Generalized Superimposed Training for RIS-aided Cell-free Massive MIMO-OFDM Networks

Hanxiao Ge, Navneet Garg, and Tharmalingam Ratnarajah

***Abstract*—In this paper, a generalized superimposed training (GST) is proposed for an uplink cell-free multiple-input multiple-output orthogonal frequency-division multiplexing (mMIMO-OFDM) system, which is aided by reconfigurable intelligent surfaces (RISs) to enhance the spectral efficiency in the system. For the GST scheme, the pilots and data symbols are transmitted simultaneously in the coherence time. This scheme is different from traditional separate transmission methods, such as regular pilots (RP) transmission. In the OFDM multi-carrier case, a part of the subcarriers is based on the GST, whereas the other part of subcarriers is used for data transmission only. The channel and data estimations are carried out and the normalized mean-squared error (NMSE), bit error rate (BER), and sum-rate in different schemes are compared. Different receiver cooperation levels are analyzed in this case, including fully centralized processing and local processing. The distributed time processing and iterative process are also used to improve the performance of the data estimation in this system.**

***Index Terms*—Cell-free, channel estimation, massive MIMO, OFDM, RIS.**

I. Introduction

CELL-FREE massive multiple-input multiple-output (mMIMO) technology has obtained large attention in the next generation mobile networks [1], which possesses the advantages of traditional mMIMO systems or cellular systems, especially in terms of the higher spectral efficiency and coverage probability [2]. In a cell-free mMIMO system, we consider multiple access points (APs) to serve multiple users. There are one or more central processing units (CPUs) in this system, which can process transmitted signals from the APs fully or partially [3]. In [1], authors come up with different levels of receiver cooperation to judge the signals, including fully centralized and local processing. The advantages of the cell-free system compared with the cellular system are discussed in [3]. In the cell-free system, the number of users is lower than the number of APs, to reduce the pilot reuse and contamination. Performance analysis of cell-free mMIMO (or cloud radio access networks) system is given in [4] and [5]. Authors in [6] have introduced the resource allocation. Low-resolution analog to digital converters (ADCs) and zero-forcing (ZF) receivers are used in [2]. The total power minimization is carried out in the cell-free mMIMO system [7].

Reconfigurable intelligent surface (RIS) is also widely used that are equipped with a large number of reflective elements. Each element can provide the phase shifts to the signals, adjusted via the RIS controller. Because of the non-line-of-sight (nLoS) component in the actual signal transmission, the spectral and energy efficiencies are decreased. The utilization of RIS can overcome this fact, as RIS elements can reflect signals in a suitable direction. In [8] and [9], the basic properties of RIS are analyzed. RIS is also used in ambient backscatter systems [10]. Authors in [11] have introduced semi-blind multiuser detection under the RIS.

Previously, authors in [12] and [13] have investigated the integration of cell-free mMIMO systems and RIS. In their systems, pilots and data symbols are transmitted with a single RIS panel; that is very limited because the systems with multiple panels can show the performance of the RIS-aided system effectively. The works in [14] and [15] have discussed the RIS-aided system with Rayleigh fading channel which ignores that the line-of-sight (LoS), is often measured. Hence, the Rician fading channel is more suitable in this case. Authors in [16] and [17] have discussed the Rician fading channel in a cell-free system, but not in the RIS-aided system.

Authors in [18]–[21] have combined the cell-free system mMIMO system with orthogonal frequency-division multiplexing (OFDM) to support multiple subcarriers case. Due to the excessively complex signal equalisation, it is difficult to process the frequency-selective fading problem in channels in a single subcarrier system, and therefore cannot sustain a high data rate [19]. Hence, we need to use the OFDM to cope with this problem. In a cell-free mMIMO-OFDM system, frequency selectivity can be eliminated by exploiting frequency-domain orthogonal pilots [20]. However, the OFDM is hardly employed in RIS-aided systems.

Note that many previous researches are based on the regular pilots (RP) transmission. Works [22]–[24] introduced the principles of RP transmission. In this scheme, pilots and data symbols are transmitted separately, and the coherence time $T = \tau_p + \tau_d$, where τ_p is the time slot for pilots transmission and τ_d is for data transmission. However, when the number of users increases, pilot reuse will become more severe and lead to pilot contamination. In [25]–[28], authors use the superimposed training (ST) scheme to cope with this problem by reducing the probability of pilot reuse, since T pilots are available and $T > \tau_p$. The performance of ST in reducing pilot contamination is limited because the length of data symbols is fixed. The standard ST scheme is used in some OFDM-related

Manuscript received April 15, 2022 revised July 14, 2022; approved for publication by Yajun Zhao, Guest Editor, September 6, 2022.

This work was supported by the UK Engineering and Physical Sciences Research Council (EPSRC) under grant number EP/T021063/1.

Authors are with Institute for Digital Communications, The University of Edinburgh, Edinburgh, EH9 3FG, UK e-mails: {hanxiao.ge, navneet.garg, t.ratnarajah}@ed.ac.uk.

H. Ge is the corresponding author.

Digital Object Identifier: 10.23919/JCN.2022.000040.

1229-2370/22/$10.00 © 2022 KICS

works, such as [29]–[32]. Authors in [3], [33]–[37] use the generalized superimposed training (GST) scheme to analyze the performance of mMIMO or cell-free mMIMO systems. Pilots contamination can be reduced effectively by adjusting the number of data streams to get better channel and data estimation performance. However, the GST scheme has not been used in the RIS-aided cell-free mMIMO-OFDM system.

Only a few of works discuss CPU processing or fully centralized processing (cooperation level 4 in [1]), and many of them are based on no-cooperation (signals are processed at local APs). However, signals processed by the CPU can reduce the channel estimation normalized mean-squared error (NMSE), bit error rate (BER), and increase the sum-rate. Hence, we also use CPU processing in the analysis of the RIS-aided systems.

A. Contributions

In this paper, we consider a RIS-aided cell-free mMIMO-OFDM system. The GST scheme is used in this system and to carry out the channel and data estimation. The receiver cooperation has four levels [1]. We compare the receiver cooperation in level 4 (fully centralized processing or CPU processing) and level 1 (local processing). The phase shifts coefficient in the RIS panel is also important, because the optimal phase can help us lower channel estimation NMSE and improve sum-rate effectively. Hence, we need to carry out the optimization of the RIS phase coefficient. The main contributions of our works are as follows.

1) GST scheme in RIS-aided systems: We propose the GST scheme in the RIS-aided cell-free mMIMO-OFDM system and then carry out the channel and data estimation. The distributed time processing and iterative process have been used to improve the data estimation performance.

2) Comparsion of cooperation levels in RIS-aided system: We discuss the receiver cooperation levels in our work. Our results show that fully centralized processing can realize better estimation performance compared with local processing. These results show that if signals transmitted by users are fully processed by the CPU, other than the local AP, pilot reuse and contamination can be reduced effectively.

3) Simulations and comparison: We compare the performance of GST, ST, and RP schemes in channel and data estimations, show that the GST scheme can reduce the NMSE and improve the sum-rate. We compare the average NMSE when we choose the optimal RIS phase, and non-RIS (traditional cell-free) in the system. We use the OFDM (multiple subcarriers case) and show the advantages of the GST scheme in reducing pilot contamination. We also plot the mean-squared error (MSE) for data estimates versus the number of iterations and show that the iterative process can improve the data detection performance.

B. Organization

Section II describes the system model of the RIS-aided cell-free mMIMO system and the process of OFDM and the equations of the systems in the multiple subcarriers form. Section III shows the process of channel estimation and data detection in local processing, and we derive equations of the channel estimation NMSE and the signal-to-interference-plus-noise ratio (SINR). Section IV provides the process of channel and data estimation in centralized processing, followed by the analysis of the ST and RP schemes in Section V. In Section VI, we analyze the process of the time distribution and the iterative process. We give our simulation results in Section VII, and the conclusion in VIII.

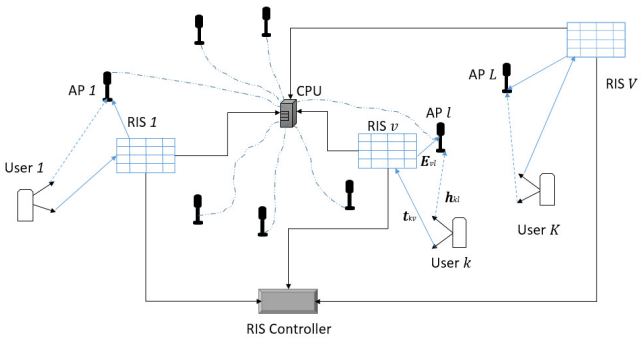

Fig. 1. Structure of the RIS-aided cell-free mMIMO system.

C. Notations

Scalars, vectors, and matrices can be represented by the lower case (a), lower case bold face (**a**), and upper case bold face (**A**). The Hermitian can be represented by $(.)^H$ and the transpose can be written as $(.)^T$. The conjugate can be given as $(.)^*$. The l_2 norm and Frobenious norm can be represented by $\|.\|$ and $\|.\|_F$ respectively. Finally, mod$(.,.)$ is the modulus operation and $\lfloor . \rfloor$ denotes truncated argument. $\mathcal{D}(\mathbf{A})$ and $\mathcal{BD}(\mathbf{A}_1, \mathbf{A}_2)$ yield a vector of diagonal entries of the argument, and the block diagonal matrix.

II. System Model

In a given geographical region, we consider a RIS-aided cell-free mMIMO-OFDM system with L APs, which serve K single antennas users, and each AP has M antennas. Let V as the number of RIS panels in this region and each panel has Q elements. For each RIS panel, the direct link connects user k and AP l. Besides, the indirect link is from user k to AP l through RIS panel v.

We use the OFDM in this system with N_s subcarriers and the coherence time is T in samples, where $T > K$. The transmitted signal are denoted by $\mathbf{X}_k = \big[\mathbf{x}_k(0), \mathbf{x}_k(1), \cdots, \mathbf{x}_k(N_s-1)\big] \in \mathbb{C}^{T\times N_s}$. The local channel matrix is $\mathbf{U}_{kl} = \mathcal{BD}\big[\mathbf{u}_{kl}(0), \mathbf{u}_{kl}(1), \cdots, \mathbf{u}_{kl}(N_s-1)\big] \in \mathbb{C}^{MN_s\times N_s}$. In order to obtain the OFDM symbol, the transmitted block signals are fed into an inverse discrete Fourier transform (IDFT). The output of the IDFT is then cycled in order to reduce intersymbol interference (ISI) and add a guard interval. After signals have been conveyed across multipath channels, the guard interval needs to be removed, and the discrete Fourier transform needs to be performed (DFT) [19], [38]. We assume that the length of the cyclic prefix (CP) is larger enough to ignore the timing and frequency offset.

TABLE I
LIST OF VARIABLES.

Variables	Desciption
$L(l)$	Number of APs
$M(m)$	Number of antennas at each AP
$K(k,j)$	Number of users
d	Number of data streams
$T(t)$	Number of coherence time slots
τ_p, τ_d	Time to transmit pilots and symbols in the RP scheme
b	Number of discarded columns from a $T \times T$ orthonormal matrix
N_s	Number of subcarriers
$F_{\max}$	Maximum number of iterations
$V(v)$	Number of RIS panels
$Q(q)$	Number of elements at RIS panels
$\mathbf{p}_k, \mathbf{S}_k, \mathbf{s}_k(n), \mathbf{Z}_k$	k-th pilot, data symbol and precoder
$\mathbf{X}_k, \mathbf{x}_k(n)$	k-th transmitted symbol
$\lambda \in (0,1)$	Power-allocation factor
$\lambda^{'}$	Wavelength
$\mathcal{Z}, \mathcal{Z}_k$	2-dimensional set and its first projection for $\mathbf{Z}_j^H \mathbf{p}_k, \forall j \neq k$
$\mathbf{u}_{kl}(n), \mathbf{u}_k(n)$	Local and centralized channel for each subcarrier
$\hat{\mathbf{u}}_{kl}(n), \hat{\mathbf{u}}_k(n)$	Local and centralized estimated channel for each subcarrier
a	Transmitted power per time slot
$\beta_{kl}, R_{kl}, \mathbf{h}_{kl}(n)$ $\overline{\mathbf{g}}_{kl}(n), \widetilde{\mathbf{g}}_{kl}(n)$	User-AP path loss, Rician factor, channel, constant and random parts
$\alpha_{kv}, \omega_{kv}, \mathbf{t}_{kv}(n)$ $\overline{\mathbf{m}}_{kv}(n), \widetilde{\mathbf{m}}_{kv}(n)$	User-RIS path loss, Rician factor, channel, constant and random parts
$\eta_{vl}, e_{vl}, \mathbf{E}_{vl}(n)$ $\overline{\mathbf{G}}_{vl}(n), \widetilde{\mathbf{G}}_{vl}(n)$	RIS-AP path loss, Rician factor, channel, constant and random parts
Φ_v, ϕ_{vq}	Diagonal matrix of phase shift ϕ_{vq}
$\mathbf{a}_M(\theta,\phi), \mathbf{a}_Q(\theta,\phi)$	Steering vectors of antenna arrays and RIS panels
$\mathbf{r}_m$	3-dimensional location of the m-th antenna
$C_{kl}^{(\mathbf{u})}(n), \mathbf{R}_{kl}^{(\mathbf{u})}(n)$, $C_k^{(\mathbf{u})}(n), \mathbf{R}_k^{(\mathbf{u})}(n)$	Variance and correlation matrices of $\mathbf{u}_{kl}(n)$ and $\mathbf{u}_k(n)$
$\mathbf{R}_{kl}^{(\Delta)}(n), \mathbf{R}_k^{(\Delta)}(n)$	Correlation matrices of $\Delta_{kl}(n)$ and $\Delta_k(n)$
$\mathbf{R}_{k,ij}^{(\mathbf{Z})}(n)$	Correlation matrix of $\mathbf{Z}_k^H \mathbf{x}_i(n) \mathbf{x}_j^H(n) \mathbf{Z}_k$
$\hat{\mathbf{S}}_{kl}, \hat{\mathbf{s}}_{kl}(n), \hat{\mathbf{S}}_k, \hat{\mathbf{s}}_k(n)$	Data estimates for local and centralized processing
$\mathbf{Y}_l(n), \mathbf{Y}(n)$	Received signal at the local AP and CPU
$\mathbf{N}_l(n), \mathbf{N}(n)$	Noise at the local AP and CPU
$\Delta_{kl}(n), \Delta_k(n)$	Channel estimation error for local and centralized processing
$\overline{\mathbf{u}}_{kl}(n), \widetilde{\mathbf{u}}_{kl}(n)$	Constant and random parts of $\mathbf{u}_{kl}(n)$
$\overline{\hat{\mathbf{u}}}_{kl}(n), \widetilde{\hat{\mathbf{u}}}_{kl}(n)$	Constant and random parts of $\hat{\mathbf{u}}_{kl}(n)$

A. Operating Processes

In this work, we discuss two levels of receiver cooperation (fully centralized processing and local processing). The operating processes of them are

1) Local processing: Users transmit signals to local APs, and APs process the signal directly, without transmitting to the CPU.

2) Fully centralized processing: APs do not process the signal at this level, and they transmit signals to the CPU. The CPU can process the collective signals from all APs.

B. Local Processing

For this cooperation level, users transmit signals to distributed local APs and local APs process signals directly, rather than transmit these signals to the central CPU.

In [1], cooperation level 1 means all signals are processed at the AP, rather than the CPU. In this case, the whole channel (including direct and indirect links) in the subcarrier n is

$$\mathbf{u}_{kl}(n) = \mathbf{h}_{kl}(n) + \sum_{v=1}^{V} \mathbf{E}_{vl}(n) \mathbf{\Phi}_v \mathbf{t}_{kv}(n), \tag{1}$$

where $\mathbf{h}_{kl}(n) \in \mathbb{C}^{M \times 1}$ is the direct channel from user k to AP l; $\mathbf{E}_{vl}(n) \in \mathbb{C}^{M \times Q}$ is the channel matrix from the v-th RIS to the l-th AP; $\mathbf{t}_{kv}(n) \in \mathbb{C}^{Q \times 1}$ is the channel vector from the k-th user to the v-th RIS; $\mathbf{\Phi}_v = \text{diag}(\phi_{v1}, \cdots, \phi_{vQ})$ is the matrix of phase shifts of the v-th RIS; and, ϕ_{vq} is the phase of the q-th element in the v-th RIS and $|\phi_{vq}| = 1, \forall v, q$ [36]. Hence, the received signal at the AP l in subcarrier n is

$$\mathbf{Y}_l(n) = \sqrt{a} \sum_{k=1}^{K} \mathbf{u}_{kl}(n) \mathbf{x}_k^H(n) + \mathbf{N}_l(n), \tag{2}$$

where $\mathbf{x}_k(n) \in \mathbb{C}^{T \times 1}$ is the transmitted signal in n-th subcarrier and $\mathbf{N}_l(n) \in \mathbb{C}^{M \times T} \sim \mathcal{CN}(\mathbf{0}, \sigma^2 \mathbf{I}_M)$ is the noise at the l-th AP, where σ is the standard deviation of the noise amplitude; a is the transmitted power per time slot.

1) Channel model for user to AP (direct link): The multpath channel of direct link from user k to AP l can be given in terms of the Rician fading as

$$\mathbf{h}_{kl}(n) = \sqrt{\beta_{kl}} \frac{\sqrt{R_{kl}} \overline{\mathbf{g}}_{kl}(n) + \widetilde{\mathbf{g}}_{kl}(n)}{\sqrt{R_{kl} + 1}}, \tag{3}$$

where β_{kl} are the large-scale fading coefficient for channel $\mathbf{h}_{kl}(n)$ in n-th subcarrier; R_{kl} being the Rician factor; if the Rician fading increases, the channel will become more deterministic. $\overline{\mathbf{g}}_{kl}(n) = \mathbf{a}_M(\theta_{g,kl}, \phi_{g,kl}) \in \mathbb{C}^{M \times 1}$ is the LoS part; $(\theta_{g,kl}, \phi_{g,kl})$ are azimuth and elevation angles at the l-th AP; $\widetilde{\mathbf{g}}_{kl}(n) \sim \mathcal{CN}(\mathbf{0}, \mathbf{I}_M)$ is the nLoS part.

The steering vector of an M-element antennas planar array is [39]

$$\mathbf{a}_M(\theta, \phi) = \left[e^{-j \mathbf{r}_1^T \mathbf{k}(\theta,\phi)}, \cdots, e^{-j \mathbf{r}_M^T \mathbf{k}(\theta,\phi)} \right]^T, \tag{4}$$

where the angles $\theta, \phi \in \left[-\frac{\pi}{2}, \frac{\pi}{2} \right]$; $\mathbf{k}(\theta,\phi) = \frac{2\pi}{\lambda^{'}} \left[\cos(\theta)\cos(\phi), \cos(\theta)\sin(\phi), \sin(\theta) \right]^T$; $\lambda^{'}$ is the wavelength. The vector $\mathbf{r}_m$ (for each $m = 1, \cdots, M$) describes the 3-dimensional location of the m-th antenna element as

$$\mathbf{r}_m = \left[\text{mod}(m-1, M_H) d_H^{'},\ 0,\ \lfloor (m-1)/M_H \rfloor d_V^{'} \right]^T, \tag{5}$$

where M_H and M_V are the number of antenna elements in horizontal and vertical directions; $d_H^{'}$ and $d_V^{'}$ are the inter-element distances in both directions.

2) Channel model for user to RIS: The channel from user k to RIS v is

$$\mathbf{t}_{kv}(n) = \sqrt{\alpha_{kv}} \frac{\sqrt{\omega_{kv}} \overline{\mathbf{m}}_{kv}(n) + \widetilde{\mathbf{m}}_{kv}(n)}{\sqrt{\omega_{kv} + 1}}, \tag{6}$$

where α_{kv} is the large scale-fading coefficient; w_{kv} is the Rician factor; $\overline{\mathbf{m}}_{kv}(n) = \mathbf{a}_Q(\theta_{u,kv}, \phi_{u,kv}) \in \mathbb{C}^{Q \times 1}$ is the LoS component and $\widetilde{\mathbf{m}}_{kv}(n)$ is the nLoS component.

3) Channel from RIS to AP: The channel matrix from RIS panel to AP in n-th subcarrier can be expressed as

$$\mathbf{E}_{vl}(n) = \sqrt{\eta_{vl}}\frac{\sqrt{e_{vl}}\overline{\mathbf{G}}_{vl}(n) + \widetilde{\mathbf{G}}_{vl}(n)}{\sqrt{e_{vl}+1}}, \tag{7}$$

where η_{vl} is the large-scale fading coefficient; e_{vl} is the Rician factor; $\overline{\mathbf{G}}_{vl}(n) = [\mathbf{a}_M(\theta_{F,vl},\phi_{F,vl})\mathbf{a}_Q^H(\theta_{E,vl},\phi_{E,vl})] \in \mathbb{C}^{M\times Q}$ is the LoS and $\widetilde{\mathbf{G}}_{vl}(n)$ is the nLoS component.

C. CPU Processing

For this cooperation level, after distributed local APs receive signals from users, then APs transmit signals to the CPU. The CPU processes collective signals from local APs.

In level 4 cooperation, signals are processed at the CPU, other than the local APs. The collective received signal from all APs can be expressed as

$$\mathbf{Y}(n) = [\mathbf{Y}_1^T(n),\cdots,\mathbf{Y}_L^T(n)]^T = \sqrt{a}\sum_{k=1}^{K}\mathbf{u}_k(n)\mathbf{x}_k^H(n)+\mathbf{N}(n),$$

where $n = 0,\cdots,N_s-1$; the centralized noise is $\mathbf{N}(n) = [\mathbf{N}_1^T(n),\cdots,\mathbf{N}_L^T(n)]^T \in \mathbb{C}^{LM\times T}$; the centralized channel $\mathbf{u}_k(n) = [\mathbf{u}_{k1}^T(n),\cdots,\mathbf{u}_{kL}^T(n)]^T \in \mathbb{C}^{LM\times 1}$.

In the OFDM transmission, the received signal at the CPU is $\mathbf{Y} = \left[\mathbf{Y}^T(0),\mathbf{Y}^T(1),\cdots,\mathbf{Y}^T(N_s-1)\right]^T \in \mathbb{C}^{LMN_s\times T}$, and the noise $\mathbf{N} = \left[\mathbf{N}^T(0),\mathbf{N}^T(1),\cdots,\mathbf{N}^T(N_s-1)\right]^T \in \mathbb{C}^{LMN_s\times T}$ [21].

D. GST Symbols

At the k-th user, the transmitted signal for both channel estimation and data estimation in n-th subcarrier is

$$\mathbf{x}_k(n) = \begin{cases} \sqrt{\lambda}\mathbf{p}_k + \sqrt{1-\lambda}\mathbf{Z}_k\mathbf{s}_k(n), & \forall n\in\xi \\ \mathbf{Z}_k\mathbf{s}_k(n), & \text{otherwise,} \end{cases} \tag{8}$$

where ξ is the index set of subcarriers on which both data symbols and pilots are transmitted; $\lambda \in (0,1)$ is the pilot power-allocation factor. $\mathbf{p}_k$ is the $T\times 1$ pilot vector; $\mathbf{Z}_k$ is the $T\times d$ orthogonal precoder matrix; b is the number of discarded columns from a $T\times T$ orthonormal matrix, where $d = T-b$; $\mathbf{s}_k(n)$ denote the $d\times 1$ data symbol vector in n-th subcarrier. The power of the transmitted GST symbols is $\mathbb{E}\{\|\mathbf{x}_k(n)\|_2^2\} = T$. $\mathbf{Z}_k^H\mathbf{p}_k = \mathbf{0}$ and $\mathbf{p}_k^H\mathbf{p}_k = T$, because $\mathbf{Z}_k$ and $\mathbf{p}_k$ are from the same orthogonal matrix. As for $\mathbf{Z}_j^H\mathbf{p}_k$, it can be written as

$$\mathbf{Z}_j^H\mathbf{p}_k = \begin{cases} T\mathbf{e}_{(k,j)}, & (k,j)\in\overline{\mathcal{Z}} \\ \mathbf{0}, & \text{otherwise,} \end{cases} \tag{9}$$

where $\mathcal{Z} = \{1,\cdots,K\}^2$; $\overline{\mathcal{Z}} = \{(k,j) : \mathbf{Z}_k \neq \mathbf{p}_j, \forall j = 1,\cdots,T-b\}$. $\mathbf{e}_{k,j}$ is a vector of zeros and ones, and the notation (k,j) denotes the location of those ones [33]. If $k\neq j$, we can get the precoder matrices product as $\mathbf{Z}_k^H\mathbf{Z}_j = T\mathbf{F}_{[k,j]}$, where $\mathbf{F}_r$ is the matrix whose rank is s and it has s ones and all the other elements are zero. The rank $r_{k,j}$ of the matrix $\mathbf{F}_{[k,j]}$ represents the number of matching columns of $\mathbf{Z}_k$ and $\mathbf{Z}_j$. For $k=j$, $\mathbf{F}_{[k,j]} = \mathbf{I}_{T-b}$, and $\mathbf{Z}_k^H\mathbf{p}_k = \mathbf{0}$, and $\mathbb{E}\{\mathbf{s}_k\mathbf{s}_k^H\} = \frac{1}{d}\mathbf{I}_d$.

Algorithm 1 Local sum-rate calculation algorithm

```
Optimize the Φ_v.
for n = 0, ···, N_s − 1 do
for k = 1, ···, K do
    Get Y_l(n) for each subcarrier.
    Compute and update channel estimation û_kl(n).
    Update data estimation ŝ_kl(n).
    Update P_kl,s(n), P_kl,SI(n), P_kl,CI(n), P_kl,N(n).
    Calculate SINR ρ_kl(n) and sum-rate R_kl(n).
  end for
  end for
```

III. Signal Processing at Local APs

To carry out the useful channel estimation, the number of equations (MT) must meet the following condition, that is,

$$MK + Kd \leq MT, \tag{10}$$

where MK and Kd means the number of channel and data variables. If this condition do not meet, there will lead to severe pilot contamination in the system [34]. For example, if $T=6$ and $K=3$, $d \leq M(\frac{T}{K}-1) = M$; for $M=K$, we have $d \leq T-K$.

A. Channel Estimation

To measure the pilot contamination, we need to carry out the channel estimation and get the NMSE.

1) Subcarriers in the set ξ: In n-th subcarrier, the estimated channel can be derived by using the least-squares (LS) criterion as

$$\begin{aligned}\hat{\mathbf{u}}_{kl}(n) =& \frac{\mathbf{Y}_l(n)\mathbf{p}_k}{T\sqrt{a\lambda}} \\ =& \mathbf{u}_{kl}(n) + \Delta_{kl}(n), \forall n\in\xi \\ \overset{(a)}{=}& \mathbf{u}_{kl}(n) + \sum_{j\in\mathbb{C}_k\setminus\{k\}}\mathbf{u}_{jl}(n) \\ &+ \sqrt{\frac{1-\lambda}{\lambda}}\sum_{j\in\mathcal{Z}_k}\mathbf{u}_{jl}(n)\mathbf{s}_j^H(n)\mathbf{e}_{k,j} + \frac{\mathbf{N}_l(n)\mathbf{p}_k}{T\sqrt{a\lambda}} \\ \overset{(b)}{=}& \mathbf{u}_{kl}(n) + \Delta_{kl}(n),\end{aligned} \tag{11}$$

where $\Delta_{kl}(n)$ is the local channel estimation error for each subcarrier; $\mathbb{C}_k\setminus\{k\}$ is the sets of users who use the same pilots; $\mathcal{Z}_k = \{j | \|\mathbf{Z}_j^H\mathbf{p}_k\| = T, \forall j\neq k\}$. The last three terms of (a) includes the pilot reuse and the noise. Hence, if these terms part increases, the channel estimation error will become large.

2) Subcarriers not in the set ξ: If the subcarrier is not in the set ξ, we could not use the LS method directly because there is no pilots in the subcarriers. We use the spline interpolation to get the estimated channel.

3) NMSE calculation: For $n \in \xi$, the mean of the error in each subcarrier $\overline{\Delta}_{kl}(n) = \mathbb{E}\{\Delta_{kl}(n)\} = \sum_{j \in \mathtt{C}_k \setminus \{k\}} \overline{\mathbf{u}}_{jl}(n)$, where $\overline{\mathbf{u}}_{jl}(n)$ is the LoS part of the channel $\mathbf{u}_{jl}(n)$. The channel estimation mean-squared error (MSE) is

$$\begin{aligned}\mathbb{E}\{\Delta_{kl}^H(n)\Delta_{kl}(n)\} =& \sum_{j,i \in \mathtt{C}_k \setminus \{k\}} \mathbb{E}\{\mathbf{u}_{il}^H(n)\mathbf{u}_{jl}(n)\} \\ &+ \frac{1-\lambda}{\lambda d} \times \sum_{j \in \mathcal{Z}_k} \sum_{i \in \mathcal{Z}_k} \mathbb{E}\{\mathbf{u}_{il}^H(n)\mathbf{u}_{jl}(n)\} \\ &+ \frac{M\sigma^2}{Ta\lambda} \\ =& \sum_{j,i \in \mathtt{C}_k \setminus \{k\}} \overline{\mathbf{u}}_{il}^H(n)\overline{\mathbf{u}}_{jl}(n) \\ &+ \frac{1-\lambda}{\lambda d} \sum_{j \in \mathcal{Z}_k} \overline{\mathbf{u}}_{jl}^H(n)\overline{\mathbf{u}}_{jl}(n) \\ &+ \sum_{j \in \mathtt{C}_k \setminus \{k\}} C_{jl}^{(\mathbf{u})}(n) \\ &+ \frac{1-\lambda}{\lambda d} \sum_{j \in \mathcal{Z}_k} C_{jl}^{(\mathbf{u})}(n) + \frac{M\sigma^2}{Ta\lambda},\end{aligned} \tag{12}$$

where $C_{jl}^{(\mathbf{u})}(n)$ is the variance of the $\mathbf{u}_{jl}(n)$, which is

$$\begin{aligned}C_{jl}^{(\mathbf{u})}(n) =& \mathbb{E}\{[\mathbf{u}_{jl}(n) - \overline{\mathbf{u}}_{jl}(n)]^H[\mathbf{u}_{jl}(n) - \overline{\mathbf{u}}_{jl}(n)]\} \\ =& \frac{M\beta_{jl}}{R_{jl}+1} + \sum_{v=1}^{V} \frac{M\eta_{vl}\alpha_{jv}Q}{e_{vl}+1} \\ &+ \sum_{v=1}^{V} \frac{\eta_{vl}e_{vl}}{e_{vl}+1} \frac{\alpha_{jv}}{\omega_{jv}+1} tr[\overline{\mathbf{G}}_{vl}^H(n)\overline{\mathbf{G}}_{vl}(n)].\end{aligned} \tag{13}$$

The trace of the correlation matrix of $\mathbf{u}_{jl}(n)$ is

$$\begin{aligned}tr[\mathbf{R}_{jl}^{(\mathbf{u})}(n)] =& \mathbb{E}\{\mathbf{u}_{jl}^H(n)\mathbf{u}_{jl}(n)\} \\ =& \overline{\mathbf{u}}_{jl}^H(n)\overline{\mathbf{u}}_{jl}(n) + C_{jl}^{(\mathbf{u})}(n).\end{aligned} \tag{14}$$

The NMSE is used to describe the performance of the channel estimation (for all subcarriers) as $NMSE_{kl} = \sum_n \frac{\mathbb{E}\{\Delta_{kl}^H(n)\Delta_{kl}(n)\}}{\mathbb{E}\{\mathbf{u}_{kl}^H(n)\mathbf{u}_{kl}(n)\}}, \forall n$.

B. Phase Shifts Optimization

We need to choose the optimal RIS phase. RIS panels can reflect beams to different directions, We assume that $\overline{\mathbf{Q}}(n) = \overline{\mathbf{G}}_{vl}^H(n)\overline{\mathbf{G}}_{vl}(n)$ and $\overline{\mathbf{K}}(n) = \overline{\mathbf{m}}_{kv}(n)\overline{\mathbf{m}}_{kv}^H(n)$ ($\overline{\mathbf{G}}$ and $\overline{\mathbf{m}}$ are the LoS parts of $\mathbf{E}_{vl}(n)$ and $\mathbf{t}_{kv}(n)$ and carry out RIS values optimization by minimizing the MSE.

Lemma 1: The MSE minimization problem solution is

$$\begin{aligned}&\underset{|\phi_{vq}|=1, \forall v,q}{\arg\min}\ tr(\mathbb{E}\{\Delta_{kl}(n)\Delta_{kl}^H(n)\}) \\ &= \frac{\mathcal{V}_{\min}\left[\overline{\mathbf{Q}}(n) \odot \overline{\mathbf{K}}^T(n)\right]}{|\mathcal{V}_{\min}\left[\overline{\mathbf{Q}}(n) \odot \overline{\mathbf{K}}^T(n)\right]|}, \forall v,\end{aligned} \tag{15}$$

where $\mathcal{V}_{\min}$ is the eigenvector of $\overline{\mathbf{Q}}(n) \odot \overline{\mathbf{K}}^T(n)$. The derivation process is given in the Appendix-A.

C. Data Estimation

After the channel estimation, we carry out the data estimation by using the LS criterion and to analyze the data errors and calculate the rate.

1) Subcarriers in the set ξ: We use the LS criterion to calculate the local data estimates as

$$\begin{aligned}\hat{\mathbf{s}}_{kl}(n) =& \arg\min_{\mathbf{s}_k(n)} \|\mathbf{Y}_l(n) - \sqrt{a(1-\lambda)}\hat{\mathbf{u}}_{kl}(n)\mathbf{s}_k^H(n)\mathbf{Z}_k^H\|_F^2 \\ \overset{(a)}{=}& \mathbf{Z}_k^H\mathbf{Y}_l^H(n)\frac{\hat{\mathbf{u}}_{kl}(n)}{\|\hat{\mathbf{u}}_{kl}(n)\|_2^2}\frac{1}{T\sqrt{a(1-\lambda)}} &\text{(16a)} \\ \overset{(b)}{=}& \Big[T\sqrt{a(1-\lambda)}\mathbf{s}_k(n)\mathbf{u}_{kl}^H(n) + \sqrt{a}\sum_{i\neq k}\mathbf{Z}_k^H\mathbf{x}_i(n)\mathbf{u}_{il}^H(n) \\ &+ \mathbf{Z}_k^H\mathbf{N}_l^H(n)\Big]\frac{\hat{\mathbf{u}}_{kl}(n)}{\|\hat{\mathbf{u}}_{kl}(n)\|_2^2}\frac{1}{T\sqrt{a(1-\lambda)}} &\text{(16b)} \\ \overset{(c)}{=}& \mathbf{s}_k(n) + \underbrace{\mathbf{s}_k(n)\left(\frac{\mathbf{u}_{kl}^H(n)\hat{\mathbf{u}}_{kl}(n)}{\|\hat{\mathbf{u}}_{kl}(n)\|_2^2} - 1\right)}_{\text{self-interference}} \\ &+ \underbrace{\sum_{i\neq k}\frac{\mathbf{Z}_k^H\mathbf{x}_i(n)}{T\sqrt{1-\lambda}}\frac{\mathbf{u}_{il}^H(n)\hat{\mathbf{u}}_{kl}(n)}{\|\hat{\mathbf{u}}_{kl}(n)\|_2^2}}_{\text{cross-interference}} &\text{(16c)} \\ &+ \frac{\mathbf{Z}_k^H\mathbf{N}_l^H(n)\hat{\mathbf{u}}_{kl}(n)}{T\sqrt{a(1-\lambda)}\|\hat{\mathbf{u}}_{kl}(n)\|_2^2}. &\text{(16d)}\end{aligned}$$

The above equation includes the actual data symbol and the noise and the interference component. We could write power equations of the transmitted signal, self-interference (SI), cross-interference (CI) and the local noise by multiplying $\|\hat{\mathbf{u}}_{kl}(n)\|_2^2$ in each term as

$$P_{kl,s}(n) = \mathbb{E}\Big\|\|\hat{\mathbf{u}}_{kl}(n)\|_2^2\mathbf{s}_k(n)\Big\|_2^2 = \mathbb{E}\|\hat{\mathbf{u}}_{kl}(n)\|_2^4 \cdot \mathbb{E}\|\mathbf{s}_k(n)\|_2^2, \tag{17}$$

$$P_{kl,SI}(n) = \mathbb{E}\Big\|\mathbf{s}_k(n)[\mathbf{u}_{kl}^H(n)\hat{\mathbf{u}}_{kl}(n) - \hat{\mathbf{u}}_{kl}^H(n)\hat{\mathbf{u}}_{kl}(n)]\Big\|_2^2, \tag{18}$$

$$P_{kl,CI}(n) = \frac{1}{1-\lambda}\mathbb{E}\Big\|\sum_{i\neq k}\frac{\mathbf{Z}_k^H\mathbf{x}_i(n)}{T}\mathbf{u}_{il}^H(n)\hat{\mathbf{u}}_{kl}(n)\Big\|_2^2, \tag{19}$$

$$P_{kl,N}(n) = \mathbb{E}\Big\|\frac{\mathbf{Z}_k^H\mathbf{N}_l^H(n)\hat{\mathbf{u}}_{kl}(n)}{T\sqrt{a(1-\lambda)}}\Big\|_2^2. \tag{20}$$

2) Subcarriers not in the set ξ: Since $\mathbf{x}_k(n) = \mathbf{Z}_k\mathbf{s}_k(n)$, the estimated data can be given as

$$\begin{aligned}\hat{\mathbf{s}}_{kl}(n) =& \mathbf{s}_k(n) + \underbrace{\mathbf{s}_k(n)\left(\frac{\mathbf{u}_{kl}^H(n)\hat{\mathbf{u}}_{kl}(n)}{\|\hat{\mathbf{u}}_{kl}(n)\|_2^2} - 1\right)}_{\text{self-interference}} \\ &+ \underbrace{\sum_{i\neq k}\frac{\mathbf{Z}_k^H\mathbf{x}_i(n)}{T}\frac{\mathbf{u}_{il}^H(n)\hat{\mathbf{u}}_{kl}(n)}{\|\hat{\mathbf{u}}_{kl}(n)\|_2^2}}_{\text{cross-interference}} \\ &+ \frac{\mathbf{Z}_k^H\mathbf{N}_l^H(n)\hat{\mathbf{u}}_{kl}(n)}{T\sqrt{a}\|\hat{\mathbf{u}}_{kl}(n)\|_2^2}.\end{aligned} \tag{21}$$

The power of SI and the signal remains unchanged, and the power equations of CI and the noise can be updated as

$$P_{kl,CI}(n) = \mathbb{E}\Big\| \sum_{i\neq k} \frac{\mathbf{Z}_k^H \mathbf{x}_i(n)}{T} \mathbf{u}_{il}^H(n)\hat{\mathbf{u}}_{kl}(n)\Big\|_2^2, \tag{22}$$

$$P_{kl,N}(n) = \mathbb{E}\Big\| \frac{\mathbf{Z}_k^H \mathbf{N}_l^H(n)\hat{\mathbf{u}}_{kl}(n)}{T\sqrt{a}}\Big\|_2^2. \tag{23}$$

3) Local sum-rate calculation: The local SINR can be expressed as

$$SINR_{kl}(n) = \frac{P_{kl,s}(n)}{P_{kl,SI}(n) + P_{kl,CI}(n) + P_{kl,N}(n)}. \tag{24}$$

We can use the SINR equation to get the local sum-rate $\mathcal{R}_l(n) = \sum_k \log_2[1 + SINR_{kl}(n)]$. The whole process of calculating the sum-rate is like Algorithm 1. The sum-rate for all subcarriers can be written as $\mathcal{R}_l = \sum_n \mathcal{R}_l(n)$.

IV. Signal Processing at the CPU

For the valid channel and data estimation at the CPU, the system need meet the essential condition as

$$MLT \geq KM + Kd, \tag{25}$$

where MLT is the number of observations and the KM and Kd are the number of channel and data estimates of K users. For example, for $T = 10$ and $K = 5$, $d \leq \frac{MLT}{K} - M = M(2L-1)$. Fully centralized processing can increase the allowable number of data streams per user in the system by a factor of $\frac{\frac{LT}{K}-1}{\frac{T}{K}-1} \approx L$. If $ML = K$, $d \leq T - \frac{K}{L}$.

A. Channel Estimation

If all the signals are processed at the CPU, the centralized estimated channel from all APs in each subcarrier can be written as $\hat{\mathbf{u}}_k(n) = [\hat{\mathbf{u}}_{k1}^T(n), \cdots, \hat{\mathbf{u}}_{kL}^T(n)]^T \in \mathbb{C}^{LM\times 1}$. The channel estimation mean-squared error (MSE) can be calculated as $\mathbb{E}\{\Delta_k^H(n)\Delta_k(n)\} = \sum_l \mathbb{E}\{\Delta_{kl}^H(n)\Delta_{kl}(n)\}$. The NMSE at the CPU can be given as $NMSE_k = \sum_n \frac{\sum_l \mathbb{E}\{\Delta_{kl}^H(n)\Delta_{kl}(n)\}}{\sum_l \mathbb{E}\{\mathbf{u}_{kl}^H(n)\mathbf{u}_{kl}(n)\}}, \forall n$.

B. Data Estimation

Like the local processing, data estimation can be carried out by using the LS criterion in the centralized processing level after the channel estimation.

1) Subcarriers in the set ξ: Towards the data detection in fully centralized processing, the data estimates via LS can be given as

$$\begin{aligned}\hat{\mathbf{s}}_k(n) =& \mathbf{s}_k(n) + \underbrace{\mathbf{s}_k(n)\left(\frac{\mathbf{u}_k^H(n)\hat{\mathbf{u}}_k(n)}{\|\hat{\mathbf{u}}_k(n)\|_2^2} - 1\right)}_{\text{self-interference}} \\ &+ \underbrace{\sum_{i\neq k} \frac{\mathbf{Z}_k^H \mathbf{x}_i(n)}{T\sqrt{1-\lambda}} \frac{\mathbf{u}_i^H(n)\hat{\mathbf{u}}_k(n)}{\|\hat{\mathbf{u}}_k(n)\|_2^2}}_{\text{cross-interference}} \\ &+ \frac{\mathbf{Z}_k^H \mathbf{N}^H(n)\hat{\mathbf{u}}_k(n)}{T\sqrt{a(1-\lambda)}\|\hat{\mathbf{u}}_k(n)\|_2^2}.\end{aligned} \tag{26}$$

Theorem 1: Like the previous method, we could also write power equations in each term in centralized processing as

$$P_{k,s}(n) = \mathbb{E}\Big\| \|\hat{\mathbf{u}}_k(n)\|_2^2 \mathbf{s}_k(n)\Big\|_2^2 = \mathbb{E}\|\hat{\mathbf{u}}_k(n)\|_2^4 \cdot \mathbb{E}\|\mathbf{s}_k(n)\|_2^2, \tag{27}$$

$$P_{k,SI}(n) = \mathbb{E}\Big\| \mathbf{s}_k(n)[\mathbf{u}_k^H(n)\hat{\mathbf{u}}_k(n) - \hat{\mathbf{u}}_k^H(n)\hat{\mathbf{u}}_k(n)]\Big\|_2^2, \tag{28}$$

$$P_{k,CI}(n) = \frac{1}{1-\lambda}\mathbb{E}\Big\| \sum_{i\neq k} \frac{\mathbf{Z}_k^H \mathbf{x}_i(n)}{T} \mathbf{u}_i^H(n)\hat{\mathbf{u}}_k(n)\Big\|_2^2, \tag{29}$$

$$P_{k,N}(n) = \mathbb{E}\Big\| \frac{\mathbf{Z}_k^H \mathbf{N}^H(n)\hat{\mathbf{u}}_k(n)}{T\sqrt{a(1-\lambda)}}\Big\|_2^2. \tag{30}$$

2) Subcarriers not in the set ξ: As $\mathbf{x}(n) = \mathbf{Z}_k \mathbf{s}_k(n)$, the estimated data can be rewritten as

$$\begin{aligned}\hat{\mathbf{s}}_k(n) =& \mathbf{s}_k(n) + \underbrace{\mathbf{s}_k(n)\left(\frac{\mathbf{u}_k^H(n)\hat{\mathbf{u}}_k(n)}{\|\hat{\mathbf{u}}_k(n)\|_2^2} - 1\right)}_{\text{self-interference}} \\ &+ \underbrace{\sum_{i\neq k} \frac{\mathbf{Z}_k^H \mathbf{x}_i(n)}{T} \frac{\mathbf{u}_i^H(n)\hat{\mathbf{u}}_k(n)}{\|\hat{\mathbf{u}}_k(n)\|_2^2}}_{\text{cross-interference}} \\ &+ \frac{\mathbf{Z}_k^H \mathbf{N}^H(n)\hat{\mathbf{u}}_k(n)}{T\sqrt{a}\|\hat{\mathbf{u}}_k(n)\|_2^2}.\end{aligned} \tag{31}$$

As the signal and the self-interference keep remained, we can only change the equation of the cross-interference and noise as

$$P_{k,CI}(n) = \mathbb{E}\Big\| \sum_{i\neq k} \frac{\mathbf{Z}_k^H \mathbf{x}_i(n)}{T} \mathbf{u}_i^H(n)\hat{\mathbf{u}}_k(n)\Big\|_2^2, \tag{32}$$

$$P_{k,N}(n) = \mathbb{E}\Big\| \frac{\mathbf{Z}_k^H \mathbf{N}^H(n)\hat{\mathbf{u}}_k(n)}{T\sqrt{a}}\Big\|_2^2. \tag{33}$$

3) Sum-rate calculation: For all subcarrers ($n \in \xi$ or $n \notin \xi$), the centralized SINR can be given as

$$SINR_k(n) = \frac{P_{k,s}(n)}{P_{k,SI}(n) + P_{k,CI}(n) + P_{k,N}(n)}. \tag{34}$$

We can use the SINR equation to get the local sum-rate $\mathcal{R}_k(n) = \sum_k \log_2[1 + SINR_k(n)]$, the sum-rate for all subcarriers can be given as $\mathcal{R} = \sum_n \mathcal{R}_k(n)$.

As for the centralized processing, we can write the collective estimated data symbol in terms of the local case as

$$\hat{\mathbf{s}}_k(n) = \sum_l \frac{\hat{\mathbf{u}}_{kl}^H(n)\hat{\mathbf{u}}_{kl}(n)}{\sum_l \hat{\mathbf{u}}_{kl}^H(n)\hat{\mathbf{u}}_{kl}(n)} \hat{\mathbf{s}}_{kl}(n). \tag{35}$$

We write the whole estimated data symbols matrix for all subcarriers as $\hat{\mathbf{S}}_k = [\hat{\mathbf{s}}_k(0), \hat{\mathbf{s}}_k(1), \cdots, \hat{\mathbf{s}}_k(N_s-1)] \in \mathbb{C}^{d\times N_s}$.

V. Comparision of ST and RP Schemes

In this section, we will introduce the standard ST and RP schemes, and derive equations of transmitted symbols $\mathbf{x}_k(n)$ and estimated data symbols $\hat{\mathbf{s}}_k(n)$ for comparison.

A. ST Scheme

In the standard ST scheme, we also transmit data symbols and the pilots at the same time [40]. The differences between this scheme and the GST scheme is that the ST scheme does not have a variable data length, and the precoder $\mathbf{F}_k$ is a $T\times T$ matrix. The number of data streams $d = T$. The transmitted signal can be expressed as

$$\mathbf{x}_k(n) = \begin{cases} \sqrt{\lambda}\mathbf{p}_k + \sqrt{1-\lambda}\mathbf{F}_k\mathbf{s}_k(n), & \forall n \in \xi \\ \mathbf{F}_k\mathbf{s}_k(n), & \text{otherwise,} \end{cases} \tag{36}$$

where $\mathbb{E}\{\mathbf{s}_k(n)\mathbf{s}_k^H(n)\} = \frac{1}{T}\mathbf{I}_T$. For the ST scheme, we could write the channel estimation error equation when $n \in \xi$ as

$$\Delta_{kl}(n) = \sum_{j\in\mathbb{C}_k\setminus\{k\}} \mathbf{u}_{jl}(n) + \sqrt{\frac{1-\lambda}{\lambda}}\sum_{j\neq k}\mathbf{u}_{jl}(n)\mathbf{s}_j^H(n)\mathbf{F}_k\mathbf{p}_k + \frac{\mathbf{N}_l(n)\mathbf{p}_k}{T\sqrt{a\lambda}}. \tag{37}$$

For local processing, we need to meet the essential condition as $MT \geq KM+KT$, and for centralized processing, we need meet $MLT \geq KM + KT$. For $n \in \xi$, the centralized data estimate can be written as

$$\hat{\mathbf{s}}_k(n) = \frac{\mathbf{Y}^H(n)\hat{\mathbf{u}}_k(n)}{\|\hat{\mathbf{u}}_k(n)\|_2^2}\frac{1}{T\sqrt{a(1-\lambda)}}. \tag{38}$$

Same as the GST scheme, we can rewrite the estimated data when $n \notin \xi$ as

$$\hat{\mathbf{s}}_k(n) = \frac{\mathbf{Y}^H(n)\hat{\mathbf{u}}_k(n)}{\|\hat{\mathbf{u}}_k(n)\|_2^2}\frac{1}{T\sqrt{a}}. \tag{39}$$

B. RP Scheme

For the RP scheme, we transmit the pilots and the data symbols separately. We assume that the coherence time $T = \tau_p + \tau_d$, where τ_p and τ_d are the time to transmit the pilots and the data symbols [34] and [3]. We need to set $\tau_p > K$ to prevent the pilot reuse. We can write the transmitted signal as

$$\mathbf{x}_k(n) = \begin{cases} \mathbf{p}_k\sqrt{\frac{\lambda T}{\tau_p}}, & \forall n \in \xi \\ \sqrt{T}\mathbf{V}_k\mathbf{s}_k(n), & \text{otherwise,} \end{cases} \tag{40}$$

where $\mathbf{V}_k$ is a $\tau_p \times d$ precoding matrix; $\|\mathbf{V}_k\|_F^2 = d$ and $\mathbb{E}\{\mathbf{s}_k(n)\mathbf{s}_k^H(n)\} = \frac{1}{d}\mathbf{I}_d$; $\mathbf{p}_k$ is a $\tau_p\times 1$ vector in the RP scheme.

When $n \in \xi$, there is no data symbol transmitted in the system. For $n \notin \xi$, we could write the estimated data for the RP scheme as

$$\hat{\mathbf{s}}_k(n) = \mathbf{V}_k^H\mathbf{Y}^H(n)\frac{\hat{\mathbf{u}}_k(n)}{\hat{\mathbf{u}}_k^H(n)\hat{\mathbf{u}}_k(n)}\frac{1}{\sqrt{aT}}. \tag{41}$$

C. Comparison of GST with Other Schemes

These schemes can be compared in terms of sum-rate and NMSE in local and centralized processing. The training time for the RP scheme is τ_p and for the GST and ST scheme is T. Because of $T > \tau_p$, the pilot contamination effect can be limited and the NMSE can be reduced.

The number of data streams for the ST scheme is T, and for the GST and RP schemes is d. Hence, the probability of generating the data errors in the ST scheme is larger than in the other schemes because of $T \geq K + d$ in local processing and $T \geq \frac{K}{L} + d$ in centralized processing, which leads the decrease of the sum-rate in the ST scheme.

D. Comparison with Other Cooperation Levels

1) Level 3: local processing and large-scale fading decoding: The data estimates of the centralized processing can be regarded as the linear summation of the local processing. Hence, data estimates can be calculated locally and then sent to the CPU. We do not need to transmit channel estimates from APs to the CPU and then carry out the data estimation [1] and [35].

2) Level 2: local processing and centralized decoding: The requirements of cooperation should be relaxed by averaging the linear combing. In this cooperation level, each AP transmit the value $\hat{\mathbf{s}}_{kl}(n) \times \|\hat{\mathbf{u}}_{kl}(n)\|^2/M$, rather than $\hat{\mathbf{s}}_{kl}(n)$, which aggregates at the CPU as

$$\frac{1}{M}\sum_{l=1}^{L}\hat{\mathbf{s}}_{kl}(n)\|\hat{\mathbf{u}}_{kl}(n)\|^2 = \hat{\mathbf{s}}_k(n) \times \frac{\|\hat{\mathbf{u}}_{kl}(n)\|^2}{M}. \tag{42}$$

VI. Distributed Time Processing and Iterations

Distributed time processing is used to reduce computing complexity at the conclusion of the coherence time block, allowing the received symbols to be evaluated as they come in and the final detection to be made once all of the symbols have been received. We also employ an iterative technique to boost the effectiveness of data estimate.

A. Distributed Time Processing

In this GST scheme, for the received signal in centralized processing $\mathbf{Y}$ spans for T slots, let the received signal $\mathbf{Y}(n) = [\mathbf{y}(n,1), \cdots, \mathbf{y}(n,T)]$. For $n \in \xi$, the channel estimation equation at the end of the time T is [34]

$$\hat{\mathbf{u}}_k(n,T) = \frac{\mathbf{Y}(n)\mathbf{p}_k}{T\sqrt{a\lambda}} = \frac{1}{T\sqrt{a\lambda}}\sum_{i=1}^{T}\mathbf{y}(n,i)p_k(i), \tag{43}$$

where $p_k(i) \in \mathbb{C}$ is the i-th entry of the $T \times 1$ pilot vector. For distributed time processing at the CPU, we can update the channel estimate as

$$\begin{aligned}\hat{\mathbf{u}}_k(n,t+1) &= \frac{1}{T\sqrt{a\lambda}}\sum_{i=1}^{t+1}\mathbf{y}(n,i)p_k(i) \\ &= \frac{\sum_{i=1}^{t}\mathbf{y}(n,i)p_k(i) + \mathbf{y}(n,t+1)p_k(t+1)}{T\sqrt{a\lambda}} \\ &= \hat{\mathbf{u}}_k(n,t) + \frac{\mathbf{y}(n,t+1)p_k(t+1)}{T\sqrt{a\lambda}}.\end{aligned} \tag{44}$$

For the data estimation, we could also write the similar equation as

Algorithm 2 Algorithm of distributed time processing and iterative processing

1: Initialize the estimated channel $\hat{\mathbf{u}}_k(n,0) = 0$, and estimated symbols $\hat{\mathbf{s}}_k(n,0) = 0, \forall k$.
2: **for** $t = 1, \cdots, T$
3: Receive observations $\mathbf{y}(n,t)$ from the CPU.
4: **for** iter$= 1, \cdots, F_{\max}$
5: Compute and update $\widetilde{\mathbf{y}}_k(n,t)$.
6: Compute and update the channel estimate $\hat{\mathbf{u}}_k(n,t)$, for $n \notin \xi$, use the spline interpolation.
7: Update the data estimate $\hat{\mathbf{s}}_k(n,t)$.
8: **end for**
9: **end for**

$$\hat{\mathbf{s}}_k(n,T) = \mathbf{Z}_k^H \mathbf{Y}^H(n) \frac{\hat{\mathbf{u}}_k(n,T)}{\|\hat{\mathbf{u}}_k(n,T)\|_2^2} \frac{1}{T\sqrt{a(1-\lambda)}} = \sum_{i=1}^{T} \mathbf{z}_k(i)\mathbf{y}^H(n,i) \frac{\hat{\mathbf{u}}_k(n,T)}{\|\hat{\mathbf{u}}_k(n,T)\|_2^2} \frac{1}{T\sqrt{a(1-\lambda)}}, \tag{45}$$

where $\mathbf{Z}_k^H = [\mathbf{z}_k^H(1), \cdots \mathbf{z}_k^H(T)]$. For $n \notin \xi$ case, the channel estimation can be carried out by the spline interpolation.

B. Iterative Estimation Process

From the above equations, we find that the data estimation performance depends on the channel estimation. We use an iterative channel and data estimation to decrease the data estimation MSE. For $n \in \xi$, the channel estimates can be updated as

$$\hat{\mathbf{u}}_k(n) \leftarrow \frac{\widetilde{\mathbf{Y}}_k(n)\mathbf{p}_k}{T\sqrt{a\lambda}}, \tag{46}$$

where the matrix $\widetilde{\mathbf{Y}}_k(n) = \mathbf{Y}(n) - \sum_{i \neq k} \hat{\mathbf{u}}_i(n)\hat{\mathbf{x}}_i^H(n) = [\widetilde{\mathbf{y}}_k(n,1), \cdots, \widetilde{\mathbf{y}}_k(n,T)]$. The update at the slot t can be expressed as

$$\hat{\mathbf{u}}_k(n,t) \leftarrow \hat{\mathbf{u}}_k(n,t-1) + \frac{\widetilde{\mathbf{y}}_k(n,t)p_k(t)}{T\sqrt{a\lambda}}, \tag{47}$$

where $\widetilde{\mathbf{y}}_k(n,t) = \mathbf{y}(n,t) - \sum_{i \neq k} \hat{\mathbf{u}}_i(n,t)\hat{x}_i^*(n,t)$, and $\hat{x}_i(n,t)$ is

$$\hat{x}_i(n,t) = \begin{cases} \sqrt{\lambda}p_k(t) + \sqrt{1-\lambda}\mathbf{z}_k(t)\hat{\mathbf{s}}_k(n), & \forall n \in \xi \\ \mathbf{z}_k(t)\hat{\mathbf{s}}_k(n), & \text{otherwise.} \end{cases} \tag{48}$$

The estimated data symbol is

$$\hat{\mathbf{s}}_k(n,T) \leftarrow \mathbf{Z}_k^H \widetilde{\mathbf{Y}}^H(n) \frac{\hat{\mathbf{u}}_k(n,T)}{\|\hat{\mathbf{u}}_k(n,T)\|_2^2} \frac{1}{T\sqrt{a(1-\lambda)}}. \tag{49}$$

The procedures of the distributed time processing and iterative estimation are given in Algorithm 2, where $F_{\max}$ be the maximum number of iterations.

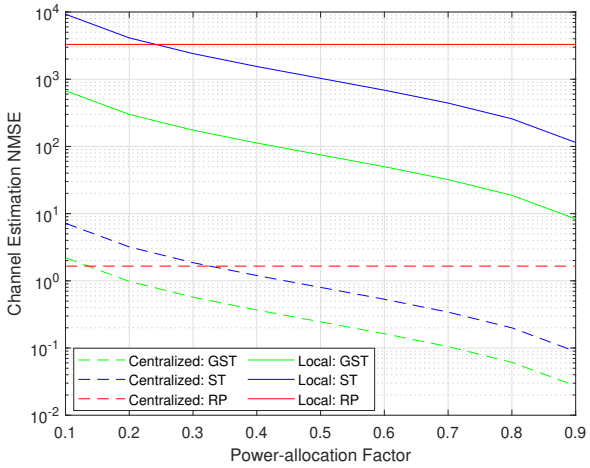

Fig. 2. Channel estimation NMSE versus λ in different training schemes with the optimal RIS phase ($T = 7, K = 5, d = 2$, SNR $= 5$ dB, $\tau_p = 2$, and $\kappa = 2$).

VII. Simulation Result

We consider $L = 100$ APs which has $M = 4$ antennas in form of 2×2 planar array. There are $V = 2$ RIS panels and each panel has $Q = 100$ elements in form of 10×10 planar array. We use the OFDM modulation and the number of subcarrier is $N_s = 64$. The large-scale fading coefficient is $\frac{10^{-U/10}}{\mu^{\delta}}$, where μ is the distance between user, RIS and APs in a 2-dimensional area with uniform distribution. δ is the path loss exponent; Our chosen channels are based on the Rician fading and Rician factor is $R = 10$ for each channel; We choose $U = 26$ dB, $\delta = 2.2$ for the direct links, and $U = 28$ dB, $\delta = 3.67$ for the indirect links. When we plot the figures of NMSE, BER and sum-rate, ξ is the index set only with odd numbers between 0 and $N_s - 1$. Meanwhile, $\kappa = 2$ is the pilots' interval.

A. Channel Estimation NMSE

In Fig. 2, we plot the NMSE in the RIS-aided cell-free mMIMO-OFDM system, both in CPU processing and local processing with the optimal RIS phase. We assume that the noise is Gaussian distributed and $\sigma^2 = 1$.

We compare the GST, ST, and RP cases in channel estimation performance, both in centralized processing and local processing. We find that GST can reduce the channel estimation NMSE effectively (almost 50% lower than the ST scheme and 98.3% at most than the RP scheme). This concludes that the GST scheme can improve the channel estimation performance obviously. The value of NMSE is possibly larger than one because of the pilot reuse.

Besides, we compare the effects of RIS panels in the channel estimation. We choose RIS-aided systems with the optimal phase and the system without RIS (the traditional cell-free system), to see the average NMSE in these cases like Fig. 3. The results show that the RIS-aided system with the optimal phase can reduce the NMSE effectively. The NMSE increases when the number of users increases, which means that the

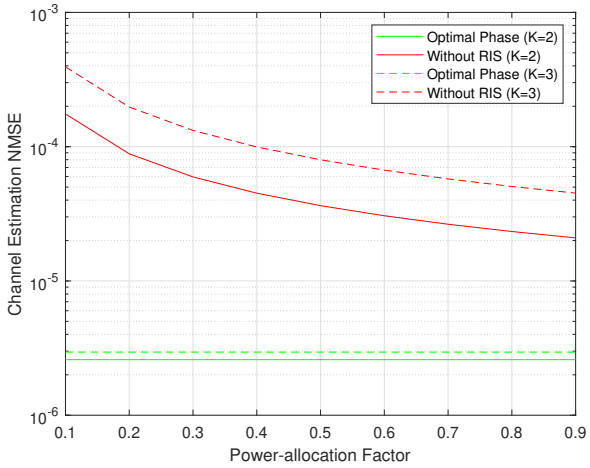

Fig. 3. Average channel estimation NMSE $\frac{1}{KL}\sum_{kl} NMSE_{kl}$ versus λ for different phase shifts and no-RIS case with GST ($T = 7, d = 2$, SNR $= 5$ dB, and $\kappa = 2$).

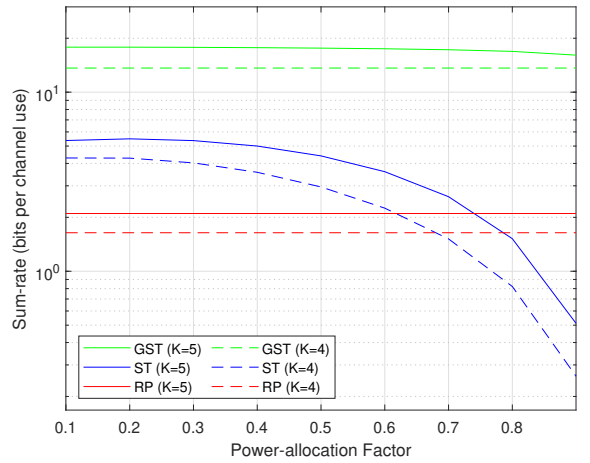

Fig. 4. For centralized processing, sum-rate versus λ for different training schemes with the optimal RIS phase ($T = 7, d = 2$, SNR $= 5$ dB, $\tau_p = 2$, and $\kappa = 2$).

pilot reuse and contamination will become severe if more users transmit signals in the system.

For the channel estimation NMSE in different cooperation levels, we conclude that centralized processing can reduce the pilot reuse and improve the channel estimation performance better than local processing at APs.

B. Sum-rate

Sum-rate is also a good parameter to measure the data estimation. We need to get the power of self-interference, cross-interference, transmitted signal, and noise.

Fig. 4 shows the concave behavior of the sum-rate with the optimal RIS phase. For the GST in centralized processing, we get the highest of the three sum-rate, which means that the GST scheme is the best scheme to improve the sum-rate in the system. The ST scheme drops dramatically when the λ increases, because data streams for the ST scheme are $d = T = 7$, for other schemes, $d = 2$. Hence, the ST scheme is easy for increasing the data transmission error.

When we increase the number of users, the sum-rate also increases. However, from Fig. 3, the NMSE and pilot contamination increase when the number of users increases. Hence, a suitable number of users should be included in the system.

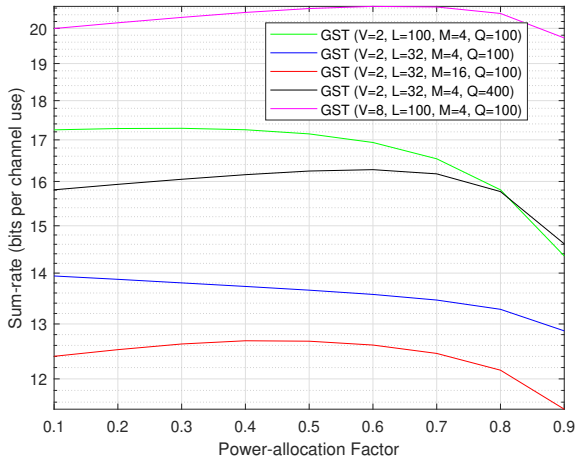

Fig. 5. For centralized processing, sum-rate versus λ for comparing different parameters ($T = 7, d = 2$, SNR $= 5$ dB, $\tau_p = 2$, and $\kappa = 2$).

From Fig. 5, it is clear that when the number of RIS panels and their elements increases, the sum-rate can be improved. When more APs are equipped in the system, the sum-rate increases, but we need to limit the number of antennas in each AP. The reason is that when we equip more antennas at a close distance, it leads to high interference and reduces the sum-rate in the system.

C. Bit Error Rate

Bit error rate is also an essential factor in data estimation. If the value of BER is low, the performance of data estimation is better.

In Fig. 6, we compare the value of BER for centralized processing, and we choose the optimal phase shift. RIS-aided system panels can reduce the BER compared with the conventional cell-free system. Our results show that when we choose the GST scheme in the centralized processing, the value of the BER is reduced significantly. In this case, the curve of the BER is convex, and we can get the lowest BER when we choose a suitable value of λ. Meanwhile, we can find that the value of the BER becomes the largest when we choose the ST scheme. Because of $d = T$, the largest data stream number leads to more bit errors.

D. Iteration Simulations

We carry out the time distributed processing and the iterative process like Algorithm 2. In Fig. 7, we plot the MSE of data estimates versus the number of iterations with the optimal RIS phase for all subcarriers $n \in \xi$. The results show that when the number of iterations increases, the MSE of data estimates decreases. When we choose the larger value of $F_{\max}$, we can

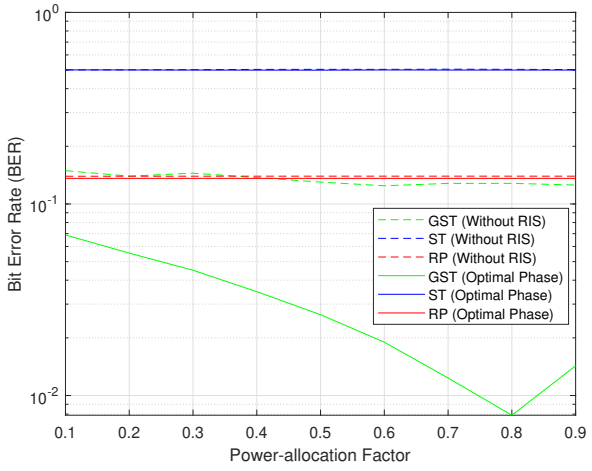

Fig. 6. For centralized processing, BER versus λ for different number of users with the optimal RIS phase ($T = 7, K = 5, d = 2$, SNR $= 5$ dB, and $\kappa = 2$)

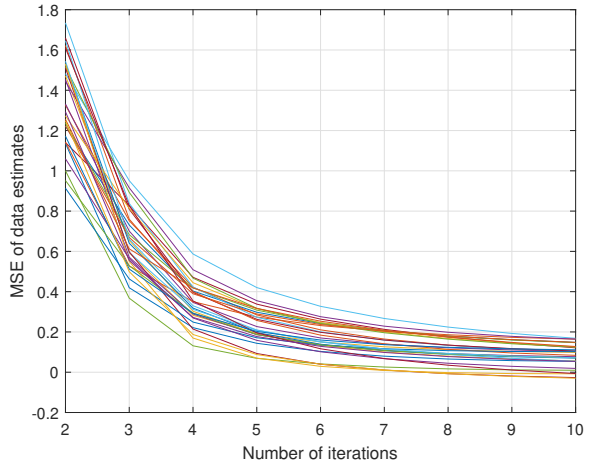

Fig. 7. For centralized processing, MSE of data estimates ($\mathbb{E}\{[\mathbf{s}_k(n,T) - \hat{\mathbf{s}}_k(n,T)]^H[\mathbf{s}_k(n,T) - \hat{\mathbf{s}}_k(n,T)]\}$) versus $F_{\max}$ with the optimal RIS phase for subcarriers $n \in \xi$ ($T = 7, K = 5, d = 2$, SNR $= 5$ dB, $\kappa = 2$, and $\lambda = 0.5$).

get the MSE convergence, and the value of the MSE of data estimates is converged.

VIII. Conclusion

In this paper, we have proposed a generalized superimposed training scheme and have used it in the RIS-aided cell-free mMIMO-OFDM system. We have analyzed the NMSE, BER, and sum-rate in multiple subcarriers cases. We have also compared this training scheme with the standard superimposed scheme and regular pilot scheme both in local and centralized scenarios. Our results have shown that a RIS-aided system with an optimal phase in centralized processing could get the best channel and data estimation performance and reduce pilot contamination. We have confirmed that the iterative process can reduce the MSE of data estimates when the number of iterations increases.

Our future work is to consider the analysis with stochastic geometry and Poisson point process (PPP) model, because the APs and users are uniformly distributed in this work and consider the effects of buildings and other blockages in the system.

Appendix

A. Proof of the Optimal Phase Shift Coefficients

As (15), the MSE minimization problem can be solved by optimizing the phase shifts of RIS panels as

$$\begin{aligned}
&\underset{|\phi_{vq}|=1,\forall v,q}{\arg\min}\ tr(\mathbb{E}\{\Delta_{kl}(n)\Delta_{kl}^H(n)\}) \\
&\overset{(a)}{=} \underset{|\phi_{vq}|=1,\forall v,q}{\arg\min} \sum_{v=1}^{V} \mathrm{tr}[\mathbf{\Phi}_v^H \overline{\mathbf{Q}}(n) \mathbf{\Phi}_v \overline{\mathbf{K}}(n)] \\
&\overset{(b)}{=} \underset{|\phi_{vq}|=1,\forall v,q}{\arg\min}\ \mathbf{v}^H[\overline{\mathbf{Q}}(n) \odot \overline{\mathbf{K}}^T(n)]\mathbf{v}, \forall v \\
&\overset{(c)}{=} \frac{\mathcal{V}_{\min}\left[\overline{\mathbf{Q}}(n) \odot \overline{\mathbf{K}}^T(n)\right]}{\left|\mathcal{V}_{\min}\left[\overline{\mathbf{Q}}(n) \odot \overline{\mathbf{K}}^T(n)\right]\right|}, \forall v,
\end{aligned} \tag{50}$$

where the vector $\mathbf{v} = [\phi_{11}, \cdots, \phi_{VQ}]^T$. From (a) to (b), the trace identity $\mathrm{tr}(\mathcal{D}(\mathbf{A})^H \mathbf{B} \mathcal{D}(\mathbf{C}) \mathbf{D}) = \mathbf{A}^H(\mathbf{B} \odot \mathbf{D}^T)\mathbf{C}$ is used. In (c), the division is element-wise division of vectors.

B. Local Processing for GST ($n \in \xi$)

For the local power of each part, the equations from (17) to (20) can be further derived as

1) Power of the signal:

$$\begin{aligned}
P_{kl,s}(n) =& \mathbb{E}\|\hat{\mathbf{u}}_{kl}(n)\|_2^4 \cdot \mathbb{E}\|\mathbf{s}_k(n)\|_2^2 = \mathbb{E}\|\hat{\mathbf{u}}_{kl}(n)\|_2^4, \\
=& \mathbb{E}|\overline{\hat{\mathbf{u}}}_{kl}^H(n)\overline{\hat{\mathbf{u}}}_{kl}(n)|^2 + \mathbb{E}|\widetilde{\hat{\mathbf{u}}}_{kl}^H(n)\overline{\hat{\mathbf{u}}}_{kl}(n)|^2 \\
&+ \mathbb{E}|\overline{\hat{\mathbf{u}}}_{kl}^H(n)\widetilde{\hat{\mathbf{u}}}_{kl}(n)|^2 + \mathbb{E}|\widetilde{\hat{\mathbf{u}}}_{kl}^H(n)\widetilde{\hat{\mathbf{u}}}_{kl}(n)|^2 \\
&+ 2\mathbb{E}\overline{\hat{\mathbf{u}}}_{kl}^H(n)\overline{\hat{\mathbf{u}}}_{kl}(n)\widetilde{\hat{\mathbf{u}}}_{kl}^H(n)\widetilde{\hat{\mathbf{u}}}_{kl}(n) \\
=& |\overline{\hat{\mathbf{u}}}_{kl}^H(n)\overline{\hat{\mathbf{u}}}_{kl}(n)|^2 + 2\overline{\hat{\mathbf{u}}}_{kl}^H(n)\mathbf{C}_{kl}^{(\hat{\mathbf{u}})}(n)\overline{\hat{\mathbf{u}}}_{kl}^H(n) \\
&+ tr[\mathbf{C}_{kl}^{(\hat{\mathbf{u}})}(n)]^2 + 2\overline{\hat{\mathbf{u}}}_{kl}^H(n)\overline{\hat{\mathbf{u}}}_{kl}(n) tr[\mathbf{C}_{kl}^{(\hat{\mathbf{u}})}(n)] \\
&+ tr\{[\mathbf{C}_{kl}^{(\hat{\mathbf{u}})}(n)]^2\} \\
=& \{\overline{\hat{\mathbf{u}}}_{kl}^H(n)\overline{\hat{\mathbf{u}}}_{kl}(n) + tr[\mathbf{C}_{kl}^{(\hat{\mathbf{u}})}(n)]\}^2 \\
&+ 2\overline{\hat{\mathbf{u}}}_{kl}^H(n)\mathbf{C}_{kl}^{(\hat{\mathbf{u}})}(n)\overline{\hat{\mathbf{u}}}_{kl}^H(n) + tr\{[\mathbf{C}_{kl}^{(\hat{\mathbf{u}})}(n)]^2\} \\
=& \{\overline{\hat{\mathbf{u}}}_{kl}^H(n)\overline{\hat{\mathbf{u}}}_{kl}(n) + tr[\mathbf{C}_{kl}^{(\hat{\mathbf{u}})}(n)]\}^2 \\
&+ 2tr\mathbf{C}_{kl}^{(\hat{\mathbf{u}})}(n)[\overline{\hat{\mathbf{u}}}_{kl}(n)\overline{\hat{\mathbf{u}}}_{kl}^H(n) \\
&+ \mathbf{C}_{kl}^{(\hat{\mathbf{u}})}(n)] - tr\{[\mathbf{C}_{kl}^{|\hat{\mathbf{u}}|}(n)]^2\} \\
=& tr[\mathbf{R}_{kl}^{|\hat{\mathbf{u}}|}(n)]^2 + 2tr[\mathbf{C}_{kl}^{(\hat{\mathbf{u}})}(n)\mathbf{R}_{kl}^{(\hat{\mathbf{u}})}(n)] \\
&- tr\{[\mathbf{C}_{kl}^{(\hat{\mathbf{u}})}(n)]^2\},
\end{aligned} \tag{51}$$

where $\overline{\hat{\mathbf{u}}}_{kl}(n) = \overline{\mathbf{u}}_{kl}(n) + \overline{\Delta}_{kl}(n)$; the correlation matrix $\mathbf{R}_{kl}^{(\hat{\mathbf{u}})}(n) = \mathbb{E}\{\hat{\mathbf{u}}_{kl}(n)\hat{\mathbf{u}}_{kl}^H(n)\} = \overline{\hat{\mathbf{u}}}_{kl}(n)\overline{\hat{\mathbf{u}}}_{kl}^H(n) + \mathbf{C}_{kl}^{(\hat{\mathbf{u}})}(n)$ is the correlation matrix of $\overline{\hat{\mathbf{u}}}_{kl}^H(n)$; the variance matrix of $\overline{\hat{\mathbf{u}}}_{kl}^H(n)$ can be given as $\mathbf{C}_{kl}^{(\hat{\mathbf{u}})}(n) = \mathbb{E}\{[\hat{\mathbf{u}}_{kl}(n) - \overline{\hat{\mathbf{u}}}_{kl}(n)][\hat{\mathbf{u}}_{kl}(n) - \overline{\hat{\mathbf{u}}}_{kl}(n)]^H\} = \mathbb{E}\{|\widetilde{\hat{\mathbf{u}}}_{kl}(n)\widetilde{\hat{\mathbf{u}}}_{kl}^H(n)|^2\}$.

2) Power of the self-interference:

$$\begin{aligned}P_{kl,SI}(n) =&\mathbb{E}\Big\|\mathbf{s}_k(n)[\mathbf{u}_{kl}^H(n)\hat{\mathbf{u}}_{kl}(n)-\hat{\mathbf{u}}_{kl}^H(n)\hat{\mathbf{u}}_{kl}(n)]\Big\|_2^2\\ =&\mathbb{E}\Big\|\mathbf{s}_k(n)\Delta_{kl}^H(n)\hat{\mathbf{u}}_{kl}(n)\Big\|\\ =&\mathbb{E}\Big|\Delta_{kl}^H(n)\mathbf{u}_{kl}(n)+\Delta_{kl}^H(n)\Delta_{kl}(n)\Big|\\ =&\mathbb{E}\Big|\Delta_{kl}^H(n)\mathbf{u}_{kl}(n)\Big|^2\\ &+2\mathbb{R}\Delta_{kl}^H(n)\mathbf{u}_{kl}(n)\Delta_{kl}^H(n)\Delta_{kl}(n)\\ &+\mathbb{E}\Big|\Delta_{kl}^H(n)\Delta_{kl}(n)\Big|^2\\ =&\mathbb{E}\Big|\Delta_{kl}^H(n)\mathbf{u}_{kl}(n)\Big|^2\\ &+2\mathbb{R}\Delta_{kl}^H(n)\mathbf{u}_{kl}(n)\Delta_{kl}^H(n)\Delta_{kl}(n)\\ &+tr[\mathbf{R}_{kl}^{(\Delta)}(n)]^2+2tr[\mathbf{C}_{kl}^{(\Delta)}(n)\mathbf{R}_{kl}^{(\Delta)}(n)]\\ &-tr\{[\mathbf{C}_{kl}^{(\Delta)}(n)]^2\},\end{aligned}\tag{52}$$

where $\mathbf{C}_{kl}^{(\Delta)}(n)=\mathbb{E}\{(\Delta_{kl}-\overline{\Delta}_{kl})(\Delta_{kl}-\overline{\Delta}_{kl})^H\}=(\frac{\beta_{jl}}{R_{jl}+1}+\sum_{v=1}^{V}\frac{\eta_{vl}\alpha_{jv}Q}{e_{vl}+1})\mathbf{I}_M+\sum_{v=1}^{V}\frac{\eta_{vl}e_{vl}}{e_{vl}+1}\frac{\alpha_{jv}}{w_{jv}+1}\overline{\mathbf{G}}_{vl}\overline{\mathbf{G}}_{vl}^H$; $\mathbf{R}_{kl}^{(\Delta)}(n)=\mathbb{E}\{\Delta_{kl}(n)\Delta_{kl}^H(n)\}$

3) Power of the cross-interference:

$$\begin{aligned}&P_{kl,CI}(n)\\ &=\frac{1}{T^2(1-\lambda)}\mathbb{E}\Big\|\sum_{i\neq k}\frac{\mathbf{Z}_k^H\mathbf{x}_i(n)}{T}\mathbf{u}_{il}^H(n)\hat{\mathbf{u}}_{kl}(n)\Big\|_2^2\\ &=\frac{1}{T^2(1-\lambda)}\sum_{i\neq k}\sum_{j\neq k}tr[\mathbb{E}\mathbf{Z}_k^H\mathbf{x}_i(n)\mathbf{x}_j^H(n)\mathbf{Z}_k]\mathbb{E}\{\hat{\mathbf{u}}_{kl}^H(n)\\ &\quad\mathbf{u}_{jl}(n)\mathbf{u}_{il}^H(n)\hat{\mathbf{u}}_{kl}(n)\}\\ &=\frac{1}{T^2(1-\lambda)}\sum_{i\neq k}\sum_{j\neq k}tr\mathbf{R}_{k,ij}^{(\mathbf{Z})}(n)\times tr[\mathbf{R}_{jl,il}^{(\mathbf{u})}(n)\mathbf{R}_{kl}^{(\hat{\mathbf{u}})}(n)],\end{aligned}\tag{53}$$

where $tr\mathbf{R}_{k,ij}^{(\mathbf{Z})}(n)=a\lambda\mathbf{e}_{ki}\mathbf{e}_{kj}^T+\frac{a(1-\lambda)}{d}\mathbf{F}_{[k,i]}\mathbf{F}_{[k,j]}^T$ is the correlation matrix of $\mathbf{Z}_k^H\mathbf{x}_i(n)\mathbf{x}_j^H(n)\mathbf{Z}_k$.

4) Power of noise:

$$\begin{aligned}&P_{kl,N}(n)\\ &=\mathbb{E}\Big\|\frac{\mathbf{Z}_k^H\mathbf{N}_l^H(n)\hat{\mathbf{u}}_{kl}(n)}{T\sqrt{a(1-\lambda)}}\Big\|_2^2\\ &=\frac{1}{T^2a(1-\lambda)}\mathbb{E}\Big\|\mathbf{Z}_k^H\mathbf{N}_l^H(n)\Big[\dot{\Delta}_{kl}(n)+\frac{\mathbf{N}_l(n)\mathbf{p}_k}{T\sqrt{a\lambda}}\Big]\Big\|^2\\ &=\mathbb{E}\Big\|\frac{\mathbf{Z}_k^H\mathbf{N}_l^H(n)\dot{\Delta}_{kl}(n)}{T^2a(1-\lambda)}\Big\|+\frac{\mathbb{E}\Big\|\mathbf{Z}_k^H\mathbf{N}_l^H(n)\mathbf{N}_l(n)\mathbf{p}_k\Big\|}{T^4a^2\lambda(1-\lambda)}\\ &=\frac{dT\sigma^2tr(\mathbf{R}_{kl,kl}^{[\dot{\Delta}]})}{T^2a(1-\lambda)}+\frac{\sigma^4MdT(T+M+1)}{T^4a^2\lambda(1-\lambda)}\\ &=\frac{dT\sigma^2}{T^2a(1-\lambda)}\Big[dT\sigma^2tr[\mathbf{R}_{kl,kl}^{(\dot{\Delta})}]+\frac{\sigma^4MTd(T+M+1)}{T^2a\lambda}\Big],\end{aligned}\tag{54}$$

where the local noise matrix can be written as $\mathbf{N}_l(n)=[\overline{\mathbf{n}}_1(n),\cdots,\overline{\mathbf{n}}_T(n)]$ and the $[\mathbf{N}_l^H(n)\mathbf{N}_l(n)]_{i,j}=\overline{\mathbf{n}}_i^H(n)\overline{\mathbf{n}}_j(n)$; $\dot{\Delta}_{kl}(n)$ is the first three terms of $(11a)$; the process to calculate the $\mathbb{E}\Big\|\mathbf{Z}_k^H\mathbf{N}_l^H(n)\mathbf{N}_l(n)\mathbf{p}_k\Big\|$ can be given as

$$\begin{aligned}&\mathbb{E}\Big\|\mathbf{Z}_k^H\mathbf{N}_l^H(n)\mathbf{N}_l(n)\mathbf{p}_k\Big\|\\ &=\sum_{i,j,k,l}\mathbb{E}\Big[\overline{\mathbf{n}}_i^H(n)\overline{\mathbf{n}}_j(n)[\mathbf{Z}_k\mathbf{Z}_k^H]_{j,k}\overline{\mathbf{n}}_k^H(n)\overline{\mathbf{n}}_l(n)[\mathbf{p}_k\mathbf{p}_k^H]_{l,i}\Big]\\ &=\sum_{i=j,k=l,l\neq k}\mathbb{E}\Big[\overline{\mathbf{n}}_i^H(n)\overline{\mathbf{n}}_i(n)[\mathbf{Z}_k\mathbf{Z}_k^H]_{i,l}\overline{\mathbf{n}}_l^H(n)\overline{\mathbf{n}}_l(n)[\mathbf{p}_k\mathbf{p}_k^H]_{l,i}\Big]\\ &\quad+\sum_{j=k,i=l,l\neq k}\mathbb{E}\Big[\overline{\mathbf{n}}_i^H(n)\overline{\mathbf{n}}_j(n)\overline{\mathbf{n}}_j^H(n)\overline{\mathbf{n}}_i(n)[\mathbf{Z}_k\mathbf{Z}_k^H]_{j,j}[\mathbf{p}_k\mathbf{p}_k^H]_{i,i}\Big]\\ &\quad+\sum_{j=k=i=l}\mathbb{E}\Big[\overline{\mathbf{n}}_i^H(n)\overline{\mathbf{n}}_i(n)\overline{\mathbf{n}}_i^H(n)\overline{\mathbf{n}}_i(n)[\mathbf{Z}_k\mathbf{Z}_k^H]_{i,i}[\mathbf{p}_k\mathbf{p}_k^H]_{i,i}\Big]\\ &=\sigma^4M^2tr(\mathbf{Z}_k\mathbf{Z}_k^H\mathbf{p}_k\mathbf{p}_k^H)+\sigma^4Mtr[\mathbf{Z}_k\mathbf{Z}_k^H]tr(\mathbf{p}_k\mathbf{p}_k^H)\\ &\quad+\sigma^4(M^2+M)tr[\mathbf{Z}_k\mathbf{Z}_k^H\mathcal{D}(\mathbf{p}_k\mathbf{p}_k^H)]\\ &=\sigma^4M\cdot Td\cdot T+\sigma^4M(M^2+M)\cdot Td\\ &=\sigma^4MTd(T+M+1).\end{aligned}\tag{55}$$

C. Centralized Processing for GST ($n\in\xi$)

For centralized processing, power equations from (27) to (30) can be further derived as

1) Power of the signal:

$$\begin{aligned}&P_{k,s}(n)\\ &=\mathbb{E}\|\hat{\mathbf{u}}_k(n)\|_2^4\cdot\mathbb{E}\|\mathbf{s}_k(n)\|_2^2=\mathbb{E}\|\hat{\mathbf{u}}_k(n)\|_2^4,\\ &=tr[\mathbf{R}_k^{(\hat{\mathbf{u}})}(n)]^2+2tr[\mathbf{C}_k^{(\hat{\mathbf{u}})}(n)\mathbf{R}_k^{|\hat{\mathbf{u}}|}(n)]-tr\{[\mathbf{C}_k^{(\hat{\mathbf{u}})}(n)]^2\}.\end{aligned}\tag{56}$$

2) Power of the self-interference:

$$\begin{aligned}P_{k,SI}(n) =&\mathbb{E}\Big\|\mathbf{s}_k(n)[\mathbf{u}_k^H(n)\hat{\mathbf{u}}_k(n)-\hat{\mathbf{u}}_k^H(n)\hat{\mathbf{u}}_k(n)]\Big\|_2^2\\ =&\mathbb{E}\Big|\Delta_k^H(n)\mathbf{u}_k(n)\Big|^2\\ &+2\mathbb{R}\Delta_k^H(n)\mathbf{u}_k(n)\Delta_k^H(n)\Delta_k(n)+tr[\mathbf{R}_k^{(\Delta)}(n)]^2\\ &+2tr[\mathbf{C}_k^{(\Delta)}(n)\mathbf{R}_k^{(\Delta)}(n)]-tr\{[\mathbf{C}_k^{(\Delta)}(n)]^2\}.\end{aligned}\tag{57}$$

3) Power of the cross-interference:

$$\begin{aligned}&P_{k,CI}(n)\\ &=\frac{1}{T^2(1-\lambda)}\mathbb{E}\Big\|\sum_{i\neq k}\frac{\mathbf{Z}_k^H\mathbf{x}_i(n)}{T}\mathbf{u}_i^H(n)\hat{\mathbf{u}}_k(n)\Big\|_2^2\\ &=\frac{1}{T^2(1-\lambda)}\sum_{i\neq k}\sum_{j\neq k}tr\mathbf{R}_{k,ij}^{(\mathbf{Z})}(n)\times tr[\mathbf{R}_{j,i}^{(\mathbf{u})}(n)\mathbf{R}_k^{(\hat{\mathbf{u}})}(n)].\end{aligned}\tag{58}$$

4) Power of the noise:

$$\begin{aligned}P_{k,N}(n) =&\mathbb{E}\Big\|\frac{\mathbf{Z}_k^H\mathbf{N}^H(n)\hat{\mathbf{u}}_k(n)}{T\sqrt{a(1-\lambda)}}\Big\|_2^2\\ =&\frac{dT\sigma^2}{T^2a(1-\lambda)}\Big[dT\sigma^2tr(\mathbf{R}_{k,k}^{(\dot{\Delta})})\\ &+\frac{\sigma^4MLTd(T+ML+1)}{T^2a\lambda}\Big]\end{aligned}\tag{59}$$

D. Local Processing for GST ($n \notin \xi$)

According to the equation (22), the power of the CI can be rewritten as

$$\begin{aligned} &P_{kl,CI}(n) \\ &= \frac{1}{T^2}\mathbb{E}\Big\| \sum_{i\neq k} \frac{\mathbf{Z}_k^H \mathbf{x}_i(n)}{T} \mathbf{u}_{il}^H(n)\hat{\mathbf{u}}_{kl}(n)\Big\|_2^2 \\ &= \frac{1}{T^2}\sum_{i\neq k}\sum_{j\neq k} tr\mathbf{R}_{k,ij}^{(\mathbf{Z})}(n) \times tr[\mathbf{R}_{jl,il}^{(\mathbf{u})}(n)\mathbf{R}_{kl}^{|\hat{\mathbf{u}}|}(n)]. \end{aligned} \tag{60}$$

For the power of the noise, we use the spline interpolation process to get the $\hat{\mathbf{u}}_{kl}(n)$ and then carry out the calculation.

E. Centralized Processing for GST ($n \notin \xi$)

As the equation (29), for subcarriers not in ξ, the power of the CI can be calculated as

$$\begin{aligned} &P_{k,CI}(n) \\ &= \frac{1}{T^2}\mathbb{E}\Big\| \sum_{i\neq k} \frac{\mathbf{Z}_k^H \mathbf{x}_i(n)}{T} \mathbf{u}_{i}^H(n)\hat{\mathbf{u}}_{k}(n)\Big\|_2^2 \\ &= \frac{1}{T^2}\sum_{i\neq k}\sum_{j\neq k} tr\mathbf{R}_{k,ij}^{(\mathbf{Z})}(n) \times tr[\mathbf{R}_{j,i}^{(\mathbf{u})}(n)\mathbf{R}_{k}^{(\hat{\mathbf{u}})}(n)]. \end{aligned} \tag{61}$$

REFERENCES

[1] E. Björnson and L. Sanguinetti, "Making cell-free massive mimo competitive with mmse process. and centralized implementation," *IEEE Trans. Wireless Commun.*, vol. 19, no. 1, pp. 77–90, 2020.

[2] X. Bai, M. Zhou, X. Qiao, Y. Zhang, and L. Yang, "Uplink performance analysis of cell-free mimo with low-resolution adcs and zf receiver," in *Proc. IEEE ICACT*, Feb 2020, pp. 125–129.

[3] H. Q. Ngo, A. Ashikhmin, H. Yang, E. G. Larsson, and T. L. Marzetta, "Cell-free massive mimo versus small cells," *IEEE Trans. Wireless Commun.*, vol. 16, pp. 1834–1850, Mar. 2017.

[4] F. A. Khan, H. He, J. Xue, and T. Ratnarajah, "Performance analysis of cloud radio access networks with distributed multiple antenna remote radio heads," *IEEE Trans. Signal Process.*, vol. 63, pp. 4784–4799, Sep. 2015.

[5] H. He, J. Xue, T. Ratnarajah, F. A. Khan, and C. B. Constantinos, "Modeling and analysis of cloud radio access networks using matern hard-core point processes," *IEEE Trans. Wireless Commun.*, vol. 15, no. 6, pp. 4074–4087, Sep. 2015.

[6] S. Mosleh, H. Almosa, E. Perrins, and L. Liu, "Downlink resource allocation in cell-free massive mimo systems," in *Proc. IEEE ICNC*, Feb 2019, pp. 883–887.

[7] Y. Zhang, H. Cao, M. Zhou, Y. Li, and L. Yang, "Power minimization for cell-free massive mimo," in *Proc. IEEE ICCE-TW*, 2019, pp. 1–2.

[8] E. Basar, M. Di Renzo, J. De Rosny, M. Debbah, M.-S. Alouini, and R. Zhang, "Wireless commun. through reconfigurable intelligent surfaces," *IEEE Access*, vol. 7, pp. 116 753–116 773, 2019.

[9] E. Björnson, Ö. Özdogan, and E. G. Larsson, "Intelligent reflecting surface versus decode-and-forward: How large surfaces are needed to beat relaying?" *IEEE Wireless Commun. Lett.*, vol. 9, no. 2, pp. 244–248, 2020.

[10] Y. Chen, "Performance of ambient backscatter systems using reconfigurable intelligent surface," *IEEE Commun. Lett.*, vol. 25, no. 8, pp. 2536–2539, 2021.

[11] N. I. Miridakis, T. A. Tsiftsis, G. Yang, P. A. Karkazis, and H. C. Leligou, "Semi-blind multiuser detection under the presence of reconfigurable intelligent surfaces," *IEEE Wireless Commun. Lett.*, pp. 1–1, 2021.

[12] M. Bashar, K. Cumanan, A. G. Burr, P. Xiao, and M. Di Renzo, "On the performance of reconfigurable intelligent surface-aided cell-free massive mimo uplink," in *Proc. IEEE GLOBECOM*, 2020, pp. 1–6.

[13] Z. Zhang and L. Dai, "A joint precoding framework for wideband reconfigurable intelligent surface-aided cell-free network," *IEEE Trans. Signal Process.*, vol. 69, pp. 4085–4101, 2021.

[14] 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," 2021.

[15] Q.-U.-A. Nadeem *et al.*, "Intelligent reflecting surface-assisted multi-user miso communication: Channel estimation and beamforming design," *IEEE Open Journal of the Commun. Society*, vol. 1, pp. 661–680, 2020.

[16] Ö. Özdogan, E. Björnson, and J. Zhang, "Performance of cell-free massive mimo with rician fading and phase shifts," *IEEE Trans. Wireless Commun.*, vol. 18, no. 11, pp. 5299–5315, 2019.

[17] H. Q. Ngo, H. Tataria, M. Matthaiou, S. Jin, and E. G. Larsson, "On the performance of cell-free massive mimo in ricean fading," in *Proc. IEEE Asilomar Conference on Signals, Systems, and Computers*, 2018, pp. 980–984.

[18] W. Jiang and H. D. Schotten, "Cell-free massive mimo-ofdm transmission over frequency-selective fading channels," *IEEE Commun. Lett.*, vol. 25, no. 8, pp. 2718–2722, 2021.

[19] J.-G. Kim and J.-T. Lim, "Map-based channel estimation for mimoofdm over fast rayleigh fading channels," *IEEE Trans. Veh. Technol.*, vol. 57, no. 3, pp. 1963–1968, 2008.

[20] S. E. Khamy, N. Korany, and H. Hassan, "Channel estimation techniques for wideband mimo-ofdm communication systems using complementary codes two-sided sequences," in *Proc. IEEE NRSC*, Sep. 2020, pp. 74–84.

[21] H. Ge, N. Garg, and T. Ratnarajah, "Channel estimation for generalized superimposed cell-free massive mimo-ofdm systems," in *Proc. IEEE SPAWC*, 2022, pp. 1–5.

[22] X. Zhang, D. Guo, K. An, X. Liang, and W. Ma, "Secure transmission in multi-pair af relaying massive mimo networks against active pilot spoofing attacks," *IEEE Access*, vol. 7, pp. 3547–3560, 2019.

[23] X. Dang, Z. Huang, Q. Li, and X. Yu, "Estimation of symbol timing in physical-layer network coding with arrival time differences," *IEEE Commun. Lett.*, vol. 21, no. 2, pp. 330–333, Feb 2017.

[24] K.-P. Chou, J.-C. Lin, and H. V. Poor, "Disintegrated channel estimation in filter-and-forward relay networks," *IEEE Trans. Commun.*, vol. 64, no. 7, pp. 2835–2847, July 2016.

[25] H. Zhang, S. Gao, D. Li, H. Chen, and L. Yang, "On superimposed pilot for channel estimation in multicell multiuser mimo uplink: Large system analysis," *IEEE Trans. Veh. Technol.*, vol. 65, no. 3, pp. 1492–1505, March 2016.

[26] I. A. Arriaga-Trejo, A. G. Orozco-Lugo, A. Veloz-Guerrero, and M. E. Guzman, "Widely linear system estimation using superimposed training," *IEEE Trans. Signal Process.*, vol. 59, no. 11, pp. 5651–5657, Nov 2011.

[27] J. Tugnait and W. Luo, "On channel estimation using superimposed training and first-order statistics," *IEEE Commun. Lett.*, vol. 7, no. 9, pp. 413–415, Sep. 2003.

[28] V. Nguyen, H. D. Tuan, H. H. Nguyen, and N. N. Tran, "Optimal superimposed training design for spatially correlated fading mimo channels," *IEEE Trans. Wireless Commun.*, vol. 7, no. 8, pp. 3206–3217, August 2008.

[29] F. Gao, B. Jiang, X. Gao, and X.-D. Zhang, "Superimposed training based channel estimation for ofdm modulated amplify-and-forward relay networks," *IEEE Trans. Commun.*, vol. 59, no. 7, pp. 2029–2039, July 2011.

[30] J. C. Estrada-Jimnez, B. G. Guzmn, M. J. Fernndez-Getino Garca, and V. P. G. Jimnez, "Superimposed training-based channel estimation for miso optical-ofdm vlc," *IEEE Trans. Veh. Technol.*, vol. 68, no. 6, pp. 6161–6166, June 2019.

[31] H. Zhang and B. Sheng, "An enhanced partial-data superimposed training scheme for ofdm systems," *IEEE Commun. Lett.*, vol. 24, no. 8, pp. 1804–1807, Aug 2020.

[32] L. He, Y.-C. Wu, S. Ma, T.-S. Ng, and H. V. Poor, "Superimposed training-based channel estimation and data detection for ofdm amplify-and-forward cooperative systems under high mobility," *IEEE Trans. Signal Process.*, vol. 60, no. 1, pp. 274–284, Jan 2012.

[33] N. Garg, A. Jain, and G. Sharma, "Partially loaded superimposed training scheme for large mimo uplink systems," *Wireless Personal Communication*, vol. 100, no. 4, pp. 1313–1338, 2018.

[34] N. Garg and T. Ratnarajah, "Generalized superimposed training scheme in cell-free massive mimo systems," *IEEE Trans. Wireless Commun.*, vol. 21, no. 9, pp. 7668–7681, Sep. 2022.

[35] N. Garg, H. Ge, and T. Ratnarajah, "Generalized superimposed training scheme in irs-assisted cell-free massive mimo systems," *IEEE J. Sel. Topics in Signal Process.*, vol. 16, no. 5, pp. 1157–1171, Aug. 2022.

[36] H. Ge, N. Garg, and T. Ratnarajah, "Generalized superimposed channel estimation for uplink ris-aided cellfree massive mimo systems," in *Proc. IEEE WCNC*, 2022, pp. 405–410.

[37] H. Ge, N. Garg, and T. Ratnarajah, "Design of generalized superimposed training for uplink cell-free massive mimo systems," in *Proc. IEEE VTC*, 2022, pp. 1–5.
[38] S. Ahmed, T. Ratnarajah, M. Sellathurai, and C. Cowan, "Iterative receiver for mimo-ofdm and their convergence behavior," *IEEE Trans. Veh. Technol.*, vol. 58, pp. 461–468, Jan. 2009.
[39] Q.-U.-A. Nadeem, A. Kammoun, A. Chaaban, M. Debbah, and M.-S. Alouini, "Asymptotic max-min sinr analysis of reconfigurable intelligent surface assisted miso systems," *IEEE Trans. Wireless Commun.*, vol. 19, no. 12, pp. 7748–7764, Dec 2020.
[40] J. Tugnait and X. Meng, "On superimposed training for channel estimation: performance analysis, training power allocation, and frame synchronization," *IEEE Trans. Signal Process.*, vol. 54, no. 2, pp. 752–765, Feb 2006.

Hanxiao Ge (Graduate Student Member, IEEE) received an M.Sc. degree in Signal Processing and Communications from The University of Edinburgh, United Kingdom, in 2020. He is currently with Institute for Digital Communications, The University of Edinburgh as a Ph.D. student. His research interests include beyond 5G wireless communications, cell-free massive MIMO systems, and reconfigurable intelligent surfaces.

Navneet Garg (Member, IEEE) received the B.Tech. degree in Electronics and Communication Engineering from College of Science & Engineering, Jhansi, India, in 2010, and the M.Tech. degree in Digital Communications from ABV-Indian Institute of Information Technology and Management, Gwalior, in 2012. He has completed the Ph.D. degree in June 2018 from the Department of Electrical Engineering at the Indian Institute of Technology Kanpur, India. From July 2018-Jan 2019, he visited The University of Edinburgh, UK. From February 2019-2020, he is employed as a Research Associate in Heriot-Watt University, Edinburgh, UK. Since February 2020, he is working as a Research Associate in The University of Edinburgh, UK. His main research interests include wireless communications, signal processing, optimization, and machine learning.

Tharmalingam Ratnarajah (Senior Member, IEEE) is currently with the Institute for Digital Communications, The University of Edinburgh, UK, as a Professor in Digital Communications and Signal Processing. His research interests include signal processing and information theoretic aspects of beyond 5G wireless networks, full-duplex radio, mmWave communications, random matrices theory, interference alignment, statistical and array signal processing, and quantum information theory. He has published over 400 publications in these areas and holds four U.S. patents. He has supervised 16 Ph.D. students and 21 post-doctoral research fellows and raised $11+ million USD of research funding.