Dimension Reduced Channel Feedback for Reconfigurable Intelligent Surface Aided Wireless Communications

Decai Shen and Linglong Dai

***Abstract*— Reconfigurable intelligent surface (RIS) has recently received extensive research interest due to its capability to intelligently change the wireless propagation environment. For RIS-aided wireless communications in frequency division duplex (FDD) mode, channel feedback at the user equipment (UE) is essential for the base station (BS) to acquire the downlink channel state information (CSI). In this paper, a dimension reduced channel feedback scheme is proposed to reduce the channel feedback overhead by exploiting the single-structured sparsity of BS-RIS-UE cascaded channel. Since different UEs share the same sparse BS-RIS channel but have their respective RIS-UE channels, there are only limited non-zero column vectors in the BS-RIS-UE cascaded channel matrix, and different UEs share the same indexes of the non-zero columns. Thus, the downlink CSI can be decomposed into *user-independent* channel information (i.e., the indexes of non-zero columns) and *user-specific* channel information (i.e., the non-zero column vectors), where the former for all UEs can be fed back by only one UE, while the latter can be fed back with a fairly low overhead by different UEs, respectively. Simulation results show that, compared with the conventional method, the proposed scheme can reduce channel feedback overhead by more than 80% for RIS-aided wireless communications.**

***Index Terms*— Reconfigurable intelligent surface, FDD, channel feedback, codebook, single-structured sparsity.**

I. Introduction

RECONFIGURABLE intelligent surface (RIS) based on metamaterial is emerging as a new promising technology for future 6G communications [2]. Instead of passively adapting to the propagation environment between the base station (BS) and the user equipments (UEs) in existing wireless communication systems, RIS can control the propagation environment by leveraging its controllable metamaterial-based elements. Besides, the noise introduced by an RIS can be nearly neglected, since normally no signal processing or radio-frequency (RF) chain is deployed at the RIS [3]. Hence, the RIS-aided wireless communication system is able to improve the spectrum efficiency and energy efficiency [4]. To achieve this performance gain, downlink channel state information (CSI) is essential for the joint beamforming optimization at the BS and RIS in RIS-aided wireless communication systems [5]–[8]. To be more specific, the RIS-aided wireless channel consists of two parts: 1) The BS-RIS-UE cascaded channel, which is the concatenation of the BS-RIS channel and the RIS-UE channel; 2) The direct BS-UE channel between the BS and the UE. Unfortunately, the channel acquisition overhead of downlink channel estimation is usually in proportion to the numbers of BS antennas and RIS elements, which are likely to be very large in RIS-aided wireless communication systems, e.g., 64 BS antennas [9] and 256 RIS elements [3]. This overhead may lead to a large gap between theoretical and actual performance in the RIS-aided wireless communication system. Hence, designing an accurate downlink CSI acquisition scheme with a low overhead is crucial for RIS-aided wireless communications.

In time division duplex (TDD) system, thanks to the channel reciprocity between the downlink and uplink channels, the downlink CSI can be easily acquired by the uplink channel estimation. In this way, the pilot overhead can be significantly reduced, which is only in proportion to the number of UEs. Thus, previous works mainly focus on the CSI acquisition schemes realized by the uplink channel estimation in TDD RIS-aided wireless communications systems [10]–[13]. However, frequency division duplex (FDD) is still widely used in current wireless communication systems, where the downlink CSI cannot be easily acquired by the uplink channel estimation, since the channel reciprocity does not hold. Considering the cost for system evolution in practice, we believe that FDD will also be an important working mode for future RIS-aided wireless communications. Hence, it is worth to design an efficient channel feedback scheme to feed the downlink CSI from UEs back to the BS, which is essential to optimize the joint beamforming in the FDD mode. In existing communication systems like wireless fidelity (Wi-Fi) and 5G [14], [15], codebook-based channel feedback schemes with reliable performance have been widely adopted. Unfortunately, the size of the conventional codebook for channel feedback exponentially increases with the number of RIS elements and BS antennas, which is unbearable in practical RIS-aided

Manuscript received December 3, 2020; revised May 2, 2021 and July 15, 2021; accepted July 18, 2021. Date of publication July 26, 2021; date of current version November 18, 2021. This work was supported in part by the National Key Research and Development Program of China (Grant No. 2020YFB1805005), in part by the National Natural Science Foundation of China (Grant No. 62031019), and in part by the European Commission through the H2020-MSCA-ITN META WIRELESS Research Project under Grant 956256. This article was presented in part at IEEE GLOBECOM 2020 [1]. The associate editor coordinating the review of this article and approving it for publication was Y.-W. P. Hong. *(Corresponding author: Linglong Dai.)*

The authors are with Beijing National Research Center for Information Science and Technology (BNRist), Department of Electronic Engineering, Tsinghua University, Beijing 100084, China (e-mail: sdc18@mails.tsinghua.edu.cn; daill@tsinghua.edu.cn).

Color versions of one or more figures in this article are available at https://doi.org/10.1109/TCOMM.2021.3100428.

Digital Object Identifier 10.1109/TCOMM.2021.3100428

wireless communication systems with a large number of BS antennas and RIS elements.

A. Prior Works

Although channel feedback has not been investigated for RIS-aided wireless communication, it has been widely studied in current wireless communication systems [14]–[20]. The downlink CSI should be acquired by channel estimation at the UE at first. Then, considering the overhead and implementation, the downlink CSI is usually approximated by an elaborate codebook before feeding back to the BS. The codebook consists of many codewords, which have the same dimension as the channel vector. To approximatively represent the downlink CSI, the codeword whose vector direction is the closest to the channel vector will be selected. To reduce the overhead for channel feedback, the index of this codeword rather than the codeword itself will be fed back to the BS. Once the index has been obtained, the BS can acquire the approximated downlink CSI according to the codeword index and the known codebook.

In [14], the random vector quantization (RVQ) codebook was proposed for multiple-input multiple-output (MIMO) channel feedback. The RVQ codebook consists of codewords that are randomly and independently generated from the uniform distribution on the complex unit sphere. It is worth to note that the required codebook size (i.e., the number of the codewords in the codebook for channel feedback) exponentially increases with the number of BS antennas. Hence, the size of RVQ codebook is usually large in 5G wireless communications due to the deployment of massive MIMO. To reduce the codebook size and the overhead for channel feedback, a more efficient codebook with better performance was proposed based on the channel statistics and the measurement matrix [15]. Specifically, a codeword of the proposed codebook was designed by multiplying a vector from RVQ codebook by the UE's channel correlation matrix. Based on the fact that such kind of channel statistics like channel correlation remains unchanged in a large time scale, the statistics-based codebook is able to achieve better performance than the RVQ codebook with the same codebook size.

However, the size of the statistics-based codebook in [15] is still large, especially when there are a huge number of antennas. To further reduce the channel feedback overhead, the channel feedback scheme based on compressive sensing (CS) was proposed in [16]. By utilizing the channel sparsity in the angular domain, which is mainly resulted from the fact that there are usually limited scatterers between the BS and the UEs, the channel vector can be compressed as a low-dimensional measurement vector by random projection at first. Then, this low-dimensional measurement vector can be fed back to the BS with a low overhead. Thanks to the ability of sparse signal recovery of CS, the sparse channel in the angular domain can be reliably recovered by using the received low-dimensional measurement vector at the BS. More recently, inspired by the excellent performance of deep learning (DL) for image compression, DL-based channel feedback schemes have been proposed [17]–[19]. Specifically, an efficient DL-based compression framework was proposed in [17], where the CSI compression, CSI encoding, and CSI recovery can be jointly tackled. To achieve better performance via extracting CSI features with multiple resolutions, a novel feedback network was proposed in [18], which was able to extract multi-resolution features and adapt to various scenarios and compression ratios. To make a trade-off between complexity and accuracy for channel feedback, a DL-based limited feedback method was studied in [19]. For this method, two individual deep neural networks (DNNs) were respectively implemented at the BS and UE to jointly design the CSI prediction and codebook optimization.

Considering that the direct BS-UE channel has been studied extensively in current wireless communications, we focus on the BS-RIS-UE cascaded channel feedback in this paper. However, the existing channel feedback schemes mentioned above cannot be directly applied to the RIS-aided wireless communications. This is mainly caused by the fact that the size of the BS-RIS-UE cascaded channel is much larger (e.g., 64×256 [3] with 64 BS antennas and 256 RIS elements) than the channel size of current massive MIMO systems (e.g., 64×1 [20] with 64 BS antennas and one UE antenna), which will result in the pretty high channel feedback overhead and large codebook size. This challenging channel feedback problem has not been investigated for the RIS-aided wireless communication in the literature.

B. Our Contributions

In this paper, channel feedback for the RIS-aided wireless communications is investigated for the first time to the best of our knowledge.[1] A dimension reduced channel feedback scheme is proposed to reduce the channel feedback overhead. The single-structured sparsity of all UEs' channels is jointly exploited to reduce per-user channel feedback overhead in the proposed scheme. This kind of channel sparsity characteristic mainly results from that different UEs share the same sparse BS-RIS channel but have their respective RIS-UE channels, i.e., the sparse BS-RIS channel is *user-independent* while the RIS-UE channel is *user-specific*. This characteristic has been utilized to improve the performance for channel estimation [21]–[23]. It is worth pointing out that channel feedback and channel estimation are the two most important core tasks for channel acquisition in RIS-aided wireless communications systems. In this work, different from the above works which focus on channel estimation, we reveal a strong interest in another core task, i.e., channel feedback, which is rarely involved but surely important for RIS-aided wireless communications.

Specifically, to reduce the channel feedback overhead, we exploit the single-structured sparsity in the hybrid spatial and angular domain, which is called as the *hybrid domain* in this paper. The BS-RIS-UE cascaded channel in the hybrid domain means that the channel is converted into the angular domain in the column dimension, while it is still in the spatial domain in the row dimension. For the hybrid-domain channel matrix, there are only limited non-zero columns. Moreover, different UEs share the same set of indexes for the limited non-zero channel vectors in the column dimension, which is called as structured sparsity, while the channel vectors

[1]Simulation codes are provided to reproduce the results presented in this paper: http://oa.ee.tsinghua.edu.cn/dailinglong/publications/publications.html.

are different for BS-RIS-UE cascaded channels of different UEs. In other words, the BS-RIS-UE cascaded channels of different UEs show the structured sparsity only in the column dimension but not in the row dimension, which is called as the *single-structured*[2] *sparsity* in this paper.

By exploiting the single-structured sparsity, we propose a dimension reduced channel feedback scheme, based on which channel feedback is decomposed as the user-independent CSI feedback and user-specific CSI feedback. Specifically, for the hybrid-domain channel matrix, what we need to feedback are the non-zero column vectors and their corresponding indexes in the column dimension. The indexes of the non-zero columns can be regarded as user-independent channel information, while the non-zero column vectors can be treated as user-specific channel information. In the proposed scheme, the user-independent channel information for all UEs can be fed back to the BS by only one UE. Then, the user-specific channel information can be fed back to the BS with the help of codebook-based feedback scheme by different UEs. Moreover, instead of utilizing a constant codebook, we design dynamic codebooks based on the channel angle information to match the time-varying BS-RIS-UE cascaded channel more accurately. Simulation results show that, compared with the conventional channel statistics-based method, the proposed scheme can reduce the channel feedback overhead by more than 80% in RIS-aided wireless communication systems.

C. Organization and Notation

Organization: The rest of this paper is organized as follows. Section II presents the channel model of the RIS-aided wireless communication, together with the basic principle for channel feedback. Section III explains the single-structured sparsity of the BS-RIS-UE cascaded channel at first, and then the dimension reduced channel feedback scheme with low overhead is proposed. Theoretical analysis for the proposed scheme is also provided in this section. Section IV provides the simulation results. Finally, we conclude this paper in Section V.

Notation: We use the lower-case and upper-case boldface letters $\mathbf{a}$ and $\mathbf{A}$ to denote a vector and a matrix, respectively; $\mathbf{A}^T$, $\mathbf{A}^H$, and $\mathbf{A}^{-1}$ denote the transpose, conjugate transpose, and inversion transpose of the matrix $\mathbf{A}$, respectively; $\|\mathbf{a}\|$ is the norm of the vector $\mathbf{a}$; $\sin^2(\measuredangle(\mathbf{a},\mathbf{b})) = 1 - \frac{|\mathbf{a}^{\mathrm{H}}\mathbf{b}|^2}{\|\mathbf{a}\|^2\|\mathbf{b}\|^2}$; $\mathbf{A}\otimes\mathbf{B}$ denotes the kronecker product of $\mathbf{A}$ and $\mathbf{B}$; $\mathcal{CN}(\mu,\sigma^2)$ represents the complex Gaussian distribution with the mean μ and the standard deviation σ; $\mathbf{A}_{(:,n)}$ denotes the n-th column of the matrix $\mathbf{A}$; $\mathbb{E}\{\cdot\}$ denotes the expectation operator; $<\mathbf{a},\mathbf{b}>$ denotes the inner product operation; $\lceil x \rceil$ denotes the smallest integer that is greater than or equal to x; Finally, $\mathrm{diag}(\mathbf{a})$ denotes the diagonal matrix with the vector $\mathbf{a}$ on its diagonal.

II. System Model

In this section, we firstly introduce the system model of RIS-aided wireless communication, based on which the channel feedback scheme and codebook design are then discussed.

[2]In this paper, "structured" means that the sparsity, non-zero column indexes, or other characteristics keep the same among different UEs' channels.

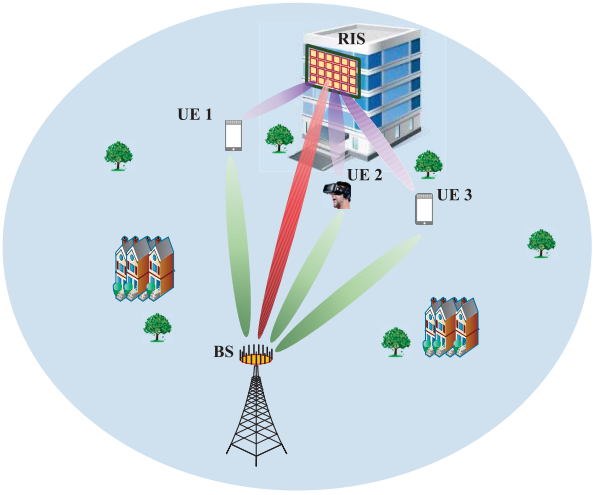

Fig. 1. The RIS-aided wireless communication system.

A. RIS-Aided Wireless Communication Model

In this paper, we consider an RIS-aided wireless communication system as illustrated in Fig. 1, where a BS with M antennas is aided by an RIS with N elements to simultaneously serve K single-antenna UEs. The downlink signal model for the k-th UE can be expressed as [24]

$$y_k = \left(\mathbf{h}_{d,k}^T + \mathbf{h}_{r,k}^T \mathbf{P}\mathbf{G}\right)\mathbf{x} + n_k, \tag{1}$$

where y_k is the received signal of the k-th UE; $\mathbf{x} \in \mathbb{C}^{M\times 1}$ is the precoded transmitted signal at the BS; $\mathbf{h}_{d,k}^T \in \mathbb{C}^{1\times M}$, $\mathbf{h}_{r,k}^T \in \mathbb{C}^{1\times N}$, and $\mathbf{G} \in \mathbb{C}^{N\times M}$ denote the direct BS-UE channel from the BS to the k-th UE, the RIS-UE channel from the RIS to the k-th UE, and the BS-RIS channel matrix from the BS to the RIS, respectively; $n_k \sim \mathcal{CN}(0,\sigma_n^2)$ is the additive white Gaussian noise (AWGN) at the k-th UE; $\mathbf{P} \in \mathbb{C}^{N\times N}$ represents the reflecting diagonal matrix of the RIS as

$$\mathbf{P} = \mathrm{diag}\left(\boldsymbol{p}^T\right) = \mathrm{diag}\left(p_1,\cdots,p_n,\cdots,p_N\right), \tag{2}$$

where $p_n = e^{j\varphi_n}$ $(\varphi_n \in [0,2\pi], n = 1,2,\cdots,N)$ represents the reflecting coefficient of the n-th RIS element. Note that the reflecting coefficient can be also represented as $p_n = \xi e^{j\varphi_n}$, where $\xi \in [0,1]$ is the amplitude coefficient. Considering the reflecting gain and hardware complexity, the default value of the amplitude coefficient is set as $\xi = 1$, which is widely considered in existing research for RIS [3].

By utilizing the property of diagonal matrix, i.e., $\mathbf{h}_{r,k}^T\mathbf{P}\mathbf{G} = \boldsymbol{p}^T\mathrm{diag}\left(\mathbf{h}_{r,k}^T\right)\mathbf{G}$, the equivalent baseband downlink channel $\mathbf{h}_{\mathrm{DL},k}^T \in \mathbb{C}^{1\times M}$ for the k-th UE can be expressed as

$$\mathbf{h}_{\mathrm{DL},k}^T = \mathbf{h}_{d,k}^T + \mathbf{h}_{r,k}^T\mathbf{P}\mathbf{G} = \mathbf{h}_{d,k}^T + \boldsymbol{p}^T\mathrm{diag}\left(\mathbf{h}_{r,k}^T\right)\mathbf{G}. \tag{3}$$

Then, we denote the BS-RIS-UE cascaded channel of the k-th UE $\mathbf{H}_k$ in the spatial domain as

$$\mathbf{H}_k \triangleq \mathrm{diag}\left(\mathbf{h}_{r,k}^T\right)\mathbf{G}. \tag{4}$$

In this paper, we assume uniform linear array (ULA) of antennas at the BS [25] and uniform planar array (UPA) of elements at the RIS [4], respectively. Considering the widely

used ray-based channel model in the literature [26], the *user-independent* BS-RIS channel[3] in the spatial domain can be expressed as

$$\mathbf{G} = \sum_{i=1}^{L_1} \alpha_i \mathbf{b}_R\left(\psi_{1,i}, \gamma_{1,i}\right) \mathbf{a}_T^H\left(\psi_i^{\mathrm{AoD}}\right), \tag{5}$$

where L_1 is the number of dominant paths, α_i denotes the complex gain of the i-th path, and $\psi_{1,i}$ ($\gamma_{1,i}$) denotes the *physical* azimuth (elevation) angle of arrival (AoA) of the i-th path, respectively. We consider an UPA with N_1 horizontal elements and N_2 vertical elements ($N = N_1 \times N_2$), and the steering vector $\mathbf{b}_R\left(\psi_{1,i}, \gamma_{1,i}\right) \in \mathbb{C}^{N\times 1}$ of the BS-RIS channel between the BS and the RIS of the i-th path can be expressed as

$$\begin{aligned}\mathbf{b}_R\left(\psi_{1,i}, \gamma_{1,i}\right) = &\left[e^{j2\pi n_1 \frac{d_{\mathrm{R}}}{\lambda}\cos\gamma_{1,i}\sin\psi_{1,i}}\right]^T_{n_1\in\mathcal{I}(N_1)} \\ &\otimes \frac{1}{\sqrt{N_1}}\frac{1}{\sqrt{N_2}}\left[e^{j2\pi n_2 \frac{d_{\mathrm{R}}}{\lambda}\sin\gamma_{1,i}}\right]^T_{n_2\in\mathcal{I}(N_2)},\end{aligned} \tag{6}$$

where d_{R} is the element spacing at the RIS, λ is the wavelength of the carrier frequency. In this paper, d_{R} is set as half wavelength, which is widely adopted in wireless communications systems [10]. The integer sequence $\mathcal{I}(n)$ can be expressed as $\mathcal{I}(n) = \{0, 1, \cdots, n-1\}$. Finally, the steering vector $\mathbf{a}_T\left(\psi_i^{\mathrm{AoD}}\right) \in \mathbb{C}^{M\times 1}$ in (5) denotes the antenna array response of the i-th path, which can be expressed as

$$\mathbf{a}_T\left(\psi_i^{\mathrm{AoD}}\right) = \frac{1}{\sqrt{M}}\left[e^{j2\pi m \frac{d_{\mathrm{B}}}{\lambda}\sin\psi_i^{\mathrm{AoD}}}\right]^T_{m\in\mathcal{I}(M)}, \tag{7}$$

where d_{B} denotes the antenna spacing at the BS, and ψ_i^{AoD} denotes the angle of departure (AoD) of the i-th path between the BS and the RIS. Note that d_{B} is also set as half wavelength, i.e., we have $d_{\mathrm{B}} = d_{\mathrm{R}} = \lambda/2$.

To express the channel more concisely, for the UPA at the RIS, we rewrite the steering vector (6) with *spacial* angles as

$$\begin{aligned}\mathbf{b}\left(\phi_{1,i}, \theta_{1,i}\right) = &\left[e^{j2\pi n_1 \phi_{1,i}}\right]^T_{n_1\in\mathcal{I}(N_1)} \\ &\otimes \frac{1}{\sqrt{N}}\left[e^{j2\pi n_2 \theta_{1,i}}\right]^T_{n_2\in\mathcal{I}(N_2)},\end{aligned} \tag{8}$$

where $\phi_{1,i} = d_{\mathrm{R}}\cos\gamma_{1,i}\sin\psi_{1,i}/\lambda$, and $\theta_{1,i} = d_{\mathrm{R}}\sin\gamma_{1,i}/\lambda$ are the normalized spacial azimuth and elevation AoAs for RIS with the range $\left[-\frac{1}{2}, \frac{1}{2}\right]$, respectively. For the ULA at the BS, the steering vector (7) with spacial angles can be similarly expressed as

$$\mathbf{a}\left(\phi_i^{\mathrm{AoD}}\right) = \frac{1}{\sqrt{M}}\left[e^{j2\pi m \phi_i^{\mathrm{AoD}}}\right]^T_{m\in\mathcal{I}(M)}, \tag{9}$$

where $\phi_i^{\mathrm{AoD}} = d_{\mathrm{B}}\sin\psi_i^{\mathrm{AoD}}/\lambda$ is the normalized spacial AoD for BS with the range $\left[-\frac{1}{2}, \frac{1}{2}\right]$. Based on (8) and (9), the BS-RIS channel $\mathbf{G}$ in (5) can be rewritten as

$$\mathbf{G} = \sum_{i=1}^{L_1} \alpha_i \mathbf{b}\left(\phi_{1,i}, \theta_{1,i}\right) \mathbf{a}^H\left(\phi_i^{\mathrm{AoD}}\right). \tag{10}$$

[3]Some researches are working on the multi-RIS assisted multi-user wireless communication systems, based on which the channels between arbitrary RIS to BS, or arbitrary two RISs are user-independent [27].

Similar to (5), the *user-specific* RIS-UE channel $\mathbf{h}_{r,k}$ between the RIS and the k-th UE in the spatial domain can be expressed as

$$\mathbf{h}_{r,k}^T = \sum_{i=1}^{L_2} \beta_{k,i} \mathbf{b}_T^H\left(\psi_{2,k,i}, \gamma_{2,k,i}\right), \tag{11}$$

where L_2 is the number of dominant paths, $\beta_{k,i}$ is the complex gain of the i-th path, $\psi_{2,k,i}$ and $\gamma_{2,k,i}$ are the physical azimuth and elevation AoDs of the i-th path, respectively, and the steering vector $\mathbf{b}_T\left(\psi_{2,k,i}, \gamma_{2,k,i}\right)$ can be expressed as

$$\begin{aligned}&\mathbf{b}_T\left(\psi_{2,k,i}, \gamma_{2,k,i}\right) \\ &= \left[e^{j2\pi n_1 \frac{d_{\mathrm{R}}}{\lambda}\cos\gamma_{2,k,i}\sin\psi_{2,k,i}}\right]^T_{n_1\in\mathcal{I}(N_1)} \\ &\quad\otimes \frac{1}{\sqrt{N_1}}\frac{1}{\sqrt{N_2}}\left[e^{j2\pi n_2 \frac{d_{\mathrm{R}}}{\lambda}\sin\gamma_{2,k,i}}\right]^T_{n_2\in\mathcal{I}(N_2)}.\end{aligned} \tag{12}$$

Similar to (8), by changing the physical angles to the spacial angles, the steering vector in (12) can be expressed as

$$\begin{aligned}\mathbf{b}\left(\phi_{2,k,i}, \theta_{2,k,i}\right) = &\left[e^{j2\pi n_1 \phi_{2,k,i}}\right]^T_{n_1\in\mathcal{I}(N_1)} \\ &\otimes \frac{1}{\sqrt{N}}\left[e^{j2\pi n_2 \theta_{2,k,i}}\right]^T_{n_2\in\mathcal{I}(N_2)},\end{aligned} \tag{13}$$

where the normalized spacial azimuth and elevation AoDs for RIS are defined as $\phi_{2,k,i} = d_{\mathrm{R}}\cos\gamma_{2,k,i}\sin\psi_{2,k,i}/\lambda$ and $\theta_{2,k,i} = d_{\mathrm{R}}\sin\gamma_{2,k,i}/\lambda$, respectively. Both of the two spacial angles have the range $\left[-\frac{1}{2}, \frac{1}{2}\right]$. Then, the user-specific RIS-UE channel $\mathbf{h}_{r,k}^T$ in (11) can be also written as

$$\mathbf{h}_{r,k}^T = \sum_{i=1}^{L_2} \beta_{k,i} \mathbf{b}^H\left(\phi_{2,k,i}, \theta_{2,k,i}\right). \tag{14}$$

According to (10) and (14), the BS-RIS-UE cascaded channel of the k-th UE $\mathbf{H}_k$ in the spatial domain as shown in (4) can be expressed as

$$\begin{aligned}\mathbf{H}_k = &\sum_{i=1}^{L_1}\sum_{j=1}^{L_2} \alpha_i \beta_{k,j} \\ &\mathrm{diag}\left(\mathbf{b}^H\left(\phi_{2,k,j}, \theta_{2,k,j}\right)\right) \mathbf{b}\left(\phi_{1,i}, \theta_{1,i}\right) \mathbf{a}^H\left(\phi_i^{\mathrm{AoD}}\right).\end{aligned} \tag{15}$$

For simplicity, we rewrite (15) as

$$\mathbf{H}_k = \sum_{i=1}^{L_1}\sum_{j=1}^{L_2} g_{i,k,j} \mathbf{b}\left(\phi_{i,k,j}^{\mathrm{AoA}}, \theta_{i,k,j}^{\mathrm{AoA}}\right) \mathbf{a}^H\left(\phi_i^{\mathrm{AoD}}\right), \tag{16}$$

where $g_{i,k,j} \triangleq \alpha_i \beta_{k,j}$ is the complex gain for the BS-RIS-UE cascaded channel, the subscripts i and j express the path index from the BS to the RIS and that from the RIS to the UE, respectively, the subscript k is the UE index, $\phi_{i,k,j}^{\mathrm{AoA}} = \phi_{1,i} - \phi_{2,k,j}$ and $\theta_{i,k,j}^{\mathrm{AoA}} = \theta_{1,i} - \theta_{2,k,j}$ are the cascaded normalized spacial azimuth and elevation angles at the RIS, which are also called as *cascaded AoAs* in this paper. Note that the number of paths for the channel between BS to RIS L_1, and that of the channel between RIS to the UE L_2 are usually much smaller than the number of BS antennas M, and the number of the RIS elements N, respectively. In other words, we have $L_1 \ll M$ and $L_2 \ll N$. This is mainly because that based on the ray-based channel model [26], L_1

only depends on the scattering geometry around the BS, and L_2 only depends on the scattering geometry around the RIS. Since the BS and the RIS are usually deployed at the high towers and the wall of skyscrapers, there are only a few scatters around the BS and RIS, based on which L_1 and L_2 are usually small. On the contrary, M and N are very large, since a huge number of antennas and elements are respectively arranged in the BS and RIS to better exploit the performance gain of the joint beamforming. Correspondingly, the AoDs at the BS and cascaded AoAs at the RIS depend on the scatterers around the BS and RIS, and the scatterers may not change their locations rapidly. Hence, these angles may remain unchanged over a large period of time, which is called as angle coherence time [20]. On the contrary, the path gains, which are related to the location changes of the UEs and the scatterers around the UEs, vary more quickly than the path angles, which usually determines the channel coherence time.

B. Channel Feedback for FDD RIS-Aided System

In FDD RIS-aided wireless communication systems, it is hard to utilize the channel reciprocity to obtain the downlink CSI by relying on the uplink channel estimation, since the frequency-dependent channel parameters are different between the downlink and uplink channels. Hence, the BS has to acquire the downlink CSI by channel feedback from the UE. The downlink channel should be estimated by the UE before channel feedback. There are several existing works of channel estimation for RIS-aided wireless communications [3], [28]–[31], hence in this paper we assume that the UE can acquire the downlink CSI perfectly, which is widely adopted in existing channel feedback researches [14], [15], [20]. Besides, it is worth to note that there is no essential difference between the direct BS-UE channel $\mathbf{h}_{d,k}^T$ of the RIS-aided wireless communications and that of the conventional massive MIMO systems, which has been extensively studied in existing channel feedback works. Thus, we focus on the channel feedback of the BS-RIS-UE cascaded channel $\mathbf{H}_k$ in this paper.

Specifically, we focus on widely adopted the quantized channel feedback scheme, based on which the channel is quantized by utilizing channel quantization codebooks at first, and then the quantized CSI (i.e., the index of selected codeword) is fed back to the BS by the UE [4] [14]. However, different from the conventional massive MIMO channel, which is usually regarded as an $M \times 1$-dimensional vector, the BS-RIS-UE cascaded channel $\mathbf{H}_k$ is an $N \times M$-dimensional matrix as shown in (4). Consequently, it is difficult to directly apply current channel feedback schemes with vector codebook to the RIS-aided wireless communication system. A straightforward solution is that, we can design a matrix codebook to feed the BS-RIS-UE cascaded channel matrix as a whole, based on which the channel feedback procedure needs to be carried out only once, which is called as the "single-time feedback" scheme in this paper. However, the complexity for the matrix codebook design (e.g., the hierarchical codebook [32]) will increase exponentially with the channel matrix size. Hence, the single-time feedback scheme will introduce unbearable complexity, especially when there are many BS antennas and RIS elements.

Another alternative solution, which is called as the "multi-time feedback" scheme in this paper, is to split the channel matrix into many channel vectors at first, and then the channel matrix feedback task can be decomposed into several channel vector feedback sub-tasks. For each sub-task of this multi-time feedback scheme, one of the channel vectors is quantized as a codeword selected from the existing vector codebook at first, and then the selected codeword index is fed back to the BS. The channel quantization and the channel feedback of the multi-time feedback scheme are discussed as follows.

Taking a column vector $\mathbf{h} \in \mathbb{C}^{N\times 1}$ from the BS-RIS-UE cascaded channel matrix $\mathbf{H}_k \in \mathbb{C}^{N\times M}$ as an example, we can divide the column vector $\mathbf{h}$ into the channel amplitude $\|\mathbf{h}\|$ and channel direction vector $\mathbf{e_h} = \frac{\mathbf{h}}{\|\mathbf{h}\|}$. Since the channel amplitude $\|\mathbf{h}\|$ is a scalar and very easy to be fed back with a low overhead, it is usually assumed to be perfectly known by the BS in existing channel feedback schemes [15]. Thus, we adopt the same assumption and only focus on the feedback of the channel direction vector $\mathbf{e_h}$. To ensure that the vector $\mathbf{e_h}$ composed with continuous complex values can be fed back to the BS, the widely used solution is to approximately represent $\mathbf{e_h}$ by selecting a codeword $\mathbf{c}_i$ from a pre-designed codebook

$$\mathcal{C}_0 = \{\mathbf{c}_1, \cdots, \mathbf{c}_i, \cdots, \mathbf{c}_{2^B}\}, \tag{17}$$

which contains 2^B different N-dimensional unit-norm column vectors (i.e., codewords), and B is the quantization bit of the codeword index. The principle for the codeword selection is that, the vector direction of the selected codeword should be as close as possible to the channel direction vector $\mathbf{e_h}$. It is obvious that the performance for channel feedback will be improved by increasing the codebook size 2^B, based on which we can divide the range of spatial directions with a higher resolution. However, the complexity for the codeword selection and the overhead for the codeword index feedback are directly related to the codebook size. Hence, the codebook design is essential for channel feedback performance. The codebook design and the codeword selection will be discussed more in detail in Section III.

After selecting the codeword, the index i of the selected codeword will be fed back to the BS from the UE. The codeword $\mathbf{c}_i$, which approximately represents the channel direction vector $\mathbf{e_h}$, can be obtained from the same codebook $\mathcal{C}_0$ as shown in (17) at the BS according to the index i. Finally, the channel column vector can be recovered by combining the codeword $\mathbf{c}_i$ and the corresponding channel amplitude. After carrying out the same procedure for all channel vectors, the BS-RIS-UE cascaded channel $\mathbf{H}_k$ can be acquired at the BS by combining those vectors as a matrix column by column.

However, the biggest obstacle for utilizing the multi-time feedback scheme for the emerging RIS-aided wireless communication is that, it will result in a unbearable channel feedback overhead. For instance, to feed the BS-RIS-UE cascaded channel matrix of size 64×256, the operation for codebook indexes feedback needs to be repeated for 256 times within the channel coherence time, which results in the high overhead for channel feedback and the low efficiency for data transmission.

[4] To finish the channel feedback, the downlink CSI will be transmitted to the BS via the uplink channel. For uplink transmission of RIS-aided wireless communications, the existing schemes for the joint beamforming optimization at the RIS and the BS can be utilized [5]–[8].

To solve this problem, we propose a more efficient scheme for BS-RIS-UE cascaded channel feedback with a low overhead, which will be discussed in the next section.

III. Proposed Dimension Reduced Channel Feedback Scheme

In this section, we introduce the single-structured sparsity of the BS-RIS-UE cascaded channel at first, based on which we then propose a dimension reduced channel feedback scheme to reduce the channel feedback overhead.

A. Single-Structured Sparsity of the BS-RIS-UE Cascaded Channel

1) Sparsity of a Single UE's BS-RIS-UE Cascaded Channel: In an RIS-aided wireless communication system, the BS and RIS are usually surrounded by limited scatterers. In other words, there are only a few AoDs at the BS and a few cascaded AoAs at the RIS. To illustrate this property, we propose the concept for hybrid domain and reveal the relationship between the hybrid-domain channel $\tilde{\mathbf{H}}_k$ and the well-known spatial-domain channel $\mathbf{H}_k$ as

$$\mathbf{H}_k = \tilde{\mathbf{H}}_k \mathbf{\Theta}_T^H, \tag{18}$$

where $\mathbf{\Theta}_T \in \mathbb{C}^{M \times G_t}$ is the dictionary matrix with the angular resolution G_t of the AoD at the BS. By utilizing the dictionary matrix, we can divide the BS AoDs into G_t grids. Note that when the typical resolution is set as $G_t = M$, the dictionary matrix $\mathbf{\Theta}_T$ becomes the $M \times M$ discrete Fourier transformation (DFT) matrix [33]. In this paper, $\tilde{\mathbf{H}}_k$ in (18) is called as the "hybrid-domain channel" (hybrid spatial- and angular-domain channel), since $\tilde{\mathbf{H}}_k$ in the column dimension is converted into the angular domain by spatial DFT, while $\tilde{\mathbf{H}}_k$ in the row dimension is still in the original spatial domain. As we can see from Fig. 2, there are only L_1 non-zero columns in the $N \times G_t$ hybrid-domain channel $\tilde{\mathbf{H}}_k$, since there are only L_1 channel paths between the BS and the RIS. Besides, the number of non-zero columns in the hybrid-domain channel $\tilde{\mathbf{H}}_k$ is determined by the number of AoDs at the BS according to the ray-based channel model [24].

The dictionary matrix $\mathbf{\Theta}_T$ can be represented as

$$\mathbf{\Theta}_T = \left[\mathbf{a}\left(\phi_1\right), \cdots, \mathbf{a}\left(\phi_{g_t}\right), \cdots, \mathbf{a}\left(\phi_{G_t}\right)\right], \tag{19}$$

where the discrete angle value of the g_t-th grid, which is also the angle of steering vector in the g_t-th column of the dictionary matrix $\mathbf{\Theta}_T$, can be expressed as $\phi_{g_t} = -\frac{1}{2} + \frac{g_t - 1}{G_t}$ with $s = 1, \cdots, G_t$. Note that the steering vector $\mathbf{a}^H\left(\phi_i^{\mathrm{AoD}}\right)$ is orthogonal with almost all columns of $\mathbf{\Theta}_T$ expect for the column with the grid angle ϕ_i^{AoD} [33].

For expression simplicity, we suppose that the value of ϕ_i^{AoD} at the BS is equal to the discrete angle value of the g_t-th angle grid, based on which we have $\mathbf{a}^H\left(\phi_i^{\mathrm{AoD}}\right)\mathbf{a}\left(\phi_{g_t}\right) = 1$. Thus, we represent the i-th non-zero column, which is also the g_t-th column of $\tilde{\mathbf{H}}_k$ as[5]

$$\tilde{\mathbf{h}}_{k,i} = \tilde{\mathbf{H}}_{k,(:,g_t)} = \sum_{j=1}^{L_2} g_{i,k,j}\mathbf{b}\left(\phi_{i,k,j}^{\mathrm{AoA}}, \theta_{i,k,j}^{\mathrm{AoA}}\right), \tag{20}$$

[5]Note that the relationship between g_t and i is not constant, since the grid index of the AoD, which determines the value of g_t, is not constant. Actually, we just use i to replace g_t in the manuscript for the sake of expression convenience. As long as there are L_1 paths between BS and the RIS, we always have $i = 1, 2, \cdots, L_1$.

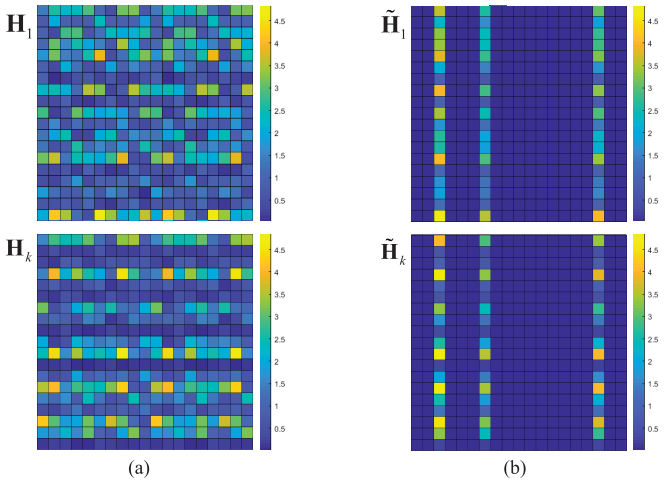

Fig. 2. The BS-RIS-UE cascaded channels: (a) in the spatial domain; (b) in the hybrid domain. The relationship between the spatial-domain channel and the hybrid-domain channel of the k-th UE can be expressed as $\mathbf{H}_k = \tilde{\mathbf{H}}_k \mathbf{\Theta}_T^H$. There are only a few non-zero columns in the hybrid-domain channel matrix, and different UEs share the same indexes of the non-zero columns, but these non-zero columns vectors with the same column index are different for different UEs, which is called as the single-structured sparsity in this paper.

where $g_t = 1, \cdots, G_t$ and $i = 1, \cdots, L_1$. We can further rewrite (20) as

$$\tilde{\mathbf{h}}_{k,i} = \mathbf{B}_{k,i}\mathbf{g}_{k,i}, \tag{21}$$

where $\mathbf{B}_{k,i} = \left[\mathbf{b}\left(\phi_{i,k,1}^{\mathrm{AoA}}, \theta_{i,k,1}^{\mathrm{AoA}}\right), \cdots, \mathbf{b}\left(\phi_{i,k,L_2}^{\mathrm{AoA}}, \theta_{i,k,L_2}^{\mathrm{AoA}}\right)\right] \in \mathbb{C}^{N \times L_2}$ is the steering matrix, and $\mathbf{g}_{k,i} = \left[g_{i,k,1}, \cdots, g_{i,k,L_2}\right]^T \in \mathbb{C}^{L_2 \times 1}$ denotes the complex gain vector of L_2 channel paths. Note that $\mathbf{b}\left(\phi_{i,k,l_2}^{\mathrm{AoA}}, \theta_{i,k,l_2}^{\mathrm{AoA}}\right)$ can be acquired according to (8).

It is worth to point that we illustrate the channel model under the assumption that the value of BS AoD is on the grid with angle resolution G_t in the above discussion, based on which the expression can be more convenience. Actually, the value of ϕ_i^{AoD} is randomly distributed in a continuous angle range and may not lie on any discrete angle grid. This condition has been adequately considered by the proposed scheme. For this realistic and general condition, we will approximate the value of AoD to the nearest discrete grid, which will result in the loss of channel feedback accuracy to some extent. Simulation results in Section IV show that the performance loss is negligible when we choose an appropriate angular resolution G_t.

2) Single-Structured Sparsity of Different UEs' BS-RIS-UE Cascaded Channels: The BS-RIS-UE cascaded channel consists of two parts: the *user-independent* BS-RIS channel and *user-specific* RIS-UE channel. Specifically, an important characteristic of the BS-RIS-UE cascaded channel is that, all UEs share the same user-independent BS-RIS channel, thus the BS-RIS-UE cascaded channels for different UEs only differ in the user-specific RIS-UE channel. In other words, different UEs share the user-independent AoDs at the BS, but they have their user-specific cascaded AoAs at the RIS. This channel property can be clearly illustrated by the hybrid-domain BS-RIS-UE cascaded channel in Fig. 2, where we can find that the limited non-zero columns of different UEs'

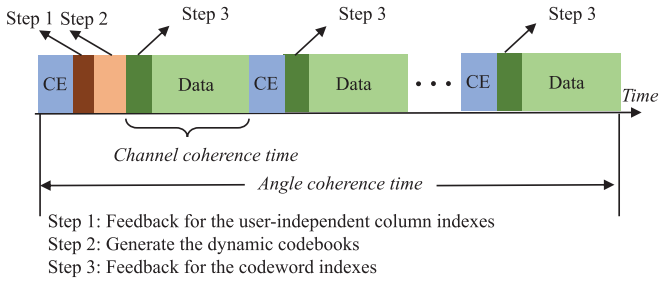

Fig. 3. Frame structure for the proposed dimension reduced channel feedback.

channel matrices have the same column indexes, which can be called as the structured sparsity. However, for a set of non-zero column vectors of different UEs' channel matrices with the same column index, those column vectors are made up of different elements, since the cascaded AoAs are different for different UEs' BS-RIS-UE cascaded channels. To sum up, the structured sparsity of the BS-RIS-UE cascaded channels of different UEs lies only in the column dimension, but not in the row dimension, which is called as "single-structured sparsity" in this paper.

According to (18), there is no essential difference to feed back the BS-RIS-UE cascaded channel $\mathbf{H}_k$ either in the spatial domain or $\tilde{\mathbf{H}}_k$ in the hybrid domain to realize the joint beamforming. However, the channel feedback overhead for those two channel matrices is quite different. We focus on the hybrid-domain channel $\tilde{\mathbf{H}}_k$ in this paper, since there are only L_1 non-zero column vectors in the hybrid-domain channel $\tilde{\mathbf{H}}_k$, and the single-structured sparsity of hybrid-domain channels is shared by all UEs. This channel characteristic does not exist in the spatial-domain channel $\mathbf{H}_k$. To reduce the channel feedback overhead, a dimension reduced channel feedback scheme will be proposed in the following Subsection III-B.

B. Proposed Dimension Reduced Channel Feedback

As mentioned above, to feed back the BS-RIS-UE cascaded channel, a multi-time feedback scheme can reduce the complexity for codebook design by splitting the channel matrix into many channel vectors. However, the multi-time feedback scheme will result in a much larger overhead, since the channel vector feedback procedure has to be carried out for N times (e.g., $N = 256$ [3]) within the channel coherence time. To reduce the channel feedback overhead, we propose a dimension reduced channel feedback scheme for the BS-RIS-UE cascaded channel with the help of single-structured sparsity. Based on the proposed scheme, we can carry out the multi-time feedback scheme only for L_1 times rather than N times, where L_1 is usually much smaller than N (e.g., $L_1 = 4$, $N = 256$).

Specifically, to feed the hybrid-domain channel $\tilde{\mathbf{H}}_k\,(k = 1, 2, \cdots, K)$ by the proposed dimension reduced channel feedback scheme, what we need to feedback are the non-zero column vectors and their indexes in the column dimension. In the proposed scheme, we decompose the task of channel feedback as the user-independent channel information feedback and user-specific channel information feedback. As shown in Fig. 3, the hybrid-domain channel $\tilde{\mathbf{H}}_k\ (k = 1, 2, \cdots, K)$ can be fed back to the BS by three steps in the proposed scheme after the downlink channel estimation (CE) [3], [28]–[31]. Note that it is very difficult to acquire the BS-RIS channel or RIS-UE channel separately, since the RIS cannot transmit or receive the pilot signal limited by the lacking of RF chains.[6] Hence, in the proposed scheme, we focus on the downlink BS-RIS-UE cascaded channel $\tilde{\mathbf{H}}_k \in \mathbb{C}^{N \times M}$. In the first step, by exploiting the single-structured sparsity, we feed the user-independent indexes of non-zero columns for all UEs' channel matrices back to the BS by only one UE. Then, the user-specific non-zero column vectors of different UEs' channel matrices are fed back to the BS by different UEs in the next two steps by utilizing dynamic codebooks. Without loss of generality, we will discuss the next two steps only for the k-th UE, since the same procedure will be carried out by all UEs. The dynamic codebooks will be generated in Step 2 to match the time-varying BS-RIS-UE cascaded channel more accurately, and the corresponding codewords for the non-zero column vectors will be selected from the dynamic codebooks in Step 3. The detailed design method of the dynamic codebooks will be discussed later. In Step 3, the indexes of the selected codewords for the corresponding non-zero column vectors, will be fed back to the BS, as mentioned in Subsection II-B.

To be more specific, we describe the three steps of the proposed dimension reduced channel feedback scheme as follows.

Step 1: Feedback for the User-Independent Columns Indexes: In this step, the L_1 user-independent indexes of the non-zero columns in the BS-RIS-UE cascaded channel $\tilde{\mathbf{H}}_k$ for all K UEs ($k = 1, 2, \cdots, K$) will be acquired and fed back to the BS. According to (18), the angular resolution of the dictionary matrix $\mathbf{\Theta}_T$ is G_t, i.e., the range of channel angles are divided by G_t grids. As mentioned in Subsection III-A, the number of column vectors in the hybrid-domain channel matrix is determined by the number of angle grids G_t, while the number of non-zero column vectors is determined by the number of channel paths between the BS and the RIS. In addition, (19) indicates that the index of the non-zero channel vector in the column dimension is also the grid index of the AoD at the BS of the corresponding channel path. In other words, we can acquire the L_1 indexes of the non-zero column vectors according to the grid indexes of L_1 AoDs at the BS of the downlink channel. Then, we can feed the indexes of the non-zero columns back to the BS via the uplink channel. In this way, the per-user overhead for channel feedback can be significantly reduced, since the number of UEs carrying out the channel feedback procedure for non-zero column indexes can be reduced from K to only 1.

Then, in the following two steps, the non-zero column vectors of different UEs' channel matrices will be fed back to the BS by different UEs, respectively.

Step 2: Generate the Dynamic Codebooks: Different from the existing channel feedback schemes with a constant codebook, we design different dynamic codebooks for the non-zero column vectors of $\tilde{\mathbf{H}}_k$ to match the time-varying channels more accurately. It is known that the channel variation is

[6]It is worth to note that the concept of active RIS has been proposed in recent research [34].

related to the changes of channel gains and the changes of channel angles, and the angle coherence time is much longer than the channel coherence time. Thus, we design dynamic codebooks by particularly considering the slow-varying channel angle information, based on which the dynamic codebooks will be only updated at the beginning of every angle coherence time and remain unchanged within the angle coherence time.

Specifically, taking the i-th non-zero column vector $\tilde{\mathbf{h}}_{k,i}$ of the hybrid-domain channel $\tilde{\mathbf{H}}_k$ as an example, we introduce the feature of the non-zero column vector and the design method of the corresponding dynamic codebook $\mathcal{C}_{k,i}$ as follows. As shown in Fig. 2, a non-zero column of the hybrid-domain channel $\tilde{\mathbf{H}}_k$ is not sparse in the row dimension, which results in a pretty high channel feedback overhead. To reduce the overhead for channel feedback, the characteristic of RIS-UE channel is exploited in the proposed scheme. Specifically, (20) indicates that $\tilde{\mathbf{h}}_{k,i}$ is composed of only L_2 channel paths between the RIS and the k-th UE. In other words, we can find from (21) that the N-dimensional non-zero column vector $\tilde{\mathbf{h}}_{k,i}$ is distributed in a vector space, which is spanned by the L_2 columns of the steering matrix $\mathbf{B}_{k,i} \in \mathbb{C}^{N\times L_2}$. This vector space is also referred to as the column space of $\mathbf{B}_{k,i}$. Note that the steering matrix is determined by the L_2 pairs of cascaded AoAs, which can be expressed as $\{(\phi_{i,k,j}^{\mathrm{AoA}}, \theta_{i,k,j}^{\mathrm{AoA}})\}_{j=1}^{L_2}$, and we usually have $L_2 \ll N$. Hence, the column space of the steering matrix $\mathbf{B}_{k,i}$ is just a subspace of a full N-dimensional space. Since the non-zero vector $\tilde{\mathbf{h}}_{k,i}$ is related to the angle information, we can design the dynamic codebook $\mathcal{C}_{k,i}$ with the help of the steering matrix $\mathbf{B}_{k,i}$. Note that the continuous cascaded AoAs with the range $\left[-\frac{\pi}{2}, \frac{\pi}{2}\right]$ usually need to be quantized for digital transmission in practice, the q-th codeword of the dynamic codebook $\mathcal{C}_{k,i}$ can be designed as

$$\mathbf{c}_{k,i,q} = \hat{\mathbf{B}}_{k,i}\mathbf{r}_{k,i,q}, \tag{22}$$

where $\mathbf{r}_{k,i,q} \in \mathbb{C}^{L_2\times 1}$ is a unit-norm vector chosen from the RVQ codebook [20] of size 2^B. Note that the RVQ codebook is per-determined and known both at the BS and the UE, and $\hat{\mathbf{B}}_{k,i} \in \mathbb{C}^{N\times L_2}$ is the quantized steering matrix which can be expressed as

$$\hat{\mathbf{B}}_{k,i} = \left[\mathbf{b}\left(\hat{\phi}_{i,k,1}^{\mathrm{AoA}}, \hat{\theta}_{i,k,1}^{\mathrm{AoA}}\right), \cdots, \mathbf{b}\left(\hat{\phi}_{i,k,L_2}^{\mathrm{AoA}}, \hat{\theta}_{i,k,L_2}^{\mathrm{AoA}}\right)\right]. \tag{23}$$

Note that $\hat{\mathbf{B}}_{k,i}$ is determined by the L_2 pairs of quantized spacial cascaded AoAs $\{(\hat{\phi}_{i,k,j}^{\mathrm{AoA}}, \hat{\theta}_{i,k,j}^{\mathrm{AoA}})\}_{j=1}^{L_2}$. For every cascaded AoA (e.g., $\phi_{i,k,j}^{\mathrm{AoA}}$), uniformly quantizing the value of continuous angle into a discrete grid with limited resolution is a simple but effective way to acquire the quantized cascaded AoA (i.e., $\hat{\phi}_{i,k,j}^{\mathrm{AoA}}$). This angle information $\{(\hat{\phi}_{i,k,j}^{\mathrm{AoA}}, \hat{\theta}_{i,k,j}^{\mathrm{AoA}})\}_{j=1}^{L_2}$ will be fed back to the BS by Step 2. With the L_2 pairs of quantized cascaded AoAs, the L_2 columns in $\hat{\mathbf{B}}_{k,i}$ can be calculated according to (8). The dynamic codebook $\mathcal{C}_{k,i}$ will be updated only once at the beginning of every angle coherence time, since the quantized steering matrix $\hat{\mathbf{B}}_{k,i}$ determined by the quantized cascaded AoAs remains constant during the angle coherence time.

In order to acquire the dynamic codebooks at the BS, the angle information at the UE also needs to be acquired at the BS by channel feedback. Thus, the L_2 pairs of quantized cascaded AoAs need to be fed back to the BS by the k-th UE via the uplink channel in this step.

By taking the same procedure for generating the dynamic codebook of the i-th non-zero column vector, the dynamic codebooks for the rest of non-zero column vectors can be acquired. Note that although we need to design different dynamic codebooks for different non-zero column vectors, it is not difficult to implement, since there are only a few non-zero columns in the hybrid-domain channel matrix. Moreover, although we have to generate the same codebook at the BS as that at the UE, we can achieve this goal with a low overhead, since we can just need to feed a few cascaded AoAs rather than the whole codebook back to the BS.

Step 3: Feedback for the Codeword Indexes: In this step, for non-zero column vectors of the hybrid-domain channel $\tilde{\mathbf{H}}_k$, we select codewords from the corresponding dynamic codebooks to approximate the non-zero column vectors. Then, the indexes of the codewords for these non-zero column vectors are fed back to the BS.

Specifically, taking the i-th non-zero column $\tilde{\mathbf{h}}_{k,i}$ of the hybrid-domain channel $\tilde{\mathbf{H}}_k$ as an example, we select the codeword for $\tilde{\mathbf{h}}_{k,i}$ from the dynamic vector codebook $\mathcal{C}_{k,i}$, which has been generated according to the angle information in Step 2. The selected codeword with the index $D_{k,i}$ is determined from

$$D_{k,i} = \underset{q\in[1,2^B]}{\arg\min} \sin^2\left(\measuredangle\left(\tilde{\mathbf{h}}_{k,i}, \mathbf{c}_{k,i,q}\right)\right). \tag{24}$$

An intuitive explanation for (24) is that, what we want to select from the codebook is an appropriate codeword $\mathbf{c}_{k,i,q}$ whose vector direction is the closest to the direction of channel vector $\tilde{\mathbf{h}}_{k,i}$ in the N-dimensional space. After selecting the codeword, we can feed back the codeword index $D_{k,i}$ to the BS with the feedback overhead of B bits. Similarly, the codeword indexes for other non-zero columns can be fed back to the BS in the same procedure like the operation for the i-th non-zero column.

After the above three steps, the BS-RIS-UE cascaded channel $\tilde{\mathbf{H}}_k$ will be recovered at the BS as follows. According to Step 1, the user-independent indexes of the non-zero column vectors can be acquired at the BS. Then, with the quantized cascaded AoAs fed back in Step 2, the BS can acquire the user-specific dynamic codebooks for all UEs. Next, with the codeword indexes from the k-th UE in Step 3, the BS can select the codewords from the corresponding dynamic codebooks to approximate the non-zero column vectors for all UEs' channel matrices. Finally, those non-zero channel vectors are arranged into the corresponding columns indicated by the indexes for non-zero columns to reconstruct the whole hybrid-domain channel $\tilde{\mathbf{H}}_k$. Moreover, the spatial-domain BS-RIS-UE cascaded channel $\mathbf{H}_k$, which can be used for the joint beamforming optimization in the RIS-aided wireless communication system, will be completely recovered as $\mathbf{H}_k = \tilde{\mathbf{H}}_k\mathbf{\Theta}_T^H$ according to (18).

Finally, we discuss the channel feedback overhead of the proposed dimension reduced channel feedback scheme from three aspects. Most importantly, by exploiting the single-structured sparsity, the number of UEs who carry out the channel feedback for indexes of the non-zero columns can be reduced from K to only 1. Moreover, the number of column

vectors which needs to be fed back can be reduced from the number of antennas M to the number of channel paths L_1 of BS-RIS channel for each UE. In addition, as shown in Fig. 3, we can carry out the first two steps only once during a large angle coherence time, since the change of angles is significantly slower than that of channel gains, based on which the channel feedback overhead can be further reduced.

The per-user overhead for the proposed channel feedback scheme consists of three parts: the feedback overhead for the user-independent indexes of non-zero columns, the feedback overhead for the user-specific cascaded AoAs, and the feedback overhead for the codeword indexes. The per-user overhead p for channel feedback of those three parts can be calculated as

$$p=\lceil \frac{\log_2(G_t)\times L_1\times(\varpi K)}{\chi\times K}+\frac{B_0\times 2\times L_1\times L_2}{\chi}+BL_1\rceil, \tag{25}$$

where the coefficient χ represents the ratio between the angle coherence time and the channel coherence time. To enhance the robustness of the user-independent channel information feedback in the practical systems, we can appoint ϖK UEs to simultaneously feed the user-independent indexes back to the BS, and the selected UE proportion ϖ for Step 1 can be neatly adjusted according to different transmission scenarios. Certainly, according to the proposed channel feedback scheme, it is feasible to appoint only one UE for Step 1 as we have discussed before.

C. Theoretical Analysis for the Proposed Scheme

In this subsection, we provide the theoretical analysis of the proposed dimension reduced channel feedback scheme. We focus on the analysis for quantization error, based on which the overhead for the proposed channel feedback scheme will also be exploited.

1) We analyze the quantization error for the non-zero column vectors at first. Without loss of generality, we take the $\tilde{\mathbf{h}}_{k,i}$ as an example. The quantization error ϱ between the original channel vector $\tilde{\mathbf{h}}_{k,i}$ and fed back channel vector $\hat{\tilde{\mathbf{h}}}_{k,i}$ can be expressed as

$$\begin{aligned}\varrho &= \mathbb{E}\left[\sin^2\left(\measuredangle\left(\hat{\tilde{\mathbf{h}}}_{k,i},\tilde{\mathbf{h}}_{k,i}\right)\right)\right]=1-\mathbb{E}\left[\frac{<\hat{\tilde{\mathbf{h}}}_{k,i},\tilde{\mathbf{h}}_{k,i}>^2}{\|\hat{\tilde{\mathbf{h}}}_{k,i}\|_2^2\|\tilde{\mathbf{h}}_{k,i}\|_2^2}\right]\\ &=1-\mathbb{E}\left[\frac{\left|\|\hat{\tilde{\mathbf{h}}}_{k,i}\|_2\mathbf{c}_{k,i,q}^H\mathbf{B}_{k,i}\mathbf{g}_{k,i}\right|^2}{\|\hat{\tilde{\mathbf{h}}}_{k,i}\|_2^2\|\tilde{\mathbf{h}}_{k,i}\|_2^2}\right]\\ &=1-\mathbb{E}\left[\left|\mathbf{r}_{k,i,q}^H\hat{\mathbf{B}}_{k,i}^H\mathbf{B}_{k,i}\tilde{\mathbf{g}}_{k,i}\right|^2\right],\end{aligned} \tag{26}$$

where $\tilde{\mathbf{g}}_{k,i}=\mathbf{g}_{k,i}/\|\tilde{\mathbf{h}}_{k,i}\|_2$.

Remark 1: Consider that $\mathbf{r}_{k,i,q}\in\mathbb{C}^{L_2\times 1}$ is randomly selected from the RVQ codebook, based on which the quantization error can be upper-bounded as

$$\varrho_{Dyna}=\mathbb{E}\left[\sin^2\left(\measuredangle\left(\hat{\tilde{\mathbf{h}}}_{k,i},\tilde{\mathbf{h}}_{k,i}\right)\right)\right]\overset{(a)}{<}2^{-\frac{B}{L_2-1}}=\bar{\varrho}, \tag{27}$$

where (a) comes from $\mathbb{E}\left[\left|\mathbf{r}_{k,i,q}^H\tilde{\mathbf{g}}_{k,i}\right|^2\right]>1-2^{-\frac{B}{L_2-1}}$ according to [14].

Based on (27), for a given ϱ_0, the overhead for feeding the codeword index in Step 3 is required as

$$B_{Dyna}>(L_2-1)\log_2\frac{1}{\varrho_0}. \tag{28}$$

According to (28), we can find that the required overhead by utilizing the dynamic codebook is proportional to the number of paths L_2. If we directly utilize the RVQ codebook to quantize the N-dimensional non-zero column vector $\tilde{\mathbf{h}}_{k,i}$, the quantization error can be represented as

$$\varrho_{Const}=1-\mathbb{E}\left[\left|\hat{\tilde{\mathbf{e}}}_{k,i},\tilde{\mathbf{e}}_{k,i}\right|^2\right], \tag{29}$$

where the N-dimensional unit-norm vector $\hat{\tilde{\mathbf{e}}}_{k,i}$ and $\tilde{\mathbf{e}}_{k,i}$ denote the direction of the channel vector $\hat{\tilde{\mathbf{h}}}_{k,i}$ and $\tilde{\mathbf{h}}_{k,i}$, respectively. Similarly, the overhead for the index feedback with the constant codebook is required as

$$B_{Const}>(N-1)\log_2\frac{1}{\varrho_0}. \tag{30}$$

From (28) and (30), we can observe that for a certain quantization error, the codebook size of the dynamic codebook and the constant codebook are proportional to 2^{L_2} and 2^N, respectively. Since $L_2\ll N$, the overhead for feeding back the non-zero vector can be significantly reduced when we utilizing the dynamic codebook.

2) Actually, the analysis for the upper bound in Remark 1 is calculated with the restriction $\hat{\mathbf{B}}_{k,i}^H\mathbf{B}_{k,i}=\mathbf{I}_{L_2}$. This restriction is considered with perfect feedback for cascaded AoAs, i.e., $\hat{\mathbf{B}}_{k,i}=\mathbf{B}_{k,i}$. However, since the feedback overhead B_0 for the cascaded AoA is limited, the equivalence relation between $\mathbf{b}\left(\hat{\phi}_{i,k,l_2}^{\mathrm{AOA}}\right)$ and $\mathbf{b}\left(\phi_{i,k,l_2}^{\mathrm{AOA}}\right)$ no longer holds.[7] In this condition, we have $\hat{\mathbf{B}}_{k,i}^H\mathbf{B}_{k,i}\approx\Upsilon\mathbf{I}_{L_2}$. Following the analysis in [35], we have

$$|\Upsilon|^2\geq 1-\frac{\left(\pi\zeta\frac{d}{\lambda}\right)^2}{3}N^2 2^{-2B_0}, \tag{31}$$

where ζ is the difference between the maximum and minimum value for cascaded AoAs. Hence, the quantization error can be rewritten as

$$\begin{aligned}\varrho &\approx 1-|\Upsilon|^2\,\mathbb{E}\left[\left|\mathbf{r}_{k,i,q}^H\tilde{\mathbf{g}}_{k,i}\right|^2\right]\\ &\leq\left(1-2^{-\frac{B}{L_2-1}}\right)\nu+2^{-\frac{B}{L_2-1}},\end{aligned} \tag{32}$$

where we denote $\nu=\frac{\left(\pi\zeta\frac{d}{\lambda}\right)^2}{3}N^2 2^{-2B_0}$ for the sake of convenience.

Remark 2: We can find that consider the quantization and limited feedback for the cascaded AoA, the upper bound of the error ϱ is related to B_0. For a given quantization error ϱ_0, we can calculate that B_0 is required as

$$B_0\geq\log_2\frac{\pi\zeta dN}{\sqrt{3}\lambda}-\frac{1}{2}\log_2(1-\frac{1-\varrho_0}{1-\bar{\varrho}}), \tag{33}$$

based on which we can observe that if we want to reduce the quantization error ϱ_0, it is essential to quantize the cascaded AoA with a larger quantization bits B_0. When the allowable upper bound ϱ_0 is close to the original error $\bar{\varrho}$ in (27),

[7] We consider the ULA deployment for RIS in this subsection.

i.e., $\varrho_0 \to \bar{\varrho}$, we have $B_0 \to \infty$, which means that the cascaded AoA should be perfectly quantized, which is consistent with the discussion in Remark 1.

3) Based on (16), the spatial-domain channel $\mathbf{H}_k \in \mathbb{C}^{N\times M}$ can be expressed by block matrix multiplication as

$$\mathbf{H}_k = \left[\tilde{\mathbf{h}}_{k,1}, \cdots, \tilde{\mathbf{h}}_{k,L_1}\right] \begin{bmatrix} \mathbf{a}^H\left(\phi_1^{\mathrm{AoD}}\right) \\ \vdots \\ \mathbf{a}^H\left(\phi_{L_1}^{\mathrm{AoD}}\right) \end{bmatrix} \triangleq \mathbf{\Xi}\mathbf{A}^H, \tag{34}$$

where the block matrix $\mathbf{\Xi} = \left[\tilde{\mathbf{h}}_{k,1}, \cdots, \tilde{\mathbf{h}}_{k,L_1}\right]$ is composed by the L_1 non-zero column vectors of $\tilde{\mathbf{H}}_k$ and $\mathbf{A} = \left[\mathbf{a}\left(\phi_1^{\mathrm{AoD}}\right), \cdots, \mathbf{a}\left(\phi_{L_1}^{\mathrm{AoD}}\right)\right]^H$ is the steering matrix. Similarly, for the proposed channel feedback scheme, the spatial-domain channel calculated by the fed back hybird-doamin channel $\hat{\tilde{\mathbf{H}}}_k$ at the BS can be expressed as $\hat{\mathbf{H}}_k = \hat{\tilde{\mathbf{H}}}_k \mathbf{\Theta}_T^H$ with $\hat{\tilde{\mathbf{H}}}_k \in \mathbb{C}^{N\times G_t}$. It is important to point that ϕ_i^{AoD} is quantized to the nearest angle grid ϕ_{gti} selected from the G_t grids by the proposed scheme in Step 1, with limited error $\delta_{AoD} = \phi_i^{\mathrm{AoD}} - \phi_{gti}$. The subscript g_{ti} is the index for the selected grid. The feedback channel $\hat{\mathbf{H}}_k$ at the BS can be rewritten as

$$\begin{aligned}\hat{\mathbf{H}}_k &= \left[\hat{\tilde{\mathbf{H}}}_{k,(:,g_{t1})}, \cdots, \hat{\tilde{\mathbf{H}}}_{k,(:,g_{tL_1})}\right] \begin{bmatrix} \mathbf{a}^H\left(\phi_{g_{t1}}\right) \\ \vdots \\ \mathbf{a}^H\left(\phi_{g_{tL_1}}\right) \end{bmatrix} \\ &\triangleq \hat{\mathbf{\Xi}}\hat{\mathbf{A}}^H. \end{aligned} \tag{35}$$

In this paper, we evaluate the quantization error between $\mathbf{H}_k$ and $\hat{\mathbf{H}}_k$ by calculating the chordal distance as a metric [36], which is a widely adopted metric of the quantization error for channel feedback research [37], [38]. The chordal distance between $\mathbf{H}_k$ and $\hat{\mathbf{H}}_k$ is denoted as

$$d(\mathbf{H}_k, \hat{\mathbf{H}}_k) = \sqrt{\sum_{j=1}^{N} \sin^2 \vartheta_j}, \tag{36}$$

where ϑ_j denotes the j-th principal angle between the two subspaces spanned by the N rows of the matrices $\mathbf{H}_k$ and $\hat{\mathbf{H}}_k$, respectively. The calculation for the principal angles can be exploited in [36], based on which we have

$$d(\mathbf{H}_k, \hat{\mathbf{H}}_k) = \left[N - \|\mathbf{H}_k \hat{\mathbf{H}}_k^H\|_F^2\right]^{1/2}. \tag{37}$$

Hence, combining (34), (35), and (37), the quantization error D evaluated by the chordal distance can be calculated as

$$D \triangleq \mathbb{E}\left[d(\mathbf{H}_k, \hat{\mathbf{H}}_k)\right]^2 = \mathbb{E}\left[N - \|\mathbf{\Xi}\mathbf{A}^H \hat{\mathbf{A}} \hat{\mathbf{\Xi}}^H\|_F^2\right]. \tag{38}$$

It is shown in [37] that the upper bound $\bar{D}$ is

$$D \le \bar{D} \approx \frac{\Gamma\left(\frac{1}{T}\right)}{T} \left(C_{L_1 N}\right)^{-\frac{1}{T}} 2^{-\frac{B}{T}}, \tag{39}$$

where $T = N(N - L_1)$, and $C_{L_1 N}$ is denoted by $\frac{1}{T!}\prod_{i=1}^{N} \frac{(L_1 - i)!}{(N-i)!}$. The Gamma function $\Gamma(x)$ is defined as $\Gamma(x) = \int_0^\infty t^{x-1} e^{-t} dt$ for any real number x.

Remark 3: It is worth to point that (39) holds with the assumption for discrete and on-grid angle, i.e., $\delta_{AoD} = 0$. In other words, we did not consider the feedback error from Step 1 in the above two Remarks. Actually, the AoD is randomly generated in the continuous range rather than a discrete angle grid. Hence, the discretization for AoD in Step 1 will bring the quantization error since $\mathbf{A} \neq \hat{\mathbf{A}}$. This error will be propagated to Step2 and Step 3, which is called as the error propagation effect in this paper. Taking the error propagation effect into account, the upper bound D_1 should be calculated similar to (32) as

$$D_1 \le \bar{D} + (N - \bar{D}) \frac{M^2}{3G_t^2} \iota, \tag{40}$$

where $\iota = \left(\pi \varsigma \frac{d}{\lambda}\right)^2$ can be obtained similar to ν, and ς is the difference between the maximum and minimum value for AoDs. We can observe from (40) that the quantization error rises caused by the discretization for AoD compared with (39). According to the channel model of this paper, the value for ι is fixed. Hence, $G_t > M$ is usually essential for our proposed scheme to reduce the feedback error brought by the AoD discretization, which can also be illustrated by Fig. 5.

IV. Simulation Results

In this section, we provide the simulation results of the proposed dimension reduced channel feedback scheme in the RIS-aided wireless communication system via Monte Carlo simulation. The number of BS antennas, RIS elements, and single-antenna UEs are set as $M = 64$, $N = 256$, and $K = 4$, respectively. The value of AoDs at the BS and the cascaded AoAs at the RIS are randomly distributed in the continuous angle range according to the channel model in Section II. The number of paths from the BS to the RIS and that from the RIS to the UE are set as $L_1 = 4$ and $L_2 = 2$, respectively. The UE proportion ϖ is set as $\varpi = 1/K = 25\%$, i.e., we consider the fully discussed scenario in this paper with one UE to carry out Step 1. According to the analysis about the angle coherence time in [20], the ratio χ is set as 10. The receiver SNR is set as 5 dB.

We choose the achievable sum-rate to evaluate the performance of the proposed channel feedback scheme. After acquiring the downlink CSI via channel feedback, cross entropy optimization (CEO) is utilized to optimize the joint beamforming at the BS and RIS [39]. Then, the achievable sum-rate rate can be expressed as [19]

$$R = \sum_{k=1}^{K} \log_2 \left(1 + \frac{\frac{\varepsilon}{K}\left|\mathbf{h}_{\mathrm{DL},k}^H \hat{\mathbf{v}}_k\right|^2}{1 + \frac{\varepsilon}{K}\sum_{i=1, i\neq k}^{K}\left|\mathbf{h}_{\mathrm{DL},k}^H \hat{\mathbf{v}}_i\right|^2}\right), \tag{41}$$

where ε is the transmit power, and $\hat{\mathbf{v}}_k \in \mathbb{C}^{M\times 1}$ is the k-th column of the normalized precoding matrix at the BS. It is worth to note that the reflecting diagonal matrix $\mathbf{P}$ for the RIS has been included in the equivalent channel $\mathbf{h}_{\mathrm{DL},k}$ as shown in (3).

Fig. 4 shows the achievable sum-rate comparison between the proposed dimension reduced channel feedback scheme and the conventional statistics-based channel feedback scheme [15] against the per-user overhead for channel feedback. We also provide the upper bound of the achievable sum-rate, which is achieved by the perfect channel state information at the transmitter (CSIT) with ideal channel feedback. For the results

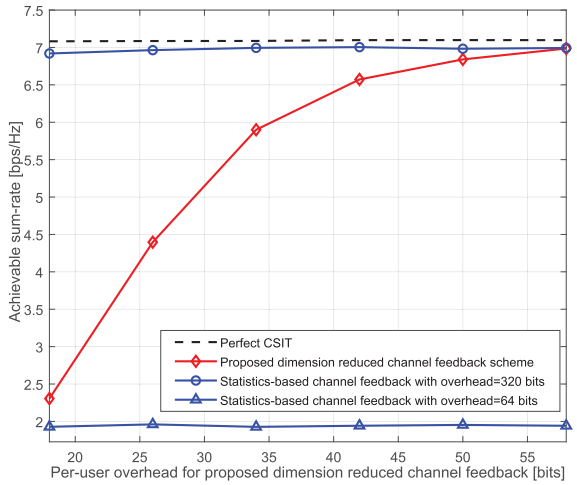

Fig. 4. The achievable sum-rate comparison between the proposed dimension reduced channel feedback scheme and the conventional feedback scheme.

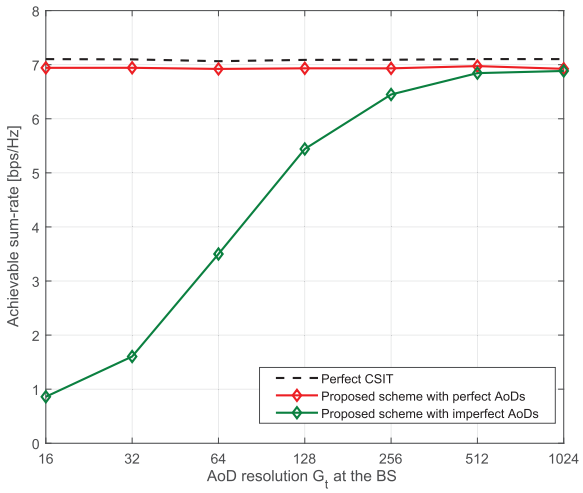

Fig. 5. The achievable sum-rate of the dimension reduced channel feedback scheme against the grid resolution G_t of AoD at the BS.

in this figure, we configure the grid resolution of AoD as $G_t = 1024$ and the quantization bits of the cascaded AoA as $B_0 = 8$ bits, and the rationality of this configuration will be observed in subsequent simulation results. Then, we enlarge the size of dynamic codebooks B from 1 bit to 11 bits, which will result in the increase of per-user overhead for channel feedback. We can find from Fig. 4 that the achievable sum-rate will be improved with the increase of the channel feedback overhead. By utilizing the proposed channel feedback scheme, the achievable sum-rate has a small gap of only 0.1 bps/Hz compared with the upper bound when the dynamic codebooks size is $B = 11$ bits, based on which the per-user overhead for the proposed channel feedback scheme can be calculated as $p = 58$ bits. However, to achieve this performance with the same achievable sum-rate gap by utilizing the conventional statistics-based channel feedback scheme, the per-user overhead for channel feedback should be 320 bits, which is 5 times of that of the proposed dimension reduced channel feedback scheme. In other words, the per-user overhead for the proposed channel feedback scheme can be reduced by more than 80% compared with the conventional statistics-based channel feedback scheme, while the same performance can still be guaranteed. On the contrary, the overhead for conventional channel feedback scheme [15] is an integer multiple for the number of BS antennas M. To evaluate the channel feedback performance with a similar overhead with $p = 58$ bits, when the overhead of 64 bits is provided for the conventional channel feedback scheme, the achievable sum-rate is only 28% of the proposed scheme. In other words, this achievable sum-rate has a gap of more than 5.13 bps/Hz compared with the upper bound.

Fig. 5 illustrates the impacts of the AoD resolution G_t at the BS on the achievable sum-rate. We adopt the quantization bits of the codeword index in the vector codebook as $B = 11$ bits according to Fig. 4. The RIS cascaded AoA quantization bits is still set as $B_0 = 8$ bits. To feed the user-independent non-zero column indexes back to the BS in Step 1, we can feed the grid index of AoD to represent the non-zero column index as mentioned in Section III. Fig. 5 shows that the achievable sum-rate of the proposed scheme has a small gap of less than 0.1 bps/Hz compared with the upper bound. However, the angles of channel paths are usually continuous rather than discretely distributed in the angular grids. Under this practical condition, the AoD of the channel path at the BS must be approximated to the nearest grid, and this procedure will result in the performance loss to some extent. Thus, the number of the grids, which is determined by the resolution G_t, will affect the channel feedback performance and consequently affect the achievable sum-rate. From Fig. 5 we can find that the achievable sum-rate will be improved with the increase of AoD resolution G_t at the BS. It is shown in Fig. 5 that when $G_t = 1024$, the performance with imperfect AoDs can almost approach the achievable sum-rate of the proposed scheme with perfect AoDs. It is worth to note that the required resolution G_t is usually much bigger than M to overcome the error propagation effect discussed in Subsection III-C. However, the grid indexes of the AoDs need to be fed back only once during the large angle coherence time, so the corresponding overhead for channel feedback is fairly low. As mentioned before, thanks to the single-structured sparsity, even if a quarter of UEs are selected to feed the grid indexes of AoDs back to the BS to improve the robustness of channel feedback, the per-user overhead for channel feedback of Step 1 is only $\lceil L_1 \times \log_2(G_t) \times \varpi/\chi \rceil = 1$ bit.

Fig. 6 presents the achievable sum-rate of the proposed dimension reduced channel feedback scheme against the number of cascaded AoA quantization bits B_0. In this figure, the quantization bits of the codeword index in the vector codebook is set as $B = 11$ bits, and the AoD resolution at the BS is set as $G_t = 1024$. It is clear from Fig. 6 that with the increase of B_0, the achievable sum-rate increases rapidly. It is worth to point out that when $B_0 = 8$ bits, the achievable sum-rate of the proposed scheme with imperfect AoAs has almost no difference with that in the ideal case of perfect AoAs. Both of them are very close to the upper bound with perfect CSIT. Since the quantized AoAs are only required to be fed back once during the large angle coherence time, the associated overhead is also very low for the proposed channel feedback scheme. The per-user overhead for channel feedback of Step 2 can be calculated as $\lceil \frac{B_0 \times 2 \times L_1 \times L_2}{\chi} \rceil = 13$ bits, where the factor 2 in numerator indicates that there are two cascaded

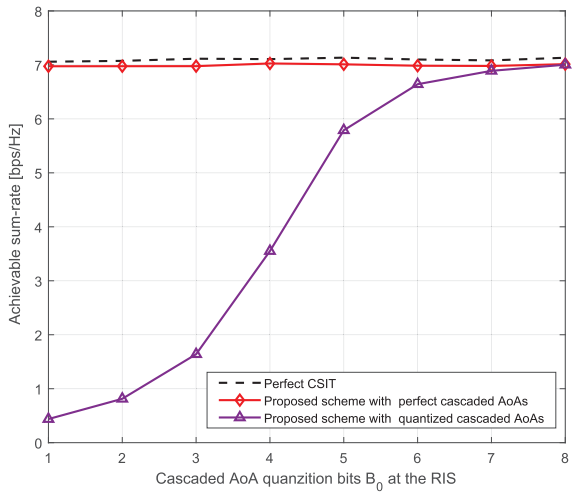

Fig. 6. The achievable sum-rate of the dimension reduced channel feedback scheme against the cascaded AoA quantization bits B_0.

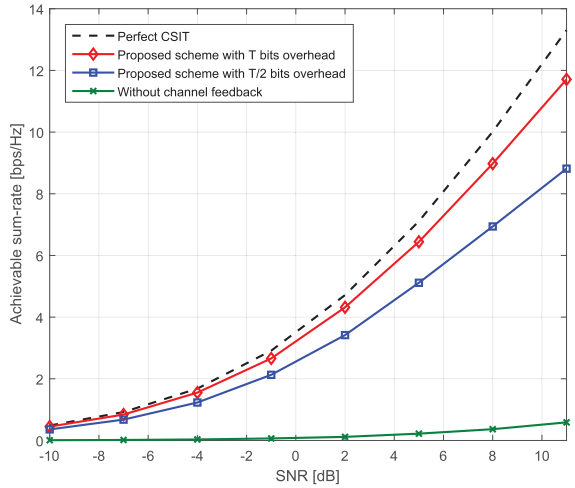

Fig. 7. The achievable sum-rate of the proposed scheme with different overheads against SNR.

AoAs (azimuth AoA and elevation AoA) of one channel path, since usually a UPA is deployed at the RIS [3].

We analyse the achievable sum-rate for the proposed scheme with different overheads against SNR in Fig. 7, based on which the trade-off between the overhead and accuracy for channel feedback can be illustrated. Specifically, in the high-SNR scenario, the inter-user interference is the main effect compared to the system noise, based on which the performance gap between the proposed scheme with high overhead and low overhead is obvious. This is mainly because that the joint beamforming, which can reduce the influence for inter-user interference, relies on the accuracy for channel feedback. However, in the low-SNR scenario, the system noise is the dominant interference, based on which the achievable sum-rate is not sensitive to the accuracy for channel feedback and the mentioned performance gap will become small.

V. Conclusion

In this paper, we investigated the channel feedback problem for RIS-aided wireless communications to reduce the channel feedback overhead by exploiting the single-structured sparsity of the BS-RIS-UE cascaded channel. We divided the channel feedback into the user-independent channel information (i.e., the indexes of the non-zero columns) feedback and user-specific channel information (i.e., the non-zero column vectors) feedback. In the proposed scheme, the user-independent channel information for all UEs can be fed back to the BS by only one UE. Then, the user-specific channel information can be fed back to the BS with the help of codebook-based feedback scheme by different UEs, respectively. Moreover, to match the time-varying BS-RIS-UE cascaded channel more accurately, we designed dynamic codebooks based on the channel angle information rather than utilizing a constant codebook. Simulation results showed that the proposed dimension reduced channel feedback scheme could reduce the channel feedback overhead by more than 80%, while the near-optimal achievable sum-rate could be guaranteed. We leave the channel feedback with the influence for imperfect downlink channel estimation in RIS-aided cell-free networks [40] for future work, in which it may be unbearable to estimate the CSI with high accuracy among every UE, every RIS and every BS.

References

[1] D. Shen and L. Dai, "Channel feedback for reconfigurable intelligent surface assisted wireless communications," in *Proc. GLOBECOM IEEE Global Commun. Conf.*, Dec. 2020, pp. 1–5.

[2] Z. Zhang *et al.*, "6G wireless networks: Vision, requirements, architecture, and key technologies," *IEEE Veh. Technol. Mag.*, vol. 14, no. 3, pp. 28–41, Sep. 2019.

[3] Q. Wu, S. Zhang, B. Zheng, C. You, and R. Zhang, "Intelligent reflecting surface-aided wireless communications: A tutorial," *IEEE Trans. Commun.*, vol. 69, no. 5, pp. 3313–3351, May 2021.

[4] E. Basar, M. Di Renzo, J. De Rosny, M. Debbah, M. Alouini, and R. Zhang, "Wireless communications through reconfigurable intelligent surfaces," *IEEE Access*, vol. 7, pp. 116753–116773, 2019.

[5] W. Yan, X. Yuan, Z.-Q. He, and X. Kuai, "Passive beamforming and information transfer design for reconfigurable intelligent surfaces aided multiuser MIMO systems," *IEEE J. Sel. Areas Commun.*, vol. 38, no. 8, pp. 1793–1808, Aug. 2020.

[6] S. Wang, Q. Li, S. X. Wu, and J. Lin, "Sum rate maximization for multiuser MISO downlink with intelligent reflecting surface," 2019, *arXiv:1912.09315*. [Online]. Available: http://arxiv.org/abs/1912.09315

[7] L. Dai *et al.*, "Reconfigurable intelligent surface-based wireless communications: Antenna design, prototyping, and experimental results," *IEEE Access*, vol. 8, pp. 45913–45923, 2020.

[8] C. Pradhan, A. Li, L. Song, B. Vucetic, and Y. Li, "Hybrid precoding design for reconfigurable intelligent surface aided mmWave communication systems," *IEEE Wireless Commun. Lett.*, vol. 9, no. 7, pp. 1041–1045, Jul. 2020.

[9] X. Gao, L. Dai, C. Yuen, and Z. Wang, "Turbo-like beamforming based on Tabu search algorithm for millimeter-wave massive MIMO systems," *IEEE Trans. Veh. Technol.*, vol. 65, no. 7, pp. 5731–5737, Jul. 2016.

[10] C. Hu, L. Dai, S. Han, and X. Wang, "Two-timescale channel estimation for reconfigurable intelligent surface aided wireless communications," *IEEE Trans. Commun.*, early access, Apr. 12, 2021, doi: 10.1109/TCOMM.2021.3072729.

[11] X. Wei, D. Shen, and L. Dai, "Channel estimation for RIS assisted wireless communications—Part II: An improved solution based on double-structured sparsity," *IEEE Commun. Lett.*, vol. 25, no. 5, pp. 1403–1407, May 2021.

[12] Z.-Q. 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.

[13] A. Taha, M. Alrabeiah, and A. Alkhateeb, "Enabling large intelligent surfaces with compressive sensing and deep learning," *IEEE Access*, vol. 9, pp. 44304–44321, 2021.

[14] N. Jindal, "MIMO broadcast channels with finite-rate feedback," *IEEE Trans. Inf. Theory*, vol. 52, no. 11, pp. 5045–5060, Nov. 2006.

[15] W. Shen, L. Dai, Y. Zhang, J. Li, and Z. Wang, "On the performance of channel-statistics-based codebook for massive MIMO channel feedback," *IEEE Trans. Veh. Technol.*, vol. 66, no. 8, pp. 7553–7557, Aug. 2017.

[16] P. H. Kuo, H. T. Kung, and P. A. Ting, "Compressive sensing based channel feedback protocols for spatially-correlated massive antenna arrays," in *Proc. IEEE Wireless Commun. Netw. Conf. (WCNC)*, Apr. 2012, pp. 492–497.

[17] Z. Liu, L. Zhang, and Z. Ding, "An efficient deep learning framework for low rate massive MIMO CSI reporting," *IEEE Trans. Commun.*, vol. 68, no. 8, pp. 4761–4772, Aug. 2020.

[18] Z. Lu, J. Wang, and J. Song, "Multi-resolution CSI feedback with deep learning in massive MIMO system," in *Proc. IEEE Int. Conf. Commun. (ICC)*, Jun. 2020, pp. 1–6.

[19] J. Jang, H. Lee, S. Hwang, H. Ren, and I. Lee, "Deep learning-based limited feedback designs for MIMO systems," *IEEE Wireless Commun. Lett.*, vol. 9, no. 4, pp. 558–561, Apr. 2020.

[20] W. Shen, L. Dai, B. Shim, Z. Wang, and R. W. Heath, Jr., "Channel feedback based on AoD-adaptive subspace codebook in FDD massive MIMO systems," *IEEE Trans. Commun.*, vol. 66, no. 11, pp. 5235–5248, Nov. 2018.

[21] 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*. [Online]. Available: http://arxiv.org/abs/1912.03619

[22] 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.

[23] B. Zheng *et al.*, "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.

[24] Q.-U.-U. Nadeem, A. Kammoun, A. Chaaban, M. Debbah, and M.-S. Alouini, "Intelligent reflecting surface assisted wireless communication: Modeling and channel estimation," 2019, *arXiv:1906.02360*. [Online]. Available: http://arxiv.org/abs/1906.02360

[25] F. Rusek *et al.*, "Scaling up MIMO: Opportunities and challenges with very large arrays," *IEEE Signal Process. Mag.*, vol. 30, no. 1, pp. 40–60, Jan. 2013.

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

[27] B. Zheng, C. You, and R. Zhang, "Efficient channel estimation for double-IRS aided multi-user MIMO system," *IEEE Trans. Commun.*, vol. 69, no. 6, pp. 3818–3832, Jun. 2021.

[28] J. He, M. Leinonen, H. Wymeersch, and M. Juntti, "Channel estimation for RIS-aided mmWave MIMO systems," in *Proc. GLOBECOM IEEE Global Commun. Conf.*, Dec. 2020, pp. 1–6.

[29] D. Zhu, J. Choi, and R. W. Heath, Jr., "Auxiliary beam pair enabled AoD and AoA estimation in closed-loop large-scale millimeter-wave MIMO systems," *IEEE Trans. Wireless Commun.*, vol. 16, no. 7, pp. 4770–4785, Jul. 2017.

[30] K. Ardah, S. Gherekhloo, A. L. F. de Almeida, and M. Haardt, "TRICE: A channel estimation framework for RIS-aided millimeter-wave MIMO systems," *IEEE Signal Process. Lett.*, vol. 28, pp. 513–517, Feb. 2021.

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

[32] Z. Xiao, T. He, P. Xia, and X.-G. Xia, "Hierarchical codebook design for beamforming training in millimeter-wave communication," *IEEE Trans. Wireless Commun.*, vol. 15, no. 5, pp. 3380–3392, May 2016.

[33] A. Alkhateeb, O. El Ayach, G. Leus, and R. W. Heath, Jr., "Channel estimation and hybrid precoding for millimeter wave cellular systems," *IEEE J. Sel. Topics Signal Process.*, vol. 8, no. 5, pp. 831–846, Oct. 2014.

[34] Z. Zhang *et al.*, "Active RIS vs passive RIS: Which will prevail in 6G?" 2021, *arXiv:2103.15154*. [Online]. Available: http://arxiv.org/abs/2103.15154

[35] X. Gao, L. Dai, Y. Ma, and Z. Wang, "Low-complexity near-optimal signal detection for uplink large-scale MIMO systems," *Electron. Lett.*, vol. 50, no. 18, pp. 1326–1328, Aug. 2014.

[36] I. S. Dhillon, R. W. Heath, Jr., T. Strohmer, and J. A. Tropp, "Constructing packings in Grassmannian manifolds via alternating projection," *Experim. Math.*, vol. 17, no. 1, pp. 9–35, 2008.

[37] N. Ravindran and N. Jindal, "Limited feedback-based block diagonalization for the MIMO broadcast channel," *IEEE J. Sel. Areas Commun.*, vol. 26, no. 8, pp. 1473–1482, Oct. 2008.

[38] S. Schwarz, M. Rupp, and S. Wesemann, "Grassmannian product codebooks for limited feedback massive MIMO with two-tier precoding," *IEEE J. Sel. Topics Signal Process.*, vol. 13, no. 5, pp. 1119–1135, Sep. 2019.

[39] X. Gao, L. Dai, Y. Sun, S. Han, and C.-L. I, "Machine learning inspired energy-efficient hybrid precoding for mmWave massive MIMO systems," in *Proc. IEEE Int. Conf. Commun. (ICC)*, Paris, France, May 2017, pp. 1–6.

[40] Z. Zhang and L. Dai, "A joint precoding framework for wideband reconfigurable intelligent surface-aided cell-free network," *IEEE Trans. Signal Process.*, early access, Jun. 17, 2021, doi: 10.1109/TSP.2021.3088755.

Decai Shen received the B.E. degree (Hons.) from Xidian University, Xi'an, China, in 2018. He is currently pursuing the Ph.D. degree in electronic engineering with Tsinghua University, Beijing, China. His research interests include reconfigurable intelligent surface (RIS) and THz massive MIMO, with the emphasis on low-overhead channel acquisition. He has received the honorary mention at the 2019 IEEE ComSoc Student Competition.

Linglong Dai received the B.S. degree from Zhejiang University, Hangzhou, China, in 2003, the M.S. degree (Hons.) from China Academy of Telecommunications Technology, Beijing, China, in 2006, and the Ph.D. degree (Hons.) from Tsinghua University, Beijing, in 2011.

From 2011 to 2013, he was a Post-Doctoral Research Fellow with the Department of Electronic Engineering, Tsinghua University, where he was an Assistant Professor from 2013 to 2016 and has been an Associate Professor since 2016. He has coauthored the book *MmWave Massive MIMO: A Paradigm for 5G* (Academic Press, 2016). He has authored or coauthored over 60 IEEE journal articles and over 40 IEEE conference papers. He also holds 19 granted patents. His current research interests include massive MIMO, reconfigurable intelligent surface (RIS), massive MIMO, millimeter-wave/terahertz communications, and machine learning for wireless communications.

Dr. Dai has received five IEEE best paper awards at the IEEE ICC 2013, the IEEE ICC 2014, the IEEE ICC 2017, the IEEE VTC 2017-Fall, and the IEEE ICC 2018. He has also received the Tsinghua University Outstanding Ph.D. Graduate Award in 2011, the Beijing Excellent Doctoral Dissertation Award in 2012, the China National Excellent Doctoral Dissertation Nomination Award in 2013, the URSI Young Scientist Award in 2014, the IEEE TRANSACTIONS ON BROADCASTING Best Paper Award in 2015, the *Electronics Letters* Best Paper Award in 2016, the National Natural Science Foundation of China for Outstanding Young Scholars in 2017, the IEEE ComSoc Asia-Pacific Outstanding Young Researcher Award in 2017, the IEEE ComSoc Asia-Pacific Outstanding Paper Award in 2018, the China Communications Best Paper Award in 2019, the IEEE ACCESS Best Multimedia Award in 2020, and the IEEE Communications Society Leonard G. Abraham Prize in 2020. He was listed as a highly cited researcher by Clarivate in 2020. He is an Area Editor of IEEE COMMUNICATIONS LETTERS, and an Editor of IEEE TRANSACTIONS ON COMMUNICATIONS and IEEE TRANSACTIONS ON VEHICULAR TECHNOLOGY. Particularly, he is dedicated to reproducible research and has made a large amount of simulation code publicly available.