrix inverse. The number of real-valued multiplications is approximately $O(11S^3 + (19M^3 + 4M^2)(K-1)(\lceil \frac{S}{M} \rceil - 1) + (12\Delta M^2 + 4\Delta^2 M + 2\Delta^3 + M^3 + 4\Delta M)(K-1))$, where $\Delta = S - (\lceil \frac{S}{M} \rceil - 1)M$. Based on the above analysis, the complexity of the proposed algorithm is in the same order of S as the benchmarks in [13], [14], [17].

C. Phase Shift Design

From Algorithm 1, we know that the MSE matrix depends on the phase shift. Therefore, we can optimize the phase shift during the prediction process to minimize MSE[4]. Let us first calculate the MSE of the kth user. According to Algorithm 1, the value of the prior covariance at the nth time slot is

$$\mathbf{P}_k^{n|n-1} = a_k^2 \mathbf{P}_k^{n-1|n-1} + \mathbf{Q}_k \triangleq \bar{\mathbf{Q}}_k^n. \tag{22}$$

The posterior covariance at the nth time slot is calculated as

$$\mathbf{P}_k^{n|n} = \left\{ \mathbf{I} - \left[\bar{\mathbf{Q}}_k^n \mathbf{F}^H \left(\mathbf{F} \bar{\mathbf{Q}}_k^n \mathbf{F}^H + \mathbf{R}_k \right)^{-1} \right] \mathbf{F} \right\} \bar{\mathbf{Q}}_k^n. \tag{23}$$

Then the MSE of the kth user can be calculated as

$$\mathrm{tr}\{\mathbf{P}_k^{n|n}\} = \mathrm{tr}\{\bar{\mathbf{Q}}_k^n\} - \mathrm{tr}\left\{ \bar{\mathbf{Q}}_k^n \mathbf{F}^H \left(\mathbf{F} \bar{\mathbf{Q}}_k^n \mathbf{F}^H + \mathbf{R}_k \right)^{-1} \mathbf{F} \bar{\mathbf{Q}}_k^n \right\}. \tag{24}$$

By summing the MSE of the kth user for all the users, the MSE minimization problem at the nth slot can be equivalently formulated as the following maximization problem.

$$\max_{\mathbf{F}} \quad \sum_{k=1}^{K} \mathrm{tr}\left\{ \bar{\mathbf{Q}}_k^n \mathbf{F}^H \left(\mathbf{F} \bar{\mathbf{Q}}_k^n \mathbf{F}^H + \mathbf{R}_k \right)^{-1} \mathbf{F} \bar{\mathbf{Q}}_k^n \right\} \tag{25a}$$

$$\text{s.t.} \quad |[\mathbf{F}]_{j,s}| = 1, \quad \forall j, s, \tag{25b}$$

where the constraint (25b) is derived from the unit-modulus constraint of phase shift $\boldsymbol{\phi}_j^n$ and the definition of $\mathbf{F}$ in (14). $[\mathbf{F}]_{j,s}$ denotes the element at the jth row and the sth column of the matrix $\mathbf{F}$.

The problem (25) is difficult to solve directly due to the nonconvex unit-modulus constraint. From the literature, we know that a popular method of solving the phase shift optimization problem is to approximate it into a semidefinite programming problem and iteratively solve it [34]. However, it is neither simple nor practical because of its high computational complexity. In order to reduce the computational complexity, a simple and effective method of designing phase shift is needed. Another general way to design phase shifts is to adopt the DFT matrix [35]. The validity of the DFT matrix has been verified in [11], [12], [16], [36]. The objective function in (25a) has a similar form to the equation (22) in [36]. The empirical study in [36] reveals that the DFT matrix is a stationary point and the MSE of local optima are close to the DFT-based solution. Although [36] considers a single user case, the DFT-based phase shift matrix is also a stationary point of our problem (25) because all the users share the same RIS. Since the DFT-based phase shift matrix is determined prior to the channel estimation process, the computational complexity is largely reduced.

In the following proposition, we show that the optimal phase shift under Rayleigh fading and orthogonal pilot sequences could be any unitary matrix that satisfies the unit-modulus constraint.

Proposition 1. When $\kappa = 0$ and $\tau = K$, i.e., in the case of Rayleigh fading channel and orthogonal pilot allocation, the optimal phase shift that minimizes MSE is any unitary matrix that satisfies the unit-modulus constraint [17].

Proof: According to Lemma 1, $\mathbf{Q}_k^n$ is reduced to $M\left(1 - a_k^2\right)\beta_{BR}\beta_{Ru}(k)\mathbf{I}_S$ when the Rician factor satisfies $\kappa = 0$. When $\tau = K$, $\mathbf{R}_k^n$ is reduced to $M\sigma^2/p\mathbf{I}_S$. That is, both $\mathbf{Q}_k^n$ and $\mathbf{R}_k^n$ are scaled identity matrices. As a result, the optimal phase shift is reduced to any unitary matrix that satisfies the unit-modulus constraint. Detailed derivations can be found in [17]. ∎

D. NMSE Analysis Under DFT Phase Shift

In this subsection, we analyze the NMSE obtained under the DFT phase shift. The NMSE of the system is defined as

$$\begin{aligned} \mathrm{NMSE} &= \mathbb{E}\left[\frac{\|\mathbf{G}^{n|n} - \mathbf{G}^n\|_F^2}{\|\mathbf{G}^n\|_F^2} \right] \\ &= \frac{\mathbb{E}\left[\sum_{k=1}^K \|\mathbf{G}_k^{n|n} - \mathbf{G}_k^n\|_F^2 \right]}{\mathbb{E}\left[\sum_{k=1}^K \|\mathbf{G}_k^n\|_F^2 \right]}. \end{aligned} \tag{26}$$

[4]The BS uses a smart controller to configure the phase shift.

The denominator in (26) can be rewritten as $\sum_{k=1}^{K}\mathbb{E}\left[\mathrm{tr}\left\{\mathbf{G}_k^n(\mathbf{G}_k^n)^H\right\}\right]$. In Appendix A, $\mathbf{G}_i^{\mathrm{cov}}$ is defined as $\mathbf{G}_i^{\mathrm{cov}} \triangleq \mathbb{E}\left\{(\mathbf{G}_i^n)^T(\mathbf{G}_i^n)^*\right\}$. Therefore, the denominator can be further rewritten as $\sum_{k=1}^{K}\mathrm{tr}\left\{\mathbf{G}_k^{\mathrm{cov}}\right\}$. Here, we omit the details and directly present the results in the following.

$$\begin{aligned}
&\sum_{k=1}^{K}\mathrm{tr}\left\{\mathbf{G}_k^{\mathrm{cov}}\right\}\\
&=\sum_{k=1}^{K}\sum_{i=1}^{S}\sum_{m=1}^{M}\Bigg(\frac{1}{\kappa+1}\beta_{Ru}(k)\frac{1}{\kappa+1}\beta_{BR}\\
&+\frac{1}{\kappa+1}\beta_{Ru}(k)\frac{\kappa}{\kappa+1}\beta_{BR}[\bar{\mathbf{H}}_1]_{m,i}^2\\
&+\frac{1}{\kappa+1}\beta_{BR}\frac{\kappa}{\kappa+1}\beta_{Ru}(k)[\bar{\mathbf{h}}_{2,k}]_i^2\\
&+(\frac{\kappa}{\kappa+1})^2\beta_{Ru}(k)\beta_{BR}|[\bar{\mathbf{h}}_{2,k}]_i|^2|[\bar{\mathbf{H}}_1]_{m,i}|^2\Bigg). \qquad (27)
\end{aligned}$$

From section III-C, the MSE of all the users, i.e., the numerator in (26) is given by $\sum_{k=1}^{K}\mathrm{tr}\{\bar{\mathbf{Q}}_k^n\} - \sum_{k=1}^{K}\mathrm{tr}\left\{\bar{\mathbf{Q}}_k^n\mathbf{F}^H\left(\mathbf{F}\bar{\mathbf{Q}}_k^n\mathbf{F}^H+\mathbf{R}_k\right)^{-1}\mathbf{F}\bar{\mathbf{Q}}_k^n\right\}$. With DFT phase shift, we have $\mathbf{F}\mathbf{F}^H = S\mathbf{I}_S$. MSE can be reduced to

$$\sum_{k=1}^{K}\mathrm{tr}\{\bar{\mathbf{Q}}_k^n\} - S\sum_{k=1}^{K}\mathrm{tr}\left\{(\bar{\mathbf{Q}}_k^n)^2\left(\mathbf{F}\bar{\mathbf{Q}}_k^n\mathbf{F}^H+\mathbf{R}_k\right)^{-1}\right\}. \qquad (28)$$

Note that in (28), the second term is still difficult to handle due to the matrices $\bar{\mathbf{Q}}_k^n$ and $\mathbf{R}_k$. However, we can obtain some asymptotic results under a special case, i.e., an orthogonal pilot allocation case. From Proposition 1, $\mathbf{R}_k^n$ is reduced to $M\sigma^2/p\mathbf{I}_S$ under orthogonal pilot allocation. The second term in (28) can be rewritten as

$$\begin{aligned}
&S\sum_{k=1}^{K}\mathrm{tr}\left\{(\bar{\mathbf{Q}}_k^n)^2\left(\mathbf{F}\bar{\mathbf{Q}}_k^n\mathbf{F}^H+M\sigma^2/p\mathbf{I}_S\right)^{-1}\right\}\\
&=S\sum_{k=1}^{K}\mathrm{tr}\left\{(\bar{\mathbf{Q}}_k^n)^2\left(\mathbf{F}\bar{\mathbf{Q}}_k^n\mathbf{F}^H+\frac{M\sigma^2}{pS}\mathbf{F}\mathbf{F}^H\right)^{-1}\right\}\\
&=\sum_{k=1}^{K}\mathrm{tr}\left\{(\bar{\mathbf{Q}}_k^n)^2\left(\bar{\mathbf{Q}}_k^n+\frac{M\sigma^2}{pS}\mathbf{I}_S\right)^{-1}\right\}. \qquad (29)
\end{aligned}$$

Remark: When considering orthogonal pilot allocation, we have NMSE $\to 0$ in high pilot transmit SNR region and large S region. This is because (29) is reduced to $\sum_{k=1}^{K}\mathrm{tr}\{\bar{\mathbf{Q}}_k^n\}$ when $\frac{\sigma^2}{p}\to 0$ or $S\to\infty$. Then the MSE derived in (28) is equal to zero. Combining (26)-(27), we have zero NMSE.

IV. Extended Kalman Filter under Hardware Impairments

A. Hardware Impairments

In practical communication scenarios, the hardware is generally non-ideal, and therefore the received signal is affected by the HWI. In this section, we consider three different types of hardware impairments in the following, including HWI at transceiver and RIS[5].

- The distortion noise at each user and the BS: They can be assumed as Gaussian distributed variables with average power being proportional to the power of transmit and received signals.
- The multiplicative phase drifts at each user and the BS: They are caused by local oscillators and can be modeled by the Wiener process [21].
- The phase error at RIS: This type of hardware impairment originates from finite precision of configuration of RIS elements and can be modeled as phase noise. We assume that the phase noise is uniformly distributed in $[-k_r\pi, k_r\pi)$, where k_r denotes the severity of the residual impairments[6] at the RIS [22]. Since the phase shift could be interfered by the phase error, the optimization of phase shift is ineffective. Therefore, we do not consider the optimization of phase shifts in this section. During the prediction process, we adopt random phase shift at RIS.

Considering the additive distortion noise and multiplicative phase drift at the transceiver as well as the phase noise at the RIS [21], [37], [38], the received pilot signal at the BS can be written as

$$\tilde{\mathbf{y}}_j^n = \mathbf{\Psi}_j^n\sum_{k=1}^{K}\mathbf{H}_1\tilde{\mathbf{\Phi}}_j^n\mathbf{h}_{2,k}^n\left(\sqrt{p}\mathbf{s}_{k,j}^n+\boldsymbol{\delta}_{k,j}^n\right)+\mathbf{\Lambda}_j^n+\mathbf{w}_j^n. \qquad (30)$$

In (30), $\boldsymbol{\delta}_{k,j}^n\in\mathbb{C}^{1\times\tau}$ denotes the distortion noise at each user with each element distributed according to $\mathcal{CN}(0,\kappa_{UE}p)$. κ_{UE} denotes the proportionality coefficient that characterizes the level of impairment at each user. The distortion noise at the BS is denoted as $\mathbf{\Lambda}_j^n\in\mathbb{C}^{M\times\tau}$ with each column distributed according to $\boldsymbol{\lambda}_j^n\in\mathbb{C}^{M\times 1}\sim\mathcal{CN}(\mathbf{0},\mathbf{\Gamma})$ where $\mathbf{\Gamma}=\kappa_{BS}p\sum_{k=1}^{K}\mathbf{I}_M\odot\mathbf{h}_{k,j,n}^c(\mathbf{h}_{k,j,n}^c)^H$ and $\mathbf{h}_{k,j,n}^c\triangleq\mathbf{H}_1\tilde{\mathbf{\Phi}}_j^n\mathbf{h}_{2,k}^n$ [21], [22], [37]. κ_{BS} denotes the proportionality coefficient that characterizes the level of impairment at the BS. The phase shift matrix with phase noise at the RIS is given as $\tilde{\mathbf{\Phi}}_j^n = \mathrm{diag}([e^{j(\theta_{j,1}^n+\Delta\theta_{j,1}^n)},e^{j(\theta_{j,2}^n+\Delta\theta_{j,2}^n)},\cdots,e^{j(\theta_{j,S}^n+\Delta\theta_{j,S}^n)}])$, where $\theta_{j,s}^n\in[0,2\pi)$ denotes the phase shift of the sth element at the jth sub-phase of the nth time slot. The phase noise of the sth element at RIS is denoted as $\Delta\theta_{j,s}^n$. The multiplicative phase drift is defined as $\mathbf{\Psi}_j^n = \mathrm{diag}([e^{j\psi_{j,1}^n},e^{j\psi_{j,2}^n},\cdots,e^{j\psi_{j,M}^n}])$, where $\psi_{j,m}^n=\epsilon_j^n+\nu_{j,m}^n$. ϵ_j^n denotes the phase drift of each user at the jth sub-phase of the nth time slot. $\nu_{j,m}^n$ denotes the phase drift of the mth BS antenna at the jth sub-phase of the nth time slot. From [21], [37], [38], ϵ_j^n and $\nu_{j,m}^n$ can be modeled as a discrete-time independent Wiener process as follows:

$$\epsilon_j^n=\epsilon_{j-1}^n+\Delta\bar{\epsilon}_j^n,\quad \nu_{j,m}^n=\nu_{j-1,m}^n+\Delta\bar{\nu}_{j,m}^n, \qquad (31)$$

where $\Delta\bar{\epsilon}_j^n\sim\mathcal{N}(0,\sigma_\epsilon^2)$ and $\Delta\bar{\nu}_{j,m}^n\sim\mathcal{N}(0,\sigma_\nu^2)$ denote the random phase increment caused by the imperfect local oscillator at each user and the mth BS antenna, respectively. According to [21], [37], [38], the phase drifts at the transceiver

[5]In order to focus on the impact of HWI on the estimation accuracy, we do not consider pilot contamination in this section.

[6]We assume that the residual impairments here are the hardware impairments that have been partially mitigated by compensation algorithms. Due to the imprecise estimates of time-variant hardware characteristics and random noise, residual HWI still exists.

update every sub-phase.

B. Joint Estimation Based on Extended Kalman Filter

By employing the LS estimation method, we have

$$\begin{aligned}&\frac{\tilde{\mathbf{y}}_j^n\left(\mathbf{s}_{k,j}^n\right)^H}{\sqrt{p}}\\&=\boldsymbol{\Psi}_j^n\mathbf{H}_1\tilde{\boldsymbol{\Phi}}_j^n\mathbf{h}_{2,k}^n+\frac{\boldsymbol{\Psi}_j^n}{\sqrt{p}}\sum_{k=1}^K\mathbf{H}_1\tilde{\boldsymbol{\Phi}}_j^n\mathbf{h}_{2,k}^n\boldsymbol{\delta}_{k,j}^n\left(\mathbf{s}_{k,j}^n\right)^H\\&+\frac{\boldsymbol{\Lambda}_j^n\left(\mathbf{s}_{k,j}^n\right)^H}{\sqrt{p}}+\frac{\mathbf{w}_j^n\left(\mathbf{s}_{k,j}^n\right)^H}{\sqrt{p}}\\&\triangleq\boldsymbol{\Psi}_j^n\mathbf{G}_k^n\tilde{\boldsymbol{\phi}}_j^n+\mathbf{d}_{k,j}^n,\end{aligned}\tag{32}$$

where $\tilde{\boldsymbol{\phi}}_j^n\triangleq[e^{j(\theta_{j,1}^n+\Delta\theta_{j,1}^n)},e^{j(\theta_{j,2}^n+\Delta\theta_{j,2}^n)},\cdots,e^{j(\theta_{j,S}^n+\Delta\theta_{j,S}^n)}]^T$ and $\mathbf{d}_{k,j}^n\triangleq\frac{\boldsymbol{\Psi}_j^n}{\sqrt{p}}\sum_{k=1}^K\mathbf{H}_1\tilde{\boldsymbol{\Phi}}_j^n\mathbf{h}_{2,k}^n\boldsymbol{\delta}_{k,j}^n\left(\mathbf{s}_{k,j}^n\right)^H+\frac{\boldsymbol{\Lambda}_j^n\left(\mathbf{s}_{k,j}^n\right)^H}{\sqrt{p}}+\frac{\mathbf{w}_j^n\left(\mathbf{s}_{k,j}^n\right)^H}{\sqrt{p}}$.

In order to facilitate the derivations later, we define a new variable as $\mathbf{g}_k^n\triangleq\mathrm{vec}(\mathbf{G}_k^n)$. $\mathrm{vec}(\cdot)$ is the vectorization operation. By utilizing $\mathbf{g}_k^n$, we can rewrite $\mathbf{G}_k^n\tilde{\boldsymbol{\phi}}_j^n$ as $((\tilde{\boldsymbol{\phi}}_j^n)^T\otimes\mathbf{I}_M)\mathbf{g}_k^n$. Therefore, (32) can be rewritten as

$$\frac{\tilde{\mathbf{y}}_j^n\left(\mathbf{s}_{k,j}^n\right)^H}{\sqrt{p}}=\boldsymbol{\Psi}_j^n\mathbf{F}_j^n\mathbf{g}_k^n+\mathbf{d}_{k,j}^n,\tag{33}$$

where $\mathbf{F}_j^n\triangleq(\tilde{\boldsymbol{\phi}}_j^n)^T\otimes\mathbf{I}_M$.

Following the same procedure in the previous section, we stack the vectorized signal $\frac{\tilde{\mathbf{y}}_j^n\left(\mathbf{s}_{k,j}^n\right)^H}{\sqrt{p}}$ for all the sub-phases as follows:

$$\tilde{\mathbf{Y}}_k^n=\tilde{\boldsymbol{\Psi}}^n\tilde{\mathbf{F}}^n\mathbf{g}_k^n+\mathbf{D}_k^n,\tag{34}$$

where

$$\begin{aligned}\tilde{\mathbf{Y}}_k^n=\Bigg[&\left(\frac{\tilde{\mathbf{y}}_1^n\left(\mathbf{s}_{k,1}^n\right)^H}{\sqrt{p}}\right)^T,\left(\frac{\tilde{\mathbf{y}}_2^n\left(\mathbf{s}_{k,2}^n\right)^H}{\sqrt{p}}\right)^T,\cdots,\\&\left(\frac{\tilde{\mathbf{y}}_J^n\left(\mathbf{s}_{k,J}^n\right)^H}{\sqrt{p}}\right)^T\Bigg]^T,\end{aligned}\tag{35}$$

$$\tilde{\boldsymbol{\Psi}}^n=\mathrm{diag}\left(\boldsymbol{\Psi}_1^n,\boldsymbol{\Psi}_2^n,\cdots,\boldsymbol{\Psi}_J^n\right),\tag{36}$$

$$\tilde{\mathbf{F}}^n=\left[(\mathbf{F}_1^n)^T,(\mathbf{F}_2^n)^T,\cdots,(\mathbf{F}_J^n)^T\right]^T,\tag{37}$$

$$\mathbf{D}_k^n=\left[\left(\mathbf{d}_{k,1}^n\right)^T,\left(\mathbf{d}_{k,2}^n\right)^T,\cdots,\left(\mathbf{d}_{k,J}^n\right)^T\right]^T.\tag{38}$$

By stacking (34) for all users into a large vector, we derive the measurement equation as follows:

$$\hat{\mathbf{Y}}^n=\hat{\boldsymbol{\Psi}}^n\hat{\mathbf{F}}^n\mathbf{g}^n+\mathbf{D}^n,\tag{39}$$

where $\hat{\mathbf{Y}}^n=[(\tilde{\mathbf{Y}}_1^n)^T,(\tilde{\mathbf{Y}}_2^n)^T,\cdots,(\tilde{\mathbf{Y}}_K^n)^T]^T\in\mathbb{C}^{JMK\times1}$, $\hat{\boldsymbol{\Psi}}^n=\mathrm{diag}(\tilde{\boldsymbol{\Psi}}^n,\tilde{\boldsymbol{\Psi}}^n,\cdots,\tilde{\boldsymbol{\Psi}}^n)\in\mathbb{C}^{JMK\times JMK}$, $\hat{\mathbf{F}}^n=\mathrm{diag}(\tilde{\mathbf{F}}^n,\tilde{\mathbf{F}}^n,\cdots,\tilde{\mathbf{F}}^n)\in\mathbb{C}^{JMK\times SMK}$, $\mathbf{g}^n=[(\mathbf{g}_1^n)^T,(\mathbf{g}_2^n)^T,\cdots,(\mathbf{g}_K^n)^T]^T\in\mathbb{C}^{SMK\times1}$, and $\mathbf{D}^n=[(\mathbf{D}_1^n)^T,(\mathbf{D}_2^n)^T,\cdots,(\mathbf{D}_K^n)^T]^T\in\mathbb{C}^{JMK\times1}$.

From (39), the phase drift is also integrated into the measurement matrix. Normally, the phase drift is unknown to the BS, which makes the estimation challenging[7]. Fortunately, we observe that the phase drift in (31) has a similar update rule as the cascaded channel. This motivates us to jointly estimate the cascaded channel and the phase drift. From [39], [40], the phase drift updates every sub-phase, i.e., every pilot sequence length. While in state equation (18), the cascaded channel updates every time slot, which means the phase drift at the transceiver updates more frequently than the channel. In order to describe the overall state equation in a unified manner, we first present the following update equations of the phase drifts by recursively updating the equations in (31)[8].

$$\epsilon_j^n=\epsilon_j^{n-1}+\Delta\epsilon_j^n,\quad\nu_{j,m}^n=\nu_{j,m}^{n-1}+\Delta\nu_{j,m}^n,\tag{40}$$

where $\Delta\epsilon_j^n\sim\mathcal{N}(0,J\sigma_\epsilon^2)$ and $\Delta\nu_{j,m}^n\sim\mathcal{N}(0,J\sigma_\nu^2)$ [39], [40].

Define $\boldsymbol{\epsilon}^n\triangleq[\epsilon_1^n,\epsilon_2^n,\cdots,\epsilon_J^n]^T$ and $\boldsymbol{\nu}^n\triangleq[\nu_{1,1}^n,\cdots,\nu_{J,1}^n,\cdots,\nu_{1,M}^n\cdots,\nu_{J,M}^n]^T$. We obtain the update equation of the phase drift in a vector form as follows:

$$\boldsymbol{\epsilon}^n=\boldsymbol{\epsilon}^{n-1}+\Delta\boldsymbol{\epsilon},\quad\boldsymbol{\nu}^n=\boldsymbol{\nu}^{n-1}+\Delta\boldsymbol{\nu},\tag{41}$$

where $\Delta\boldsymbol{\epsilon}\triangleq[\Delta\epsilon_1,\Delta\epsilon_2,\cdots,\Delta\epsilon_J]^T$ and $\Delta\boldsymbol{\nu}=[\Delta\nu_{1,1},\cdots,\Delta\nu_{J,1},\cdots,\Delta\nu_{1,M},\cdots,\Delta\nu_{J,M}]^T$.

Next, we present the new state equation of the vectorized channel vector $\mathbf{g}_k^n$ as follows:

$$\mathbf{g}_k^n=\tilde{\mathbf{A}}_k\mathbf{g}_k^n+\mathbf{z}_k^n,\tag{42}$$

where $\tilde{\mathbf{A}}_k\triangleq a_k\mathbf{I}_{SM}$ and $\mathbf{z}_k^n\triangleq\mathrm{rvec}(\mathbf{U}_k^n)$. $\mathrm{rvec}(\cdot)$ denotes the row vectorization of a matrix.

Combining (41)-(42), we rearrange the overall state equation as $\boldsymbol{\Upsilon}^n=\tilde{\mathbf{A}}\boldsymbol{\Upsilon}^{n-1}+\mathbf{Z}^n$, where the new state vector $\boldsymbol{\Upsilon}^n$ is defined by stacking the phase drift and channel variables into one vector as follows:

$$\boldsymbol{\Upsilon}^n\triangleq\left[(\mathbf{g}_1^n)^T,(\mathbf{g}_2^n)^T,\cdots,(\mathbf{g}_K^n)^T,(\boldsymbol{\epsilon}^n)^T,(\boldsymbol{\nu}^n)^T\right]^T,\tag{43}$$

$$\tilde{\mathbf{A}}=\mathrm{diag}\left(\tilde{\mathbf{A}}_1,\tilde{\mathbf{A}}_2,\cdots,\tilde{\mathbf{A}}_K,\mathbf{I}_J,\mathbf{I}_{JM}\right),\tag{44}$$

$$\mathbf{Z}^n=\left[(\mathbf{z}_1^n)^T,(\mathbf{z}_2^n)^T,\cdots,(\mathbf{z}_K^n)^T,(\Delta\boldsymbol{\epsilon})^T,(\Delta\boldsymbol{\nu})^T\right]^T.\tag{45}$$

By utilizing the new state vector $\boldsymbol{\Upsilon}^n$ and defining $\boldsymbol{\varphi}(\boldsymbol{\Upsilon}^n)=\hat{\boldsymbol{\Psi}}^n\hat{\mathbf{F}}^n\mathbf{g}^n$, the measurement equation (39) can be rewritten as $\hat{\mathbf{Y}}^n=\boldsymbol{\varphi}(\boldsymbol{\Upsilon}^n)+\mathbf{D}^n$. Since the phase drifts need to be jointly estimated with the cascaded channel, $\boldsymbol{\varphi}(\boldsymbol{\Upsilon}^n)$ is a nonlinear function of the state vector $\boldsymbol{\Upsilon}^n$. The nonlinearity of the measurement equation is caused by the multiplicative phase shift matrix $\hat{\boldsymbol{\Psi}}^n$. Specifically, the overall state vector $\boldsymbol{\Upsilon}^n$ in (43) consists of the channel vector $\mathbf{g}^n$ and the phase drift vector

[7]Note that in (39) the phase noise vector is integrated into the matrix $\hat{\mathbf{F}}^n$, which is part of the measurement matrix. Generally, the phase noise at RIS is modeled as a random variable following a uniform distribution [22], [23]. The BS only has the information of phase shift but not the information of phase noise because the phase noise is a random variable. Therefore, when applying the Kalman filter to estimate the channel, the BS uses the designed phase shift without phase noise to estimate the cascaded channel although there will be NMSE performance degradation.

[8]Since each time slot is divided into J sub-phases, the new update equations in (40) are derived recursively by means of (31) with J times.

$[(\boldsymbol{\epsilon}^n)^T, (\boldsymbol{\nu}^n)^T]^T$. While the function $\boldsymbol{\varphi}(\boldsymbol{\Upsilon}^n) = \hat{\boldsymbol{\Psi}}^n \hat{\mathbf{F}}^n \mathbf{g}^n$ is the product of the channel vector $\mathbf{g}^n$ and the phase drift matrix $\hat{\boldsymbol{\Psi}}^n$. Since the measurement equation is nonlinear, the Kalman filter cannot be directly applied. One approach to solve the nonlinear problem is the extended Kalman filter. Through this approach, we first need to linearize the nonlinear equation by applying the first-Taylor approximation. Then we can apply KF to the linearized model.

The linearization requires computing a Jacobian matrix. Since the new state vector can be rewritten as $\boldsymbol{\Upsilon}^n = [(\mathbf{g}^n)^T, (\boldsymbol{\epsilon}^n)^T, (\boldsymbol{\nu}^n)^T]^T$, the Jacobian matrix of the nonlinear function $\boldsymbol{\varphi}$ with respect to $\boldsymbol{\Upsilon}^n$ and evaluated at $\boldsymbol{\Upsilon}^{n|n-1}$ can be computed as

$$\begin{aligned}\boldsymbol{\Omega}^n &= \frac{\partial \boldsymbol{\varphi}(\boldsymbol{\Upsilon}^n)}{\partial \boldsymbol{\Upsilon}^n}\Big|_{\boldsymbol{\Upsilon}^n=\boldsymbol{\Upsilon}^{n|n-1}} \\ &= \left[\frac{\partial \boldsymbol{\varphi}(\boldsymbol{\Upsilon}^n)}{\partial \mathbf{g}^n}, \frac{\partial \boldsymbol{\varphi}(\boldsymbol{\Upsilon}^n)}{\partial \boldsymbol{\epsilon}^n}, \frac{\partial \boldsymbol{\varphi}(\boldsymbol{\Upsilon}^n)}{\partial \boldsymbol{\nu}^n}\right] \\ &= \left[\hat{\boldsymbol{\Psi}}^n \hat{\mathbf{F}}^n, \frac{\partial \boldsymbol{\varphi}(\boldsymbol{\Upsilon}^n)}{\partial \boldsymbol{\epsilon}^n}, \frac{\partial \boldsymbol{\varphi}(\boldsymbol{\Upsilon}^n)}{\partial \boldsymbol{\nu}^n}\right]. \end{aligned} \tag{46}$$

With the derived Jacobian matrix, we can apply extended Kalman filter to jointly estimate the phase drift and the cascaded channel, as shown in Algorithm 2 below, where $\tilde{\mathbf{Q}}^n$ and $\tilde{\mathbf{R}}^n$ denote the covariance matrices of the state noise $\mathbf{Z}^n$ and the measurement noise $\mathbf{D}^n$, respectively. The expressions of $\tilde{\mathbf{Q}}^n$ and $\tilde{\mathbf{R}}^n$ can be derived following a similar method as in Appendix A.

Algorithm 2 Joint Channel and HWI Prediction Based on Extended Kalman Filter

1: **Initialization**: $\mathbf{g}^0 = \mathbf{0}_{SMK\times 1}$, $\boldsymbol{\epsilon}^0$, $\boldsymbol{\nu}^0$, $\boldsymbol{\Upsilon}^0 = [(\mathbf{g}^0)^T, (\boldsymbol{\epsilon}^0)^T, (\boldsymbol{\nu}^0)^T]^T$, $\mathbf{P}^0 = \mathbf{I}_{SMK+J+JM}$, $n = 1$, and the total number of time slots N.
2: **repeat**
3: **Prediction**: $\boldsymbol{\Upsilon}^{n|n-1} = \tilde{\mathbf{A}} \times \boldsymbol{\Upsilon}^{n-1|n-1}$ and $\mathbf{P}^{n|n-1} = \tilde{\mathbf{A}}\mathbf{P}^{n-1|n-1}\tilde{\mathbf{A}}^H + \tilde{\mathbf{Q}}^n$.
4: **Kalman Gain**: $\mathbf{S}^n = \boldsymbol{\Omega}^n \mathbf{P}^{n|n-1}(\boldsymbol{\Omega}^n)^H + \tilde{\mathbf{R}}^n$, $\mathbf{K}^n = \mathbf{P}^{n|n-1}(\boldsymbol{\Omega}^n)^H(\mathbf{S}^n)^{-1}$.
5: **Correction**: $\boldsymbol{\Upsilon}^{n|n} = \boldsymbol{\Upsilon}^{n|n-1} + \mathbf{K}^n(\hat{\mathbf{Y}}^n - \boldsymbol{\varphi}(\boldsymbol{\Upsilon}^{n|n-1}))$ and $\mathbf{P}^{n|n} = \mathbf{P}^{n|n-1} - \mathbf{K}^n\boldsymbol{\Omega}^n\mathbf{P}^{n|n-1}$.
6: Set $n = n + 1$.
7: **until** $n = N$

V. SIMULATION RESULTS

In this section, we evaluate the NMSE performance of the proposed algorithms through simulations. Numerical results are obtained using MATLAB, and the results are averaged over 100 independent channel realizations. The total number of time slots is $N = 200$. We consider a two-dimensional (2D) plane network where the BS and RIS are located at $(-10, 0)$ and $(0, 10)$, respectively. In addition, K single-antenna users are uniformly distributed in a $20\text{m} \times 10\text{m}$ rectangle area. The path loss models from the BS to RIS and from RIS to users are $30 + 22\log_{10}(d(m))$, where d denotes the distance. The AoAs/AoDs are uniformly distributed within $[0, \frac{1}{2}\pi]$ for the elevation angle and within $[0, 2\pi]$ for the azimuth angle. The Rician factor is 10 dB. The coherence time of the BS-RIS channel is 10 times that of the RIS-user channel, i.e., $T_H = 10T_h$. The Doppler shift for different users is uniformly distributed as $0.01 + 0.001 \times \mathcal{U}(0, 1)$. The noise power is set as $\sigma^2 = -120$ dBm. The number of users K, the number of BS antennas M, the number of RIS elements S, and the severity of different types of HWI will be specified in each experiment. The NMSE of KF and EKF algorithms can be computed as $\text{NMSE}_{\text{KF}} = \mathbb{E}\left[\frac{\|\mathbf{G}^{\text{n|n}} - \mathbf{G}^{\text{n}}\|_{\text{F}}^2}{\|\mathbf{G}^{\text{n}}\|_{\text{F}}^2}\right]$ and $\text{NMSE}_{\text{EKF}} = \mathbb{E}\left[\frac{\|\mathbf{g}^{\text{n|n}} - \mathbf{g}^{\text{n}}\|^2}{\|\mathbf{g}^{\text{n}}\|^2}\right]$, where $\|\cdot\|_F$ and $\|\cdot\|$ represent the Frobenius norm and the l_2 norm, respectively.

We denote the proposed Algorithm 1 and Algorithm 2 as "Proposed KF algorithm" and "Proposed EKF algorithm", respectively. The "Proposed KF algorithm" adopts the DFT matrix as the phase shift. We evaluate the performance of the proposed algorithms by comparing them with the following benchmark schemes:

- Unitary phase [17]: In this scheme, the authors consider a single-user case and do not consider pilot contamination. Kalman filter is used to estimate the cascaded channel. Therefore, we iteratively estimate the cascaded channel for each user. Also, Rayleigh fading is considered in this paper. The optimal phase shift in this case is any unitary matrix that satisfies the power constraint.
- LMMSE [13]: In [13], the authors consider multiple users, pilot contamination, and Rayleigh fading. Since the optimal phase shift design in the scheme is equal phase shifts, we replace the equal phase shifts with varying random phase shifts for a fair comparison, and thus the BS can obtain sufficient observations. Particularly, instead of estimating the cascaded channel of the BS-RIS-user link, the authors estimate the aggregate channel that includes phase shift. The authors adopt the LMMSE estimator to estimate the aggregate channel.
- Random phase: This scheme adopts the same Kalman filter as the "Proposed KF algorithm", but the RIS phase shifts are randomly generated from $[0, 2\pi]$ during the whole estimation process.
- LMMSE, three phase [14]: The authors of [14] proposed a three-phase channel estimation framework, in which the direct link and the cascaded channel of a typical user are estimated in phases I and II, respectively. The cascaded channels of the rest of users are sequentially estimated in phase III by utilizing the common BS-RIS channel. When each user transmits a pilot symbol during phase III, only selected RIS elements are switched on to reflect the pilot symbol and thus multiple time slots are needed to complete the estimation. When comparing with this framework, the estimation of the direct link is neglected to ensure consistency.
- LMMSE, ON/OFF [24]: When considering hardware impairments, we compare our proposed EKF algorithm with this scheme to show the advantages. In this scheme, the distortion noise and phase noise are considered while the phase drift is not considered. We will show that this benchmark scheme is inferior to the proposed scheme even though the phase drift is not taken into

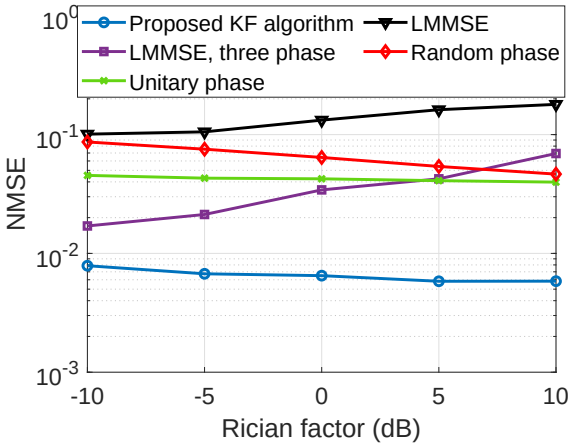

Fig. 3. NMSE versus Rician factor when $p = 10$ dB, $K = 25$, $\tau = 20$, $S = 20$, $M = 8$ and no HWI.

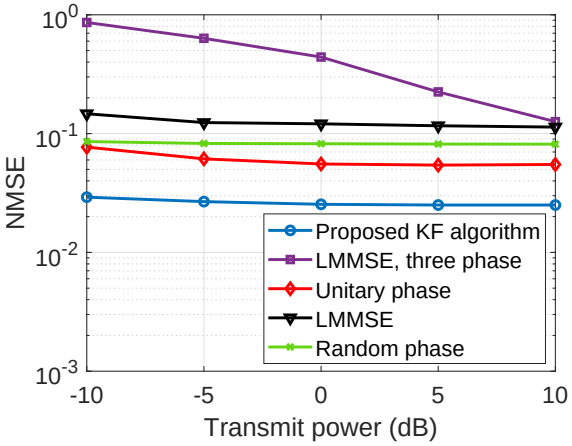

Fig. 4. NMSE versus transmit power of pilots when $K = 25$, $\tau = 20$, $S = 20$, $M = 8$ and no HWI.

consideration. Particularly, the authors use the ON/OFF switch method and the LMMSE method to consecutively estimate the cascaded channel of each RIS element.

A. Without HWI

Fig. 3 shows the NMSE performance versus the Rician factor. The RIS adopts a 5×4 UPA. The NMSE of the proposed KF algorithm and the random phase scheme improve when the Rician factor increases. This is because when κ becomes larger, the LoS components of BS-RIS and RIS-user channels become more dominant, thus enhancing the received signal power at the BS. The NMSE of the Unitary phase scheme slightly decreases when the Rician factor increases. This is because [17] assumes Rayleigh fading, and thus the Rician factor is not involved in the KF algorithm and its phase shift design. On the contrary, the NMSE of the LMMSE scheme and the "LMMSE, three phase" scheme increases with Rician factor. This can be explained as follows. Both two schemes adopt the LMMSE estimator in which the covariance matrices of the BS-RIS and RIS-user channels are utilized to compute the second moment of the cascaded channel. To adapt to our scenario, the Rician factor is involved in the covariance matrices. However, the two schemes consider Rayleigh fading assumption. The LMMSE estimator obtained under Rayleigh fading does not work well in the case of Rician fading. Furthermore, the LMMSE-based schemes are inferior to all the Kalman-based schemes. This is reasonable because the LMMSE-based schemes do not have the information on time correlation that can be exploited to improve the channel estimation.

Fig. 4 demonstrates the NMSE performance versus the transmit power of pilot sequences. It can be observed that the NMSE performance of all the schemes improves with the increase in p. The proposed KF algorithm achieves a much better NMSE performance over other benchmark schemes even when suffering from pilot contamination. This is because the time-varying characteristic is exploited by the proposed algorithm. The performance gaps between the proposed algorithm, the benchmark [17], and the random phase scheme

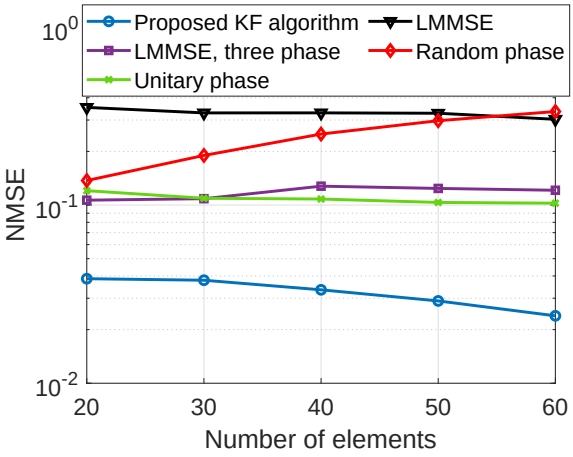

Fig. 5. NMSE versus the number of elements when $p = 10$ dB, $K = 12$, $M = 8$, $\tau = M$, and no HWI.

show the superiority of the DFT matrix over unitary phase shift and random phase shift. The "LMMSE, three phase" scheme achieves the worst performance. The reason is that only selected RIS elements are switched on during estimation and the full RIS aperture gain is not utilized.

In Fig. 5, we compare the NMSE performance in terms of the number of elements. The RIS is an $S_H \times 10$ UPA, where S_H increases from 2 to 6. The proposed KF algorithm outperforms all the other benchmarks, which indicates the benefit of the DFT matrix and the utilization of time-varying characteristics. The NMSE performance of the proposed KF algorithm, the "Unitary phase", and the LMMSE scheme improve with the number of elements, especially for the proposed KF algorithm. The NMSE performance of the "Random phase" scheme degrades when the number of elements increases. The reason is that the observations obtained with random phase shift are not sufficient for the estimation, and the contaminated signal significantly increases when S increases. It is seen that the NMSE of the "LMMSE, three phase" scheme first slightly increases, and then varies a little after $N > 40$. This is reasonable because when S increases, each user needs

TABLE I
PILOT OVERHEAD COMPARISON

Scheme	Proposed protocol	Protocol in [17]	Protocol in [13]	Protocol in [14]
Overhead	$S\tau$	SK	τ	$K+S+(K-1)\left\lceil\frac{S}{M}\right\rceil$

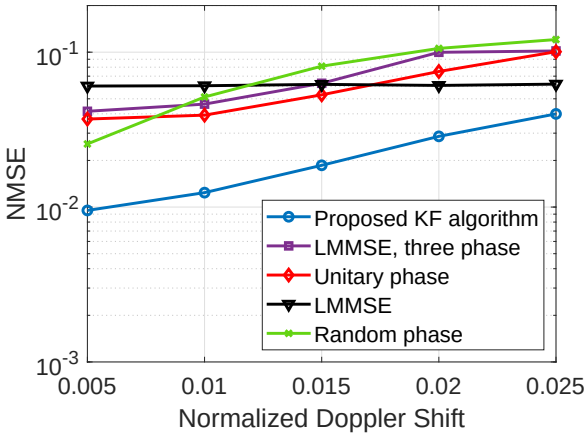

Fig. 6. NMSE versus the number of elements when $p = 10$ dB, $K = 25$, $\tau = S = 20$, $M = 8$ and no HWI.

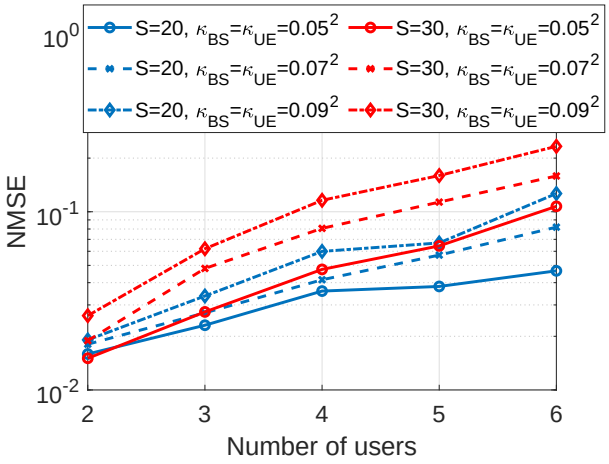

Fig. 7. NMSE of the proposed EKF algorithm versus the number of users when $M = 8$, $p = 10$ dB and with HWI.

more time slots to complete the estimation of the cascaded channels of all the RIS elements. As the estimation at each time slot is independent, the total estimation error of each user accumulates when the number of time slots increases.

Fig. 6 illustrates the NMSE versus the normalized Doppler shift. A higher Doppler shift means a higher moving speed of users. When the speed of users increases, as expected, the channel tracking becomes difficult and the estimation accuracy decreases. From Fig. 6, the NMSE of all the schemes except the "LMMSE" scheme increases with the increase in the normalized Doppler shift. Although both the "LMMSE" scheme and the "LMMSE, three phase" scheme assume Rayleigh fading and use the same LMMSE estimator, the two schemes show different trends. The reason is that the "LMMSE" scheme only utilizes the variance matrices of the BS-RIS and RIS-user channels. As a result, the NMSE of the "LMMSE" scheme does not change with the Doppler shift. The "LMMSE, three phase" scheme utilizes the covariance matrix of the RIS-user channel, and thus the channel dynamics are taken into consideration.

Table I compares the total pilot overhead consumed for estimating the channel in one round. It can be observed that the presented protocol consumes fewer pilot symbols than the protocol proposed in [17] when $\tau < K$. The reduced pilot overhead is equal to $S(K-\tau)$ which is considerable when the number of elements is large. The protocol in [13] only consumes τ pilot symbols because the BSs estimate the aggregate channel integrated with the fixed phase shift matrix. However, the estimated aggregate channel cannot be utilized when the phase shift is changing. The protocol in [14] also consumes fewer pilot symbols than the proposed protocol. This is because they utilize the strong correlation between different users due to the common BS-RIS channel. However, the protocol is developed based on the assumption that the cascaded channel of the typical user is perfectly estimated, which is not practical.

B. With HWI

In the following, we compare the NMSE performance in the presence of HWI. The severity of the residual impairments at the RIS is fixed to $k_r = 0.08$ [41]. Fig. 7 indicates the NMSE performance of the proposed EKF algorithm versus the number of users under different levels of distortion noise and various numbers of elements. We observe that when the number of users becomes larger, the NMSE increases under different values of S and different levels of HWI. This is because the received pilot signal at the BS is impaired by the increased phase drift and distortion noise at users when K increases. The NMSE performance degrades when the severity of distortion noise at the BS and each user grows. Different from Fig. 5, a larger number of elements induces performance degradation due to the increased phase error at RIS.

Fig. 8 shows the NMSE versus the number of BS antennas M under different levels of distortion noise and phase drift. When M increases, the performance degrades due to the increased distortion noise at the BS. When the phase drift is fixed to $\sigma_\epsilon^2 = \sigma_\nu^2 = 2.47 \times 10^{-5}$, the NMSE degrades when the distortion noises at the BS and users, i.e., κ_{BS} and κ_{UE}, increase from 0.05^2 to 0.09^2. Similar observations can be found when $\sigma_\epsilon^2 = \sigma_\nu^2 = 4.94 \times 10^{-5}$. Furthermore, the curves with the same mark, i.e., with the same level of distortion noise at the BS and users, achieve worse NMSE performance when the phase drift becomes more severe.

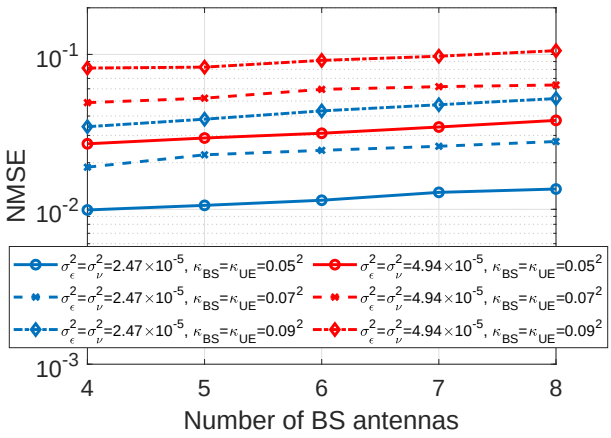

Fig. 8. NMSE of the proposed EKF algorithm versus number of antennas when $K = 2$, $p = 10$ dB, $S = 20$ and with HWI.

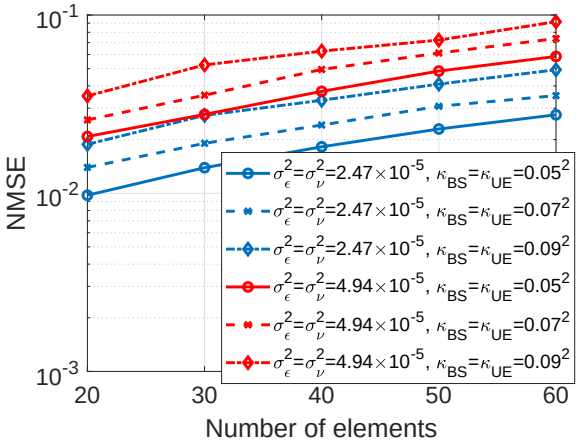

Fig. 9. NMSE of the proposed EKF algorithm versus number of elements when $K = 2$, $p = 10$ dB, $M = 8$ and with HWI.

In Fig. 9, we show the NMSE versus the number of elements under different levels of distortion noise and phase drift. Different from the proposed KF algorithm in Fig. 5, the NMSE of the proposed EKF algorithm increases with S as HWI becomes more severe. The estimation accuracy is largely impacted by the phase error at RIS. Similar to Fig. 8, the NMSE performance further degrades when the phase drift and distortion noise become more severe.

In Fig. 10, we compare the proposed EKF algorithm with the scheme proposed in [24] in terms of transmit power. Both algorithms can achieve a better NMSE performance when the transmit power increases. Both algorithms reach error floors around $p = 10$ dB due to the HWI-induced interference. The NMSE performance degrades when the severity of the distortion noise increases. The error floors go higher when increasing the severity of the distortion noise. It is observed that there is a large performance gap between the two algorithms. The proposed algorithm achieves a much better performance under different values of κ_{BS} and κ_{UE}. This is reasonable because the "LMMSE, ON/OFF" scheme adopts the ON/OFF scheme without fully utilizing the RIS aperture. The figures

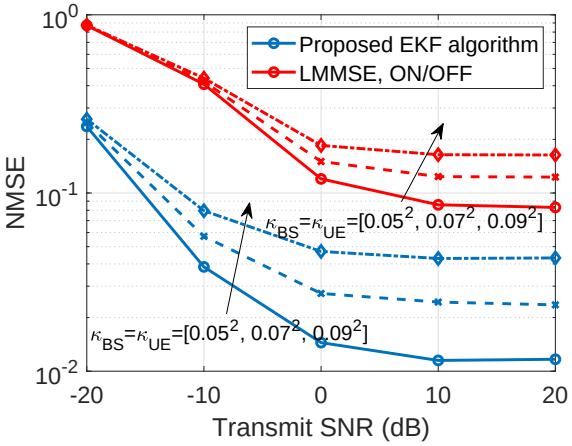

Fig. 10. NMSE versus transmit power of pilots when $K = 2$, $M = 8$, $S = 20$ and with HWI.

above indicate that the RIS-assisted system has poor NMSE performance under lower hardware quality.

VI. Conclusion

In this paper, we have investigated channel estimation in an RIS-assisted multi-user network. Considering users are mobile, we utilized KF to track the cascaded channel based on time correlation. We derived the expressions of state and measurement covariance matrices in the case of Rician fading and pilot contamination. We adopted the DFT matrix as the phase shift matrix for the sake of practicality and presented NMSE analysis and some asymptotic results. We then extended the channel estimation problem to a more practical scenario where three different kinds of hardware impairments are considered. Due to the non-linearity of HWI, we proposed to jointly estimate the cascaded channel and HWI by EKF. Extensive numerical results demonstrated the superiority of the proposed algorithms over benchmark schemes and provided valuable insights into the effect of various HWI on estimation accuracy. Without HWI, it can be observed that the NMSE can be improved by increasing the transmit power of pilot sequences and the number of elements. When the LoS of RIS-BS and RIS-user links are strengthened, the estimation accuracy can be enhanced. When the moving speed of users increases, the ability to track the channel deteriorates. The proposed EKF algorithm is able to track the cascaded channel well even with severe HWI. When the number of users, RIS elements, and BS antennas increase, the HWI at the transceiver and RIS becomes more severe resulting in the degradation of estimation accuracy.

Appendix A
Proof of Lemma 1

Based on the definition of $\mathbf{Q}^n$, it can be calculated as $\mathbf{Q}^n = \mathrm{Cov}(\mathbf{U}^{\mathrm{n}}, \mathbf{U}^{\mathrm{n}})$ and $\mathrm{Cov}(\cdot)$ takes the covariance. For simplicity, we first calculate the covariance matrix for the kth user, i.e., $\mathbf{Q}_k^n$. $\mathbf{Q}^n$ can be obtained by stacking the covariance matrices of all the users. For the kth user, $\mathbf{Q}_k^n$ can be calculated as

$$\mathbf{Q}_k^n = \mathrm{Cov}(\mathbf{U}_{\mathrm{k}}^{\mathrm{n}}, \mathbf{U}_{\mathrm{k}}^{\mathrm{n}})$$

$$= \mathbb{E}\left\{\text{diag}\left(\mathbf{u}_k^n\right)\mathbf{H}_1^T\left[\text{diag}\left(\mathbf{u}_k^n\right)\mathbf{H}_1^T\right]^H\right\} - \mathbb{E}\left\{\text{diag}\left(\mathbf{u}_k^n\right)\mathbf{H}_1^T\right\}\left[\mathbb{E}\left\{\text{diag}\left(\mathbf{u}_k^n\right)\mathbf{H}_1^T\right\}\right]^H, \quad (47)$$

where $\mathbb{E}\left\{\text{diag}\left(\mathbf{u}_k^n\right)\mathbf{H}_1^T\right\} = (1-a_k)\frac{\kappa}{\kappa+1}\sqrt{\beta_{Ru}(k)}\sqrt{\beta_{BR}}\,\text{diag}(\bar{\mathbf{h}}_{2,k})(\bar{\mathbf{H}}_1)^T$.

Define $\mathbf{X}_k \triangleq \text{diag}\left(\mathbf{u}_k^n\right)\mathbf{H}_1^T$ and $\mathbf{X}_k^{\text{cov}} \triangleq \mathbb{E}\left\{\mathbf{X}_k\mathbf{X}_k^H\right\}$. The element at the ith row and the jth column of the matrix $\mathbf{X}_k^{\text{cov}}$ is given by

$$[\mathbf{X}_k^{\text{cov}}]_{i,j} = \begin{cases} \sum_{m=1}^{M}\mathbb{E}\{[\mathbf{X}_k]_{i,m}\}\mathbb{E}\{[\mathbf{X}_k]_{j,m}^*\}, & i \neq j, \\ \sum_{m=1}^{M}\left(\mathbb{V}\{[\mathbf{X}_k]_{i,m}\} + \mathbb{E}\{[\mathbf{X}_k]_{i,m}\}\mathbb{E}^*\{[\mathbf{X}_k]_{i,m}\}\right), & i = j. \end{cases} \quad (48)$$

By expanding $\mathbf{u}_k^n$ and $\mathbf{H}_1$ into $\mathbf{u}_k^n = [u_{k,1}^n, \cdots, u_{k,S}^n]^T$ and $\mathbf{H}_1 = [H_{1,1}, \cdots, H_{1,S}; \cdots; H_{M,1}, \cdots, H_{M,S}]$, respectively, we have $[\mathbf{X}_k]_{i,m} = u_{k,i}^n H_{m,i}$. Particularly, $\mathbb{E}\{[\mathbf{X}_k]_{i,m}\}$ and $\mathbb{V}\{[\mathbf{X}_k]_{i,m}\}$ are given by

$$\mathbb{E}\{[\mathbf{X}_k]_{i,m}\} = \mathbb{E}\{u_{k,i}^n\}\mathbb{E}\{H_{m,i}\} = (1-a_k)\sqrt{\frac{\kappa}{\kappa+1}}\sqrt{\beta_{Ru}(k)}[\bar{\mathbf{h}}_{2,k}]_i\sqrt{\frac{\kappa}{\kappa+1}}\sqrt{\beta_{BR}}[\bar{\mathbf{H}}_1]_{m,i}, \quad (49)$$

$$\mathbb{V}\{[\mathbf{X}_k]_{i,m}\} = \mathbb{V}\{u_{k,i}^n\}\mathbb{V}\{H_{m,i}\} + \mathbb{V}\{u_{k,i}^n\}\mathbb{E}^2\{H_{m,i}\} + \mathbb{V}\{H_{m,i}\}\mathbb{E}^2\{u_{k,i}^n\}, \quad (50)$$

where $\mathbb{V}\{u_{k,i}^n\} = \left(1-a_k^2\right)\frac{1}{\kappa+1}\beta_{Ru}(k)$ and $\mathbb{V}\{H_{m,i}\} = \frac{1}{\kappa+1}\beta_{BR}$. Therefore, the expression of $\mathbf{Q}_k^n$ can be proved by combining (47)-(50).

$\mathbf{R}^n$ can be calculated as $\mathbf{R}^n = \text{Cov}(\mathbf{V}^{\text{n}}, \mathbf{V}^{\text{n}})$. Similarly, we calculate the covariance matrix for the kth user as $\mathbf{R}_k^n = \text{Cov}(\mathbf{V}_{\text{k}}^{\text{n}}, \mathbf{V}_{\text{k}}^{\text{n}})$. Apparently, $\mathbf{R}_k^n \in \mathbb{C}^{J\times J}$ is a diagonal matrix. Considering the same structure of $\mathbf{v}_{k,j}^n, \forall j$, we start with the element at the jth row and the jth column of $\mathbf{R}_k^n$ as follows:

$$[\mathbf{R}_k^n]_{j,j} = \mathbb{E}\left\{\left(\mathbf{v}_{k,j}^n\right)^T\left[\left(\mathbf{v}_{k,j}^n\right)^T\right]^H\right\} - \mathbb{E}\left\{\left(\mathbf{v}_{k,j}^n\right)^T\right\}\left(E\left\{\left(\mathbf{v}_{k,j}^n\right)^T\right\}\right)^H. \quad (51)$$

The first term in (51) can be calculated as

$$\mathbb{E}\left\{\left(\mathbf{v}_{k,j}^n\right)^T\left[\left(\mathbf{v}_{k,j}^n\right)^T\right]^H\right\} = \mathbb{E}\left\{\sum_{i\in C_{k,j}}\left(\boldsymbol{\phi}_j^n\right)^T\left(\mathbf{G}_i^n\right)^T\left(\mathbf{G}_i^n\right)^*\left(\boldsymbol{\phi}_j^n\right)^*\right\} + \mathbb{E}\left\{\left(\mathbf{s}_{k,j}^n\right)^*\left(\mathbf{w}_1^n\right)^T\left(\mathbf{w}_1^n\right)^*\left(\mathbf{s}_{k,j}^n\right)^T/p\right\}, \quad (52)$$

where $\mathbb{E}\left\{\left(\mathbf{s}_{k,j}^n\right)^*\left(\mathbf{w}_1^n\right)^T\left(\mathbf{w}_1^n\right)^*\left(\mathbf{s}_{k,j}^n\right)^T/p\right\} = M\sigma^2/p$ and $\mathbb{E}\left\{\sum_{i\in C_{k,j}}\left(\boldsymbol{\phi}_j^n\right)^T\left(\mathbf{G}_i^n\right)^T\left(\mathbf{G}_i^n\right)^*\left(\boldsymbol{\phi}_j^n\right)^*\right\} = \sum_{i\in C_{k,j}}\left(\boldsymbol{\phi}_j^n\right)^T\mathbb{E}\left\{\left(\mathbf{G}_i^n\right)^T\left(\mathbf{G}_i^n\right)^*\right\}\left(\boldsymbol{\phi}_j^n\right)^*$. The second term in (51) can be calculated as

$$\mathbb{E}\left\{\left(\mathbf{v}_{k,j}^n\right)^T\right\} = \mathbb{E}\left\{\sum_{i\in C_{k,j}}\left(\boldsymbol{\phi}_j^n\right)^T\left(\mathbf{G}_i^n\right)^T\right\} + \mathbb{E}\left\{\left(\mathbf{s}_{k,j}^n\right)^*\left(\mathbf{w}_j^n\right)^T/\sqrt{p}\right\} = \sum_{i\in C_{k,j}}\left(\boldsymbol{\phi}_j^n\right)^T E\left\{\left(\mathbf{G}_i^n\right)^T\right\} + \left(\mathbf{s}_{k,j}^n\right)^* E\left\{\left(\mathbf{w}_j^n\right)^T\right\}/\sqrt{p}. \quad (53)$$

As $\mathbf{w}^n$ has a zero mean, we only need to calculate $\mathbb{E}\left\{\left(\mathbf{G}_i^n\right)^T\right\}$ as follows:

$$\mathbb{E}\left\{\left(\mathbf{G}_i^n\right)^T\right\} = \mathbb{E}\left\{\text{diag}\left(\mathbf{h}_{2,i}^n\right)\right\}\mathbb{E}\left\{\left(\mathbf{H}_1\right)^T\right\} = \frac{\kappa}{\kappa+1}\sqrt{\beta_{Ru}(i)}\sqrt{\beta_{BR}}\,\text{diag}(\bar{\mathbf{h}}_{2,i})(\bar{\mathbf{H}}_1)^T. \quad (54)$$

We define $\mathbf{G}_i^{\text{cov}} \triangleq \mathbb{E}\left\{\left(\mathbf{G}_i^n\right)^T\left(\mathbf{G}_i^n\right)^*\right\} = \mathbb{E}\left\{\text{diag}\left(\mathbf{h}_{2,i}^n\right)\mathbf{H}_1^T\left[\text{diag}\left(\mathbf{h}_{2,i}^n\right)\mathbf{H}_1^T\right]^H\right\}$, and it can be calculated following the similar way as the first term in (47). Therefore, $\mathbf{R}_k^n$ is a diagonal matrix with its diagonal elements being given by combining (51)-(54).

REFERENCES

[1] P. Zhang, J. Chen, X. Yang, N. Ma, and Z. Zhang, "Recent research on massive MIMO propagation channels: A survey," *IEEE Commun. Mag.*, vol. 56, no. 12, pp. 22–29, Dec. 2018.

[2] T. E. Bogale and L. B. Le, "Massive MIMO and mmWave for 5G wireless HetNet: Potential benefits and challenges," *IEEE Veh. Technol. Mag.*, vol. 11, no. 1, pp. 64–75, Mar. 2016.

[3] K. Samdanis and T. Taleb, "The road beyond 5G: A vision and insight of the key technologies," *IEEE Netw.*, vol. 34, no. 2, pp. 135–141, Mar. 2020.

[4] V. Petrov *et al.*, "Dynamic multi-connectivity performance in ultra-dense urban mmWave deployments," *IEEE J. Sel. Areas Commun.*, vol. 35, no. 9, pp. 2038–2055, Sep. 2017.

[5] L. Sanguinetti, E. Björnson, and J. Hoydis, "Toward massive MIMO 2.0: Understanding spatial correlation, interference suppression, and pilot contamination," *IEEE Trans. Commun.*, vol. 68, no. 1, pp. 232–257, Jan. 2020.

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

[7] A. Shojaeifard *et al.*, "MIMO evolution beyond 5G through reconfigurable intelligent surfaces and fluid antenna systems," *Proc. IEEE*, vol. 110, no. 9, pp. 1244–1265, Sep. 2022.

[8] 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 Commun. Surveys Tuts.*, vol. 24, no. 2, pp. 1035–1071, 2nd Quart. 2022.

[9] M. Jian and Y. Zhao, "A modified off-grid SBL channel estimation and transmission strategy for RIS-assisted wireless communication systems," *Proc. Int. Wireless Commun. Mobile Comput. (IWCMC)*, Jun. 2020, pp. 1848–1853.

[10] D. Mishra and H. Johansson, "Channel estimation and low-complexity beamforming design for passive intelligent surface assisted MISO wireless energy transfer," *Proc. IEEE Int. Conf. Acoust., Speech Signal Process. (ICASSP)*, May 2019, pp. 4659–4663.

[11] Q. -U. -A. Nadeem, H. Alwazani, A. Kammoun, A. Chaaban, M. Debbah, and M. -S. Alouini, "Intelligent reflecting surface-assisted multi-user MISO communication: Channel estimation and beamforming design," *IEEE Open J. Commun. Soc.*, vol. 1, pp. 661–680, 2020.

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

[13] T. V. Chien, H. Q. Ngo, S. Chatzinotas, M. D. Renzo, and B. Ottersten, "Reconfigurable intelligent surface-assisted cell-free massive MIMO systems over spatially-correlated channels," *IEEE Trans. Wireless Commun.*, vol. 21, no. 7, pp. 5106–5128, Jul. 2022.

[14] Z. Wang, L. Liu, and S. Cui, "Channel estimation for intelligent reflecting surface assisted multiuser communications: Framework, algorithms, and analysis," *IEEE Trans. Wireless Commun.*, vol. 19, no. 10, pp. 6607–6620, Oct. 2020.

[15] E. P. Simon, L. Ros, H. Hijazi, J. Fang, D. P. Gaillot, and M. Berbineau, "Joint carrier frequency offset and fast time-varying channel estimation for MIMO-OFDM systems," *IEEE Trans. Veh. Technol.*, vol. 60, no. 3, pp. 955–965, Mar. 2011.

[16] P. Cai, J. Zong, X. Luo, Y. Zhou, S. Chen, and H. Qian, "Downlink channel tracking for intelligent reflecting surface-aided FDD MIMO systems," *IEEE Trans. Veh. Technol.*, vol. 70, no. 4, pp. 3341–3353, Apr. 2021.

[17] Z. Mao, M. Peng, and X. Liu, "Channel estimation for reconfigurable intelligent surface assisted wireless communication systems in mobility scenarios," *China Commun.*, vo. 18, no. 3, pp. 29–38, Mar. 2021.

[18] T. Li, H. Tong, Y. Xu, X. Su, and G. Qiao, "Double IRSs aided massive MIMO channel estimation and spectrum efficiency maximization for high-speed railway communications," *IEEE Trans. Veh. Technol.*, vol. 71, no. 8, pp. 8630–8645, Aug. 2022.

[19] Y. Wei, M. -M. Zhao, A. Liu, and M. -J. Zhao, "Channel tracking and prediction for IRS-aided wireless communications," *IEEE Trans. Wireless Commun.*, vol. 22, no. 1, pp. 563–579, Jan. 2023.

[20] Y. Liu, E. Liu, R. Wang, B. Lu, and Z. Han, "Beamforming design and performance evaluation for reconfigurable intelligent surface assisted wireless communication systems with non-ideal hardware," 2020, *arXiv:2006.00664.*

[21] Z. Xing, R. Wang, J. Wu, and E. Liu, "Achievable rate analysis and phase shift optimization on intelligent reflecting surface with hardware impairments," *IEEE Trans. Wireless Commun.*, vol. 20, no. 9, pp. 5514–5530, Sep. 2021.

[22] A. Papazafeiropoulos, C. Pan, P. Kourtessis, S. Chatzinotas, and J. M. Senior, "Intelligent reflecting surface-assisted MU-MISO systems with imperfect hardware: Channel estimation and beamforming deisgn," *IEEE Trans. Wireless Commun.*, vol. 21, no. 3, pp. 2077–2092, Mar. 2022.

[23] N. Ginige, K. B. S. Manosha, N. Rajatheva, and M. Latva-aho, "Untrained DNN for channel estimation of RIS-assisted multi-user OFDM system with hardware impairments," in *IEEE 32nd Annu. Int. Symp. Pers. Indoor Mobile Radio Commun. (PIMRC)*, Sep. 2021, pp. 561–566.

[24] Y. Liu, E. Liu, R. Wang, and Y. Geng, "Channel estimation and power scaling law of large reflecting surface with non-ideal hardware," 2020, *arXiv:2004.09761.*

[25] S. Ahmed, S. Lambotharan, A. Jakobsson, and J. A. Chambers, "Parameter estimation and equalization techniques for communication channels with multipath and multiple frequency offsets," *IEEE Trans. Commun.*, vol. 53, no. 2, pp. 219–223, Feb. 2005.

[26] S. Ahmed, S. Lambotharan, A. Jakobsson, and J. A. Chambers, "MIMO frequency-selective channels with multiple-frequency offsets: Estimation and detection techniques," *IEE Proc. on Commun.*, vol. 152, no. 4, pp. 489–494. Aug. 2005.

[27] P. Zhang, J. He, J. Dong, and L. Sha, "Joint frequency offset and channel estimation for NB-IoT systems," *J. Netw. Netw. Appl.*, vol. 2, no. 2, pp. 61–67, May 2022.

[28] V. Arya and K. Appaiah, "Kalman filter based tracking for channel aging in massive MIMO systems," in *Proc. Int. Conf. Signal Process. Commun. (SPCOM)*, Jul. 2018, pp. 362–366.

[29] Q. Du, Z. Dong, H. Xu, N. Wei, and J. Liu, "Uplink channel estimation for intelligent reflecting surface aided direct and reflected users," in *Proc IEEE VTC-Spring*, Jun. 2022, pp. 1–6.

[30] S. Jung, J. -W. Lee, and C. Lee, "RSS-based channel estimation for IRS-aided wireless energy transfer system," *IEEE Internet Things J.*, vol. 8, no. 19, pp. 14860–14873, Oct. 2021.

[31] X. Zhou, Z. Yang, T. Zhang, and Y. Jiang, "Channel estimation and projection for RIS-assisted MIMO using Zadoff-Chu sequences," 2022, *arXiv:2202.10038.*

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

[33] H. Cheng, Y. Xia, Y. Huang, L. Yang, and D. P. Mandic, "Joint channel estimation and Tx/Rx I/Q imbalance compensation for GFDM systems," *IEEE Trans. Wireless Commun.*, vol. 18, no. 2, pp. 1304–1317, Feb. 2019.

[34] Y. Byun, H. Kim, S. Kim, and B. Shim, "Channel estimation and phase shift control for UAV-carried RIS communication systems," *IEEE Trans. Veh. Technol.*, early access, May 10, 2023, doi: 10.1109/TVT.2023.3274871.

[35] C. Ruan, Z. Zhang, H. Jiang, J. Dang, L. Wu, and H. Zhang, "Approximate message passing for channel estimation in reconfigurable intelligent surface aided MIMO multiuser systems," *IEEE Trans. Commun.*, vol. 70, no. 8, pp. 5469–5481, Aug. 2022.

[36] N. K. Kundu and M. R. McKay, "Channel estimation for reconfigurable intelligent surface aided MISO communications: From LMMSE to deep learning solutions," *IEEE Open J. Commun. Soc.*, vol. 2, pp. 471–487, 2021.

[37] Q. Zhang, T. Q. S. Quek, and S. Jin, "Scaling analysis for massive MIMO systems with hardware impairments in Rician fading," *IEEE Trans. Wireless Commun.*, vol. 17, no. 7, pp. 4536–4549, Jul. 2018.

[38] C. Li, A. Sezgin, and Z. Han, "On the impact of oscillator phase noise in an IRS-assisted MISO TDD systems," in *25th International ITG Workshop on Smart Antennas (WSA)*, Nov. 2021, pp. 220–225.

[39] N. Hajiabdolrahim, S. R. Aghdam, and T. Eriksson, "An extended kalman filter framework for joint phase noise, CFO and sampling time error estimation," in *IEEE 31st Annu. Int. Symp. Pers. Indoor Mobile Radio Commun. (PIMRC)*, Aug. 2020, pp. 1–6.

[40] D. -N. Nguyen and K. T. Truong, "Phase impairment estimation for mmWave MIMO systems," in *8th NAFOSTED Conference on Information and Computer Science (NICS)*, Dec. 2021, pp. 543–548.

[41] J. Dai, F. Zhu, C. Pan, H. Ren, and K. Wang, "Statistical CSI-based transmission design for reconfigurable intelligent surface-aided massive MIMO systems with hardware impairments," *IEEE Wireless Commun. Lett.*, vol. 11, no. 1, pp. 38–42, Jan. 2022.