Generalized Superimposed Training Scheme In IRS-assisted Cell-free Massive MIMO Systems

Generalized Superimposed Training Scheme In IRS-assisted Cell-free Massive MIMO Systems

Navneet Garg, *Member, IEEE*, Hanxiao Ge, *Student Member, IEEE*, and Tharmalingam Ratnarajah, *Senior Member, IEEE*.

***Abstract*—In this paper, for a cell-free massive multi-input multi-output (MIMO) system assisted with intelligent reflective surface (IRS) panels, a generalized superimposed pilot (GSP) training scheme is proposed, where the available number of pilots are equal to the coherence time slots, and the transmitting data symbols are spread over the coherence time with the help of simple precoding. Further, in order to keep the system scalable, a low complexity approach is employed for processing, and the corresponding rate components are analyzed. It is shown that with careful design of precoding and number of data symbols, the GSP symbols can provide much better channel estimation and data detection performance, as compared to the regular pilot scheme and the conventional superimposed scheme. These results, verified via simulations, shows that the centralized processing improves the data detection performance than localized processing. The pilot contamination effect, is significantly reduced due to the availability of larger number of pilots, as compared to the regular pilots transmission. For four IRS panels in the system, the proposed scheme is shown to reduce the channel estimation errors by 74% approximately.**

I. Introduction

In the recent years, cell-free massive multi-input multi-output (CF-mMIMO) systems have gained considerable attention for next generation wireless networks (beyond 5G or 6G), since it is able to achieve all merits of traditional distributed large-scale MIMO and network MIMO systems with simple precoding schemes [1]–[3]. These merits include tremendous macro-diversity and coverage ratio, high spectral and energy efficiencies, low interference [4], [5]. In such systems, the spatial data detection and the system capacity heavily depends on the CSI quality. To improve the channel estimation performance in CF-mMIMO system and to make the system scalable to the number of users and access points, low complexity matched filter based approaches are investigated in [6]. However, as the number of user equipments (UEs) exceeds the number of available pilots, the reuse of pilots among UEs gives rise to the pilot contamination, which bottlenecks both the channel estimation and data detection, and does not vanish even in the asymptotic regime $L \to \infty$ [7], where L denotes the number of access points (APs). To minimize the pilot contamination, several pilot assignment (PA) methods have been studied, including random PA, greedy PA, Tabu-search-based PA [8], structured PA based on geographical locations [9], graph coloring based PA [10]. Dynamic cooperation clustering has also been used to reuse the pilots in an efficient way in [6]. In [11], sparse channel matrix estimates are improved using deep neural networks for mmWave systems. In [12], [13], pilot power control is proposed to reduce the pilot contamination using convex approximation approach, whereas the rate-optimized power allocation is considered in [14], [15] via geometric programming. Energy efficiency maximization for power control is investigated for mmWave system in [16]. Authors in [17], [18] have analyzed the performance for Rician fading channels.

Note that the above works employ regular pilots (RPs) transmission, where the pilot and data symbols are sent separately in the coherence interval $(T = \tau_p + \tau_d)$, where τ_p and τ_d are the number of time slots used for pilot and data transmission. The value of τ_p decides the number of available pilots, which in turn decide the strength of pilot contamination. In contrast to RPs, the superimposed pilots (SPs) have been investigated in literature to have interesting features in traditional large-scale MIMO systems [4], with orthogonal frequency-division multiplexing (OFDM) in [19]–[22]. In the SP scheme [23], pilot and data symbols are transmitted simultaneously over a coherence time block. Particularly, the SP scheme benefits from suppressing pilot contamination by reducing the possibility of pilot reuse, since T pilots are available and $T > \tau_p$ [24], [25]. However, the correlation between pilot and data symbols reduces the quality of channel estimation and deteriorates the data detection process. This performance analysis has been verified via the asymptotic and closed-form analytical expressions of sum rate for an uplink massive MIMO system [26]–[28]. In spite of the performance degradation, the above works indicate that the SP scheme outperform RPs in terms of the achievable sum rate.

Interestingly, to improve the SP scheme's performance, a general SP framework has been proposed in our previous work [7]. The related works include the literature [21], [29]–[31]. In [29], information theoretic bounds are obtained for separate and joint channel and data estimations. The work in [30] focus on single carrier precoding optimization for peak to average power ratio, whereas [21], [31] is oriented towards multi-carrier transmissions. In the GSP scheme, instead of sending the same number of data symbols as the number of pilots (T), the number of data symbols is reduced, optimized, and further precoded to avoid the possible pilot contamination in both the channel estimation and the data detection. This general SP scheme achieves significantly better performance than the conventional SP scheme, in terms of both the channel estimation mean-squared error (MSE) and the sum rates.

This work was supported by U.K. Engineering and Physical Sciences Research Council (EPSRC) under Grant EP/T021063/1. Authors are with Institute of digital communications, School of engineering, The University of Edinburgh, Edinburgh, UK, EH9 3FG. Emails: {ngarg, t.ratnarajah}@ed.ac.uk.

A. Intelligent reflective surfaces

Parallelly, in the recent years, reconfigurable intelligent surfaces (IRS) are also investigated, where the surface at a geographical location equipped a large number of reflective elements. Each element can provide the phase shifts of the signals, which can be configured via a central or the IRS controller [32]. Due to the presence of the non-line-of-sight (nLOS) component of the multi-path fading channel in the signal transmission, the spectral efficiency decreases. The utilization of IRS can improve the performance, since a properly configured IRS can align or reflect the signal in a suitable direction to improve the channel strength [33], [34]. For beyond 5G and 6G, IRS based applications are investigated in [35]–[37]. Further, with IRS, the works like stochastic geometry analysis for centralized and local processing systems [38], [39], joint-transmit beamforming with reflectors designs [40], spectral-efficiency maximization [41], and more (e.g., energy harvesting, relaying, secrecy rate) are carried out. The required number of reflective elements is investigated in [42]. These studies show performance improvements due to the presence of an IRS. The motivation of this work is to investigate the improvements for a multi-cell system in presence of multiple-IRSs.

In literature [17], [43]–[46], authors have investigated the integration of CF-mMIMO systems with an IRS panel for Rayleigh fading channels. Therefore, in such a system, it will be interesting to compare the performances of RPs, SPs and GSPs, especially for Rician fading channels. These superimposed schemes are attractive for the most wireless networks, including future next-generation beyond 5G systems, where high data rates are desirable [31], [33].

B. Contributions

In this paper, a CF-mMIMO system is considered with multiple users, where the access points are connected via a central processing unit (CPU) [5]. For the proposed general SP (GSP) scheme, which is a generalization from [7] and for the CF-mMIMO system with a CPU, we consider two scenarios, viz., centralized and localized processing, and analyze the channel estimation MSE and sum rates. Note that these scenarios correspond to the cooperation levels $L4$ and $L1$ of [5]. For the SP or GSP scheme, the performance of levels $L2$ and $L3$ is similar to $L4$. With the same transmit power constraint, the channel and data estimates of RPs and conventional SPs are obtained. Simulation results verify the expressions and shows significant gains for GSPs, as compared to RPs and conventional SPs. Contributions of the article can be summarized as follows.

1) The GSP scheme: Instead of choosing the same number of data symbols as the number of pilots, the generalized SP (GSP) scheme is proposed [7], where any number (from 1 to T) of data symbols can be used for reliable communication. Moreover, the data symbols are precoded in the general scheme, making the channel estimation and data detection less susceptible to pilot contanimations.

2) Analysis for IRS-aided CF-mMIMO system: For both the centralized and localized processing scenarios, the MSEs of channel estimates, the powers of self-interference (SI) and cross-interference (CI), and the sum rates are derived to evaluate the general SP scheme. The analysis shows that the centralized operation with CPU can improve the rates significantly, and asymptotically for large number of APs, signal-to-interference-plus-noise-ratio (SINR) can be seen to be improved at least by a factor of L.

3) IRS optimization: The phase shifts at IRSs are optimized to minimize the channel estimation error variance. The optimization problem is simplified to a quadratic program in terms of channel correlation matrices.

4) Comparison and simulations: For the same transmit power constraint as in the general SP scheme, the estimates of channels and data for both the centralized and localized scenarios are obtained for different possible cases based on number of users, number of data streams and the number of coherence time slots. In simulations, the analytical expressions are verified, and the different performance indicators show the improved performance of the general SP scheme for CF-mMIMO system.

C. Organization

Section II describes the details the cell-free system and the general SP scheme. The analysis for centralized and localized processing is given in Section III. Different pilot schemes are compared in Section IV, followed by simulations and conclusion in Section V and VI, respectively.

D. Notations

Scalars, vectors and matrices are represented by lower case (a), lower case bold face ($\mathbf{a}$) and upper case bold face ($\mathbf{A}$) letters, respectively. Conjugate, transpose, Hermitian transpose, element-wise Hadamard product, and Kronecker product of matrices are denoted by $(\cdot)^*$, $(\cdot)^T$, $(\cdot)^\dagger$, $\odot$ and $\otimes$, respectively. $\mathcal{CN}(\mu, \mathbf{R})$ represents a circularly symmetric complex Gaussian random vector with mean μ and covariance matrix $\mathbf{R}$. The notations $\|\cdot\|_2$ and $\|\cdot\|_F$ denote the l_2 norm and Frobenious norm, respectively. The notation vec$(\mathbf{X})$ denotes the vector obtained by stacking the columns of matrix $\mathbf{X}$. $\mathcal{D}(a_1, a_2)$ and $\mathcal{BD}(\mathbf{A}_1, \mathbf{A}_2)$ denote a block diagonal matrix with respectively scalars a_1, a_2 and matrices $\mathbf{A}_1, \mathbf{A}_2$ as its diagonal components. $\mathbb{O}(\mathbf{X})$ denotes the unitary part of the QR decomposition of $\mathbf{X}$ [26]. $\lambda_{\max}(\mathbf{A})$ and $\nu_{\max}(\mathbf{A})$ denote the maximum eigenvalue and the corresponding eigenvector of $\mathbf{A}$. The Kronecker delta $\delta_{kj} = 1$, when $k = j$, and is 0 otherwise. mod(a, b), $\lceil a \rceil$, $\lfloor a \rfloor$ and $1_{a \in \mathcal{A}}$ respectively denote the remainder of a/b, ceil and floor value of a, and the indicator function.

II. System Model

In a given geographical region, we consider IRS-aided cell-free mMIMO system with L APs and a CPU as shown in Figure 1, which serve K single antenna users, and each AP has M antennas. This region consists of N_B blockages, through which multi-path fading channels are determined. The surfaces of a few of these blockages are equipped with the IRS panels.

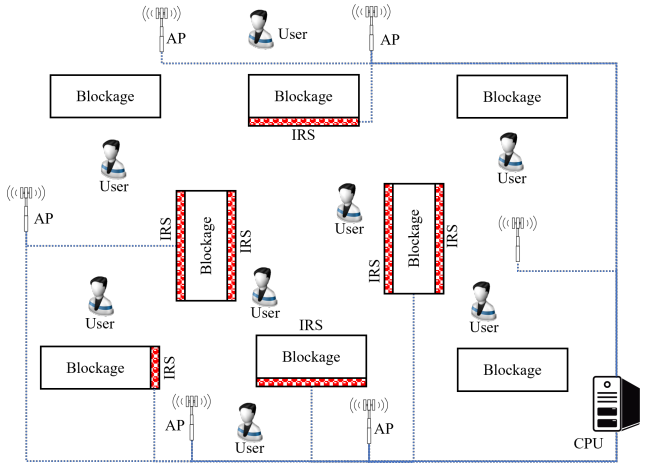

Figure 1. CF-mMIMO system with APs, users, CPU, IRSs and blockages.

Let N_{IRS} be the number of IRS panels deployed in this region, and each IRS panel is assumed to have Q number of passive elements. Given T number of slots in the coherence time, the received signal at the l^{th} AP can be written as

$$\mathbf{Y}_l = \sum_k \mathbf{h}_{lk}\mathbf{x}_k^H + \mathbf{W}_l = \mathbf{H}_l\mathbf{X}^H + \mathbf{W}_l, \tag{1}$$

where $\mathbf{x}_k$ is the $T \times 1$ transmitted vector from the k^{th} user, for each $k = 1, \ldots, K$; the matrix $\mathbf{W}_l$ is the $M \times T$ additive Gaussian noise at the l^{th} AP with zero mean and σ^2 variance across its each entry; the $M \times 1$ vector $\mathbf{h}_{lk}$ denotes the effective multi-path channel. The matrices $\mathbf{H}_l = [\mathbf{h}_{l1}, \ldots, \mathbf{h}_{lK}]$ and $\mathbf{X} = [\mathbf{x}_1, \ldots, \mathbf{x}_K]$ are obtained by concatenating the respective vector entries. The effective channel can be expressed as

$$\mathbf{h}_{lk} = \mathbf{g}_{lk} + \sum_i \mathbf{E}_{li}\Phi_i\mathbf{u}_{ik} \tag{2a}$$

$$= \mathbf{g}_{lk} + \sum_i \left(\mathbf{E}_{li} \odot \mathbf{1}_M\mathbf{u}_{ik}^T\right)\boldsymbol{\phi}_i \tag{2b}$$

$$= \mathbf{g}_{lk} + \mathbf{U}_{lk}\boldsymbol{\phi}, \tag{2c}$$

where $i = 1, \ldots, N_{IRS}$; the vector $\mathbf{g}_{lk}$ represents the $M \times 1$ multi-path channel between the l^{th} AP and the k^{th} user via $N_B - N_{IRS}$ number of blockages; $\mathbf{E}_{li}$ is an $M \times Q$ multi-path channel from i^{th} IRS panel to the l^{th} AP; $\mathbf{u}_{ik}$ denotes the $Q \times 1$ multi-path channel from the k^{th} user to the i^{th} IRS panel; the diagonal matrix $\Phi_i = \mathcal{D}(\phi_{i1}, \ldots, \phi_{iQ})$ such that $|\phi_{iq}| = 1, \forall i, \forall q = 1, \ldots, Q$ and $\boldsymbol{\phi}_i^T = [\phi_{i1}, \ldots, \phi_{iQ}]$;

$$\mathbf{U}_{lk} \coloneqq \left[\mathbf{E}_{l1} \odot \mathbf{1}_M\mathbf{u}_{1k}^T, \ldots, \mathbf{E}_{lN_{IRS}} \odot \mathbf{1}_M\mathbf{u}_{N_{IRS}k}^T\right] \tag{3}$$

is of size $M \times QN_{IRS}$ and $\boldsymbol{\phi}^T = \left[\boldsymbol{\phi}_1^T, \ldots, \boldsymbol{\phi}_{N_{IRS}}^T\right] \in \mathbb{C}^{QN_{IRS} \times 1}$. When IRS panel is not present, we set $\Phi_i = \rho\mathbf{I}_Q, \forall i = 1, \ldots, N_{IRS}$, where ρ denotes the attenuation introduced by a blockage in absence of a IRS panel. However, when IRS panels are deployed, the values of IRS panel coefficients are obtained from the channel statistics to minimize the channel estimation MSE. The details of each channel model and the transmitted symbol vector $\mathbf{x}$ is provided in the subsequent subsections, respectively. To generalize, we assume Rician fading for all channel models. A comprehensive list of variables is given in Table I.

Variables (& indices)	Description
L (l, m)	number of APs
K (k, j)	number of users
M	number of antennas at each AP
d (n)	number of data symbols
T (t)	number of coherence time slots
N_{IRS} (i)	number of IRS panels
Q (q)	number of elements in each IRS
N_B	number of blockages
$\mathbf{p}_k, \mathbf{s}_k, \mathbf{Z}_k$	k^{th} pilot, data and precoder
$\mathbf{x}_k$	superimposed symbol of k^{th} user
$\lambda \in (0, 1)$	power fraction
$\mathcal{C}_t$	a set of users using t^{th} pilot
$\mathcal{Z}$, $\mathcal{Z}_j$	a 2D-set and its first projection for $\mathbf{Z}_k^H\mathbf{p}_j \neq \mathbf{0}, \forall k \neq j$
P	transmit power per time slot
$\mathbf{F}, \mathbf{f}_t$	Fourier matrix and its column
$\mathbf{h}_{lk} = \bar{\mathbf{h}}_{lk} + \tilde{\mathbf{h}}_k$	effective channel, constant, and random parts
$\mathbf{g}_{lk}, G_{lk}, \beta_{g,lk}, \bar{\mathbf{g}}_{lk}, \tilde{\mathbf{g}}_{lk},$	AP-user channel, Rician factor, path loss, constant and random parts
$\mathbf{E}_{li}, E_{li}, \beta_{E,li}, \bar{\mathbf{E}}_{li}, \tilde{\mathbf{E}}_{li},$	AP-IRS channel, Rician factor, path loss, constant and random parts
$\mathbf{u}_{ik}, U_{ik}, \beta_{u,ik}, \bar{\mathbf{u}}_{ik}, \tilde{\mathbf{u}}_{ik},$	IRS-user channel, Rician factor, path loss, constant and random parts
Φ_i, ϕ_{iq}	diagonal matrix of phase shifts ϕ_{iq}
$\mathbf{a}_M(\theta, \phi)$, λ_0	steering vector and wavelength
$\mathbf{R}_{lk,mj}^{[\mathbf{h}]}, \mathbf{C}_{lk}^{[\mathbf{h}]}$	cross-correlation $(\mathbf{h}_{lk}, \mathbf{h}_{mj})$ and covariance matrices $(\mathbf{h}_{lk})$
$\mathbf{R}_{k,j_1j_2}^{[\mathbf{Z}]}$	correlation of two superimposed symbols $(\mathbf{Z}_k^H\mathbf{x}_{j_1}, \mathbf{Z}_k^H\mathbf{x}_{j_2})$
$\hat{\mathbf{h}}_{lk}$, Δ_{lk}, $\bar{\Delta}_{lk}$	channel estimate, estimation error and its mean
$\mathbf{Y}_l, \mathbf{W}_l$	received signal and noise at l^{th} AP
$\mathbf{R}_{lk,mj}^{[\hat{\mathbf{h}}]}, \mathbf{C}_{lk}^{[\hat{\mathbf{h}}]}$	cross-correlation $(\hat{\mathbf{h}}_{lk}, \hat{\mathbf{h}}_{mj})$ and covariance matrices $(\hat{\mathbf{h}}_{lk})$
$\mathbf{R}_{lk,mj}^{[\Delta]}, \mathbf{C}_{lk,mk}^{[\Delta]}$	cross-correlation $(\Delta_{lk}, \Delta_{mj})$ and cross-covariance $(\Delta_{lk}, \Delta_{mk})$
$\mathbf{R}_{k,j}^{[\Delta]}, \mathbf{C}_k^{[\Delta]}$	cross-correlation (Δ_k, Δ_j) and covariance (Δ_k)
$\mathbf{R}_{k,j}^{[\hat{\mathbf{h}}]}, \mathbf{C}_k^{[\hat{\mathbf{h}}]}$	cross-correlation $(\hat{\mathbf{h}}_k, \hat{\mathbf{h}}_j)$ and covariance matrices $(\hat{\mathbf{h}}_k)$
$\hat{\mathbf{s}}_{lk}$, $\hat{\mathbf{s}}_k$	data estimate at l^{th} AP and the CPU

Table I
LIST OF VARIABLES.

A. Channel model for user to AP ($\mathbf{g}_{lk}$)

The multi-path channels passing from $N_B - N_{IRS}$ number of blockages can be given in terms of Rician channel model as

$$\mathbf{g}_{lk} = \sqrt{\frac{\beta_{g,lk}}{G_{lk}^+}}\left[\sqrt{G_{lk}}\bar{\mathbf{g}}_{lk} + \tilde{\mathbf{g}}_{lk}\right] \quad , \tag{4}$$

where $\beta_{g,lk}$ and G_{lk} denote the large-scale fading and the Rician factor; $\bar{\mathbf{g}}_{lk} = \mathbf{a}_M(\vartheta_{g,lk}, \varphi_{g,lk})$ and $\tilde{\mathbf{g}}_{lk} \sim \mathcal{CN}(0, \mathbf{I}_M)$ represent the LOS and NLOS components; the angles $\vartheta_{g,lk}$ and $\varphi_{g,lk}$ are azimuth and elevation angles between the AP and the k^{th} user. The notation G_{lk}^+ denotes plus one value, that is, $G_{lk}^+ = G_{lk} + 1$. The steering vector can be given as [47]

$$\mathbf{a}_M^T(\vartheta, \varphi) = \left[1, e^{-j\frac{2\pi}{\lambda_0}\mathbf{r}_2^T\mathbf{u}(\vartheta,\varphi)}, \ldots, e^{-j\frac{2\pi}{\lambda_0}\mathbf{r}_M^T\mathbf{u}(\vartheta,\varphi)}\right],$$

where $\mathbf{r}_m \in \mathbb{R}^3$ represents the location of m^{th} antenna element in the antenna plane, $\mathbf{r}_1 = [0, 0, 0]$, and $\mathbf{u}^T(\vartheta, \varphi) =$

$[\cos\vartheta\cos\varphi, \sin\vartheta\cos\varphi, \sin\varphi]$.

B. Channel model for user to IRS ($\mathbf{u}_{ik}$)

The channel from a user to an IRS panel can either be LOS or NLOS. Thus, the channel $\mathbf{u}_{ik}$ can be defined similar to the above as

$$\mathbf{u}_{ik} = \sqrt{\frac{\beta_{u,ik}}{U_{ik}^{+}}}\left[\sqrt{U_{ik}}\bar{\mathbf{u}}_{ik} + \tilde{\mathbf{u}}_{ik}\right] \quad , \tag{5}$$

where $\bar{\mathbf{u}}_{ik} = \mathbf{a}_Q(\vartheta_{u,ik}, \varphi_{u,ik})$, $\tilde{\mathbf{u}}_{ik} \sim \mathcal{CN}(0, \mathbf{I}_M)$ and $\vartheta_{u,ik}, \varphi_{u,ik}$ are azimuth and elevation angles at the i^{th} IRS panel.

C. Channel model for IRS to AP ($\mathbf{E}_{li}$)

The $M \times Q$ channel between the l^{th} AP and the i^{th} IRS panel can be written as

$$\mathbf{E}_{li} = \sqrt{\frac{\beta_{E,li}}{E_{li}^{+}}}\left[\sqrt{E_{li}}\bar{\mathbf{E}}_{li} + \tilde{\mathbf{E}}_{li}\right], \tag{6}$$

where $\bar{\mathbf{E}}_{li} = \mathbf{a}_M(\vartheta_{A,ik}, \varphi_{A,ik})\,\mathbf{a}_Q(\vartheta_{E,ik}, \varphi_{E,ik})^H$; $\mathrm{vec}\left[\tilde{\mathbf{E}}_{li}\right] \sim \mathcal{CN}(0, \mathbf{I}_{MQ})$; the angles $(\vartheta_{A,ik}, \varphi_{A,ik})$ and $(\vartheta_{E,ik}, \varphi_{E,ik})$ are the angles at the AP and the IRS panel respectively.

D. Effective channel from AP to user ($\mathbf{h}_{lk}$)

Based on the above channel models, the effective channel from the k^{th} user to the l^{th} AP can be expanded from the Equation (2a) as

$$\begin{aligned}\mathbf{h}_{lk} &= \sqrt{\frac{\beta_{g,lk}}{G_{lk}^{+}}}\left[\sqrt{G_{lk}}\bar{\mathbf{g}}_{lk} + \tilde{\mathbf{g}}_{lk}\right] \\ &+ \sum_i \sqrt{\frac{\beta_{E,li}}{E_{li}^{+}}\frac{\beta_{u,ik}}{U_{ik}^{+}}}\left[\sqrt{E_{li}}\bar{\mathbf{E}}_{li} + \tilde{\mathbf{E}}_{li}\right]\Phi_i\left[\sqrt{U_{ik}}\bar{\mathbf{u}}_{ik} + \tilde{\mathbf{u}}_{ik}\right] \\ &= \bar{\mathbf{h}}_{lk} + \tilde{\mathbf{h}}_{lk}\end{aligned} \tag{7}$$

where $\bar{\mathbf{h}}_{lk} = \sqrt{\frac{\beta_{g,lk}G_{lk}}{G_{lk}^{+}}}\bar{\mathbf{g}}_{lk} + \bar{\mathbf{U}}_{lk}\boldsymbol{\phi}$ and

$$\begin{aligned}\tilde{\mathbf{h}}_{lk} = &\sqrt{\frac{\beta_{g,lk}}{G_{lk}^{+}}}\tilde{\mathbf{g}}_{lk} + \sum_i \sqrt{\frac{\beta_{E,li}}{E_{li}^{+}}\frac{\beta_{u,ik}}{U_{ik}^{+}}} \\ &\times\left[\tilde{\mathbf{E}}_{li}\Phi_i\sqrt{U_{ik}}\bar{\mathbf{u}}_{ik} + \sqrt{E_{li}}\bar{\mathbf{E}}_{li}\Phi_i\tilde{\mathbf{u}}_{ik} + \tilde{\mathbf{E}}_{li}\Phi_i\tilde{\mathbf{u}}_{ik}\right]\end{aligned} \tag{8}$$

constitute the constant and random part the channel respectively with $\bar{\mathbf{U}}_{lk} = \left[\sqrt{\frac{\beta_{E,li}E_{li}}{E_{li}^{+}}\frac{\beta_{u,ik}U_{ik}}{U_{ik}^{+}}}\bar{\mathbf{E}}_{li} \odot \mathbf{1}_M\bar{\mathbf{u}}_{ik}^T, \forall i = 1, \ldots, N_{IRS}\right]$. It can be noted that the random part comprises of sum of Gaussian random variables and a product of Gaussian. Therefore, the resulting distribution will not be Gaussian. However, we can approximate this part to Gaussian, which is present in the following proposition.

Proposition 1. *The NLOS part of the effective Rician channel can be considered a correlated Gaussian random variable, that is, the effective channel distribution is* $\mathbf{h}_{lk} = \bar{\mathbf{h}}_{lk} + \tilde{\mathbf{h}}_{lk} \sim \mathcal{CN}\left(\bar{\mathbf{h}}_{lk}, \mathbf{C}_{h,lk}\right)$, *where the covariance matrix given as*

$$\begin{aligned}\mathbf{C}_{lk}^{[\mathbf{h}]} &= \mathbb{E}\left\{\left(\mathbf{h}_{lk} - \bar{\mathbf{h}}_{lk}\right)\left(\mathbf{h}_{lk} - \bar{\mathbf{h}}_{lk}\right)^H\right\} \\ &= \left[\frac{\beta_{g,lk}}{G_{lk}^{+}} + \sum_i \frac{\beta_{E,li}\beta_{u,ik}Q}{E_{li}^{+}}\right]\mathbf{I} + \sum_i \frac{\beta_{E,li}E_{li}}{E_{li}^{+}}\frac{\beta_{u,ik}}{U_{ik}^{+}}\bar{\mathbf{E}}_{li}\bar{\mathbf{E}}_{li}^H,\end{aligned} \tag{9}$$

and $\mathbf{R}_{lk,mj}^{[\mathbf{h}]} = \mathbb{E}\left\{\mathbf{h}_{lk}\mathbf{h}_{mj}^H\right\} = \bar{\mathbf{h}}_{lk}\bar{\mathbf{h}}_{mj}^H + \delta_{lm}\delta_{kj}\mathbf{C}_{lk}^{[\mathbf{h}]}$.

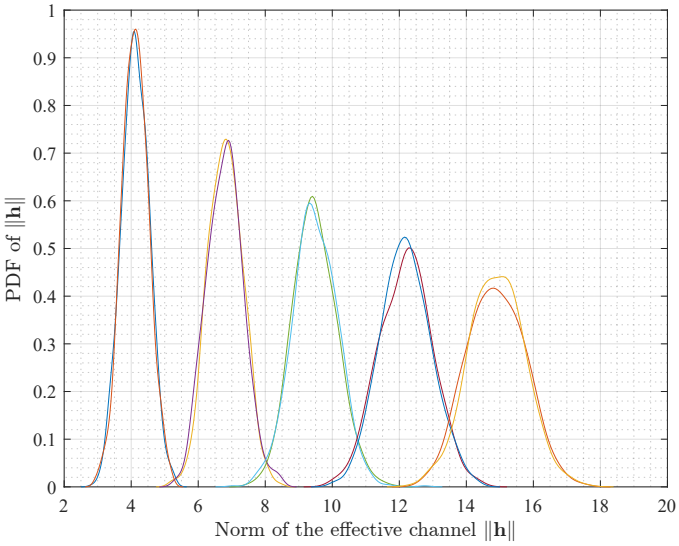

Figure 2. The PDFs of the norm of effective channel with and without Gaussian approximation with $Q = M = 2$.

The above results simplifies the distribution of the effective channel to Rician fading. To verify the proposition, we have plotted the PDF of the norm of the effective channel $\|\mathbf{h}_{lk}\|_2$ for different number of IRS panels with $Q = M = 2$ and $\beta_* = 1$ in Figure 2. It can be seen that the effective channel values agrees with the Gaussian-approximated values, for each number of IRS panels. Therefore, we use this proposition to derive the theoretical analysis in the following.

E. The GSP symbol

At the k^{th} user, the transmitted superimposed symbol for both the channel estimation and data transfer can be given as

$$\mathbf{x}_k = \sqrt{P}\left(\mathbf{p}_k\sqrt{\lambda} + \mathbf{Z}_k\mathbf{s}_k\sqrt{1-\lambda}\right), \tag{10}$$

where P is the transmit power per time slot for each user; $\lambda \in [0, 1]$ is the pilot power allocation factor; $\mathbf{p}_k$ is the $T \times 1$ pilot vector; $\mathbf{Z}_k$ is a $T \times d$ orthogonal precoder matrix; and the $\mathbf{s}_k$ denote the $d \times 1$ data vector having zero mean and covariance as $\mathbb{E}\left\{\mathbf{s}_k\mathbf{s}_k^H\right\} = \frac{1}{d}\mathbf{I}_d$. The pilot and precoder matrix can be chosen as follows. Let $\mathbf{F} = [\mathbf{f}_1, \ldots, \mathbf{f}_T]$ be the $T \times T$ orthogonal matrix such that $\mathbf{f}_k^H\mathbf{f}_j = T\delta_{kj}$. Then, each pilot can be selected as

$$\mathbf{p}_k = \mathbf{f}_{\underline{k}}, \forall k = 1, \ldots, K, \tag{11}$$

where $\underline{k} = \mathrm{mod}(k-1, T) + 1$, assigning the available T pilots in a round robin manner when pilots are reused ($K > T$). Let $\mathcal{C}_t = \{k : \underline{k} = t, k = 1, \ldots, K\}$ for each $t = 1, \ldots, T$ denote the set of users, who use the t^{th} pilot.

Precoding matrix: The precoding matrix for the k^{th} transmitting user ($\mathbf{Z}_k$) can be chosen orthogonal to the pilot $\mathbf{p}_k$ as

$$\frac{\mathbf{Z}_k^H \mathbf{p}_j}{T} = \begin{cases} \mathbf{0}, & (k,j) \in \bar{\mathcal{Z}}, \\ \mathbf{e}_{(kj)}, & (k,j) \in \mathcal{Z}, \end{cases} \tag{12}$$

where $\mathcal{Z} = \{1,\ldots,K\}^2 \setminus \bar{\mathcal{Z}}$;
$\bar{\mathcal{Z}} = \{(k,j) : j \in \mathcal{C}_{\underline{k}}, \forall k = 1,\ldots,K\}$;
$\mathbf{e}_{(kj)}$ is a vector of zeros and ones with ones $e_{kj} = \|\mathbf{e}_{(kj)}\|_1$.

For further usage, we define $\mathcal{Z}_j$ be the projection of $\mathcal{Z}$ in the second dimension, i.e.,

$$\mathcal{Z}_j = \left\{k : \mathbf{Z}_k^H \mathbf{p}_j \neq \mathbf{0}, \forall k \neq j\right\},$$

which shows that the sets $\mathcal{C}_{\underline{k}} \setminus \{k\}$ and $\mathcal{Z}_k$ are disjoint by construction. For example, with $T = 2$, $K = 3$ and $d = 1$, we have $\mathbf{p}_1 = \mathbf{p}_3 = \mathbf{f}_1 = \mathbf{Z}_2$, $\mathbf{Z}_1 = \mathbf{Z}_3 = \mathbf{p}_2 = \mathbf{f}_2$. Then, $\mathcal{C}_1 = \{1,3\}$ and $\mathcal{Z}_1 = \left\{k : \mathbf{Z}_k^H \mathbf{p}_1 \neq \mathbf{0}, \forall k \neq 1\right\} = \{2\}$. The above allocation distributes the data symbols over the whole coherence time when more users are present, i.e., $T - K < d$; and for $T - K \geq d$, data symbols are spread over only $T - K$ dimensions in order to provide the accurate channel estimation with pilot contamination. If $T - K \geq d$ and $M \geq K$, the data symbols can be estimated in a much better and reliable manner.

The two other products can be defined as $\mathbf{p}_k^H \mathbf{p}_j = T\delta_{kj}$ and

$$\frac{\mathbf{Z}_k^H \mathbf{Z}_j}{T} = \begin{cases} \mathbf{J}_{[kj]}, & k \neq j, \\ \mathbf{I}_d, & k = j, \end{cases} \tag{13}$$

where $\mathbf{J}_n$ is a rank-n permutation matrix with n-ones. The transmit power constraint at the k-th user can be given as

$$\begin{aligned} \mathbb{E}\|\mathbf{x}_k\|^2 &= P \cdot \mathbb{E}\|\mathbf{p}_k\sqrt{\lambda} + \mathbf{Z}_k\mathbf{s}_k\sqrt{1-\lambda}\|^2 \\ &= P \cdot \left[\|\mathbf{p}_k\|^2\lambda + \mathbb{E}\|\mathbf{Z}_k\mathbf{s}_k\|^2(1-\lambda)\right] = PT, \end{aligned}$$

where $\|\mathbf{Z}_k\|_F^2 = Td$. The product of precoder matrix and the superimposed vector can be given as

$$\mathbf{R}_{k,j_1j_2}^{[\mathbf{z}]} = \mathbb{E}\frac{\mathbf{Z}_k^H \mathbf{x}_{j_1}}{T\sqrt{P}} \frac{\mathbf{x}_{j_2}^H \mathbf{Z}_k}{T\sqrt{P}} \tag{14}$$

$$= \lambda \mathbf{e}_{kj_1}\mathbf{e}_{kj_2}^T + \frac{1-\lambda}{d}\delta_{j_1j_2}\mathbf{J}_{[kj_1]}\mathbf{J}_{[kj_2]}^T. \tag{15}$$

III. Channel and data estimations

In the cell-free system, APs are connected via fronthaul connections to a CPU that has higher computational resources. Hence, the APs can cooperate to provide the better performance to UEs. The l^{th} AP receives the signal, and can use the available channel estimaties to detect the data signals locally, or fully delegate the data detection to the CPU, which can combine the inputs from all APs to provide superior performance [5]. In the following, we first provide the analysis of localized processing based channel estimation and data detection. Thereafter, the CPU based centralized processing analyzed.

A. Localized processing

To get the meaningful channel or data estimates, the received signal equation should satisfy the necessary condition that the number of equations (MT) must be at least equal to the number of variables, that is,

$$MK + Kd \leq MT, \tag{16}$$

where MK and Kd stand for number of channel entries and data symbols, respectively. If the above condition is not satisfied, it leads to pilot and data contamination. For example, for $K = 3$ and $T = 6$, $d \leq M\left(\frac{T}{K} - 1\right) = M$. For $M = K$, we have $d \leq T - K$.

1) Channel estimation: At the l^{th} AP, the channel estimates can be obtained using least squares (LS) method as

$$\hat{\mathbf{h}}_{lk} = \arg\min_{\mathbf{h}_{lk}} \left\|\mathbf{Y}_l - \sqrt{P\lambda}\mathbf{h}_{lk}\mathbf{p}_k^H\right\|_F^2 \tag{17a}$$

$$= \frac{\mathbf{Y}_l\mathbf{p}_k}{T\sqrt{P\lambda}} = \mathbf{h}_{lk} + \Delta_{lk}, \tag{17b}$$

where the estimation error Δ_{lk} is given as

$$\Delta_{lk} = \frac{1}{T\sqrt{P\lambda}}\sum_{j\neq k}\mathbf{h}_{lj}\mathbf{x}_j^H\mathbf{p}_k + \frac{\mathbf{W}_l\mathbf{p}_k}{T\sqrt{P\lambda}} \tag{18a}$$

$$= \sum_{j\in\mathcal{C}_{\underline{k}}\setminus\{k\}}\mathbf{h}_{lj} + \sqrt{\frac{1-\lambda}{\lambda}}\sum_{j\in\mathcal{Z}_k}\mathbf{h}_{lj}s_{(jk)}^* + \mathbf{w}_{lk}, \tag{18b}$$

with $s_{(jk)}^* = \mathbf{s}_j^H\mathbf{e}_{(jk)}$ and $\mathbf{w}_{lk} = \frac{\mathbf{W}_l\mathbf{p}_k}{T\sqrt{P\lambda}} \sim \mathcal{CN}\left(\mathbf{0}, \frac{\sigma^2}{PT\lambda}\mathbf{I}_M\right)$. In the above equation, the factor $\frac{\sigma^2}{PT\lambda}$ denotes the inverse value of the pilot SNR over T time slots. The mean of channel estimation error can be given as

$$\bar{\Delta}_{lk} = \mathbb{E}\{\Delta_{lk}\} = \sum_{j\in\mathcal{C}_{\underline{k}}\setminus\{k\}}\bar{\mathbf{h}}_{lj}, \tag{19}$$

since $\mathbb{E}\left\{s_{(jk)}^*\right\} = 0$. It can be seen that the non-zero mean only arises due to the pilot reuse. Then the correlation can be obtained as

$$\mathbf{R}_{lk,mj}^{[\Delta]} = \mathbb{E}\Delta_{lk}\Delta_{mj}^H = \bar{\Delta}_{lk}\bar{\Delta}_{mj}^H + \delta_{kj}\mathbf{C}_{lk,mk}^{[\Delta]}, \tag{20}$$

where $\mathbf{C}_{lk,mk}^{[\Delta]} = \mathbb{E}\left\{\left(\Delta_{lk} - \bar{\Delta}_{lk}\right)\left(\Delta_{mk} - \bar{\Delta}_{mk}\right)^H\right\}$ is obtained in the following result.

Lemma 2. *The cross covariance of the channel estimation error can be derived as*

$$\begin{aligned} \mathbf{C}_{lk,mk}^{[\Delta]} &= \frac{1-\lambda}{\lambda d}\sum_{j\in\mathcal{Z}_k}\bar{\mathbf{h}}_{lj}\bar{\mathbf{h}}_{mj}^H \\ &+ \delta_{lm}\left[\sum_{j\in\mathcal{C}_{\underline{k}}\setminus\{k\}}\mathbf{C}_{lj}^{[\mathbf{h}]} + \frac{1-\lambda}{\lambda d}\sum_{j\in\mathcal{Z}_k}\mathbf{C}_{lj}^{[\mathbf{h}]} + \frac{\sigma^2}{TP\lambda}\mathbf{I}_M\right]. \end{aligned} \tag{21}$$

Proof: Proof is given in Appendix-A. ∎

The above result relates the error covariance as a function of the covariance matrices of other effective channels, and it shows that as the portion of data in the GSP symbol is

decreased, the MSE decreases. For further computations, we define the following correlations as

$$\begin{aligned}\mathbf{R}_{lk,mj}^{[\hat{\mathbf{h}}]} &= \mathbb{E}\hat{\mathbf{h}}_{lk}\hat{\mathbf{h}}_{mj}^{H} \\ &= \left(\bar{\mathbf{h}}_{lk}+\bar{\Delta}_{lk}\right)\left(\bar{\mathbf{h}}_{mj}+\bar{\Delta}_{mj}\right)^{H}+\delta_{lm}\delta_{kj}\mathbf{C}_{lk}^{[\hat{\mathbf{h}}]},\end{aligned}$$

where $\mathbf{C}_{lk}^{[\hat{\mathbf{h}}]} = \mathbf{C}_{lk}^{[\mathbf{h}]} + \mathbf{C}_{lk,lk}^{[\Delta]}$. For the analysis with Rician fading, the following result is derived, which enables us to derive the analytical expressions in a organized manner to better understand the effect of correlations.

Lemma 3. *For $\mathbf{x} \sim \mathcal{CN}(\bar{\mathbf{x}}, \mathbf{C})$, the forth moment can be given as*

$$\mathbb{E}\|\mathbf{x}\|^4 = f_4(\bar{\mathbf{x}}, \mathbf{C}) = tr(\mathbf{R})^2 + 2tr(\mathbf{CR}) - tr(\mathbf{C}^2), \quad (22)$$

where $\mathbf{R} = \bar{\mathbf{x}}\bar{\mathbf{x}}^H + \mathbf{C}$. If $\bar{\mathbf{x}} = \mathbf{0}$ and $\mathbf{C} = \mathbf{I}_M$, $\mathbb{E}\|\mathbf{x}\|^4 = tr(\mathbf{C})^2 + tr(\mathbf{C}^2) = M^2 + M$.

Proof: Proof is given in Appendix-B. ∎

The above correlation value mainly depends on $\mathbf{C}$ and $\|\bar{\mathbf{x}}\|_2^2$. Thus, to get higher moment value, any of two can be improved.

2) IRS optimization: The objective of IRS panels is to improve the channel strengths. In a given region, coefficients of the IRS panels can be optimized in order to minimize the channel estimation error at the APs. Thus, after obtaining the expression of channel estimation error variance for a given value of IRS panel coefficients in the above, these coefficients will be optimized using the second order statistics. The IRS values optimization problem can be cast and solved in the following lemma.

Lemma 4. *The solution of the MSE minimization problem is obtained as*

$$\arg\min_{|\phi_{iq}|=1,\forall i,q} \sum_{lk} tr\mathbf{C}_{lk,lk}^{[\Delta]} = -\mathbf{U}_{\phi}^{-1}\mathbf{c}_{\phi} \odot \frac{1}{\left|\mathbf{U}_{\phi}^{-1}\mathbf{c}_{\phi}\right|}, \quad (23)$$

where $\mathbf{U}_{\phi} = \sum_{lk,j\in\mathcal{Z}_k} \bar{\mathbf{U}}_{lj}^H \bar{\mathbf{U}}_{lj}$ and $\mathbf{c}_{\phi}^H = \sum_{lk,j\in\mathcal{Z}_k} \sqrt{\frac{\beta_{g,lj}G_{lj}}{G_{lj}^{+}}}\bar{\mathbf{g}}_{lj}^H \bar{\mathbf{U}}_{lj}$.

Proof: Proof is provided in Appendix-C. ∎

Since the dominant strength of $tr\mathbf{C}_{lk}^{[\Delta]}$ is due to pilot reuse components, the above expression shows that one can configure the reflective elements using the LoS channel information to reduce the pilot contamination in the effective channel estimates. The above result is more suitable for pilot reuse case. Thus, for no-pilot reuse $T \geq K + d$, we get $\mathcal{Z}_k = \emptyset$; in that case, the values $\phi_{iq} = 1, \forall i, q$ are set.

3) Data detection: Towards the data estimates at the l^{th} AP for the k^{th} user, the data estimates via LS can be expressed as

$$\hat{\mathbf{s}}_{lk} = \arg\min_{\mathbf{s}_k} \left\|\mathbf{Y}_l - \sqrt{P(1-\lambda)}\hat{\mathbf{h}}_{lk}\mathbf{s}_k^H\mathbf{Z}_k^H\right\|_F^2 \quad (24a)$$

$$= \mathbf{Z}_k^H\mathbf{Y}_l^H\frac{\hat{\mathbf{h}}_{lk}}{\|\hat{\mathbf{h}}_{lk}\|_2^2}\cdot\frac{1}{T\sqrt{P(1-\lambda)}} \quad (24b)$$

$$= \mathbf{Z}_k^H\left[\sum_i \mathbf{x}_i\mathbf{h}_{li}^H\right]\frac{\hat{\mathbf{h}}_{lk}}{\|\hat{\mathbf{h}}_{lk}\|_2^2}\cdot\frac{1}{T\sqrt{P(1-\lambda)}} \quad (24c)$$

$$= \mathbf{s}_k + \underbrace{\mathbf{s}_k\left(\frac{\mathbf{h}_{lk}^H\hat{\mathbf{h}}_{lk}}{\|\hat{\mathbf{h}}_{lk}\|_2^2}-1\right)}_{:=\mathbf{s}_{lk,SI}} + \underbrace{\sum_{i\neq k}\frac{\mathbf{Z}_k^H\mathbf{x}_i}{T\sqrt{P}}\frac{\mathbf{h}_{li}^H\hat{\mathbf{h}}_{lk}}{\|\hat{\mathbf{h}}_{lk}\|_2^2}\frac{1}{\sqrt{1-\lambda}}}_{:=\mathbf{s}_{lk,CI}}$$

$$+ \frac{\mathbf{Z}_k^H\mathbf{W}^H\hat{\mathbf{h}}_{lk}}{\|\hat{\mathbf{h}}_{lk}\|_2^2}\cdot\frac{1}{T\sqrt{P(1-\lambda)}}, \quad (24d)$$

which respectively consists of desired signal term ($\mathbf{s}_k$), self-interference (SI) term ($\mathbf{s}_{lk,SI}$), cross-interference (CI) term ($\mathbf{s}_{lk,CI}$) and noise vector. The SI term depends on the accuracy of the channel estimates, that is, for perfect channel estimation, $\mathbf{s}_{lk,SI} = \mathbf{0}$. The CI terms depend on both the channel estimation of the desired channel and the transmission from other users. To analyze the effect of these terms, the following result has been derived. Note that for simplicity, the factor $\|\hat{\mathbf{h}}_{lk}\|_2^2$ is multiplied to both sides in the above equation.

Theorem 5. *For localized processing at the l^{th} AP, the power of desired signal, SI, CI and noise for the k^{th} user can be obtained as*

$$P_{lk,S} = f_4(\bar{\mathbf{h}}_{lk}+\bar{\Delta}_{lk}, \mathbf{C}_{lk,lk}^{[\hat{\mathbf{h}}]}), \quad (25)$$

$$\begin{aligned}P_{lk,SI} &= tr\left(\mathbf{R}_{lk,lk}^{[\mathbf{h}]}\mathbf{R}_{lk,lk}^{[\Delta]}\right) + f_4(\bar{\Delta}_{lk}, \mathbf{C}_{lk,lk}^{[\Delta]}) \\ &\quad + 2\Re\bar{\Delta}_{lk}^H\bar{\mathbf{h}}_{lk}tr\mathbf{R}_{lk,lk}^{[\Delta]},\end{aligned} \quad (26)$$

$$P_{lk,CI} = \frac{M}{1-\lambda}\sum_{j_1\neq k}\sum_{j_2\neq k} tr\mathbf{R}_{k,j_1j_2}^{[\mathbf{Z}]} \times \frac{1}{M}tr\left[\mathbf{R}_{lj_2,lj_1}^{[\mathbf{h}]}\mathbf{R}_{lk,lk}^{[\hat{\mathbf{h}}]}\right], \quad (27)$$

$$P_{lk,N} = \frac{\sigma^2 Md}{TP(1-\lambda)}\left[\frac{tr(\mathbf{R}_{lk,lk}^{[\hat{\mathbf{h}}]})}{M} + \frac{\sigma^2(M+1)}{T^2P\lambda}\right]. \quad (28)$$

Proof: Proof is given in Appendix-D. ∎

In the above results, the signal power ($P_{lk,S}$) depends both on LoS channel and the variance of channel estimates. With pilot reuse, the variance of estimates becomes; however, the power of SI ($P_{lk,SI}$) also becomes large, thereby reducing the effective information rates. In the SI, the covariance matrix $\mathbf{C}_{lk,lk}^{[\Delta]}$ plays a major role as $\mathbf{R}_{lk,lk}^{[\Delta]}$ is also composed of $\mathbf{C}_{lk,lk}^{[\Delta]}$. In the power of CI ($P_{lk,CI}$), cross-correlation between channels and the superimposed symbol design are the main components. Superimposed symbol design depends on system variables (T, K, d and M); thus, the value $tr\mathbf{R}_{k,j_1j_2}^{[\mathbf{Z}]}$ is limited and fixed for a given system. The strength of cross-correlation is due to the LoS components in the Rician channels. LoS components depends on the steering vectors. For a large array ($M \to \infty$), inner-product of two steering vectors diminishes, as the difference between their corresponding angles is increased. Since each AP has limited antennas, the cross-correlation factor is dominant in the CI. The power of effective

noise terms mainly depends on the transmit power (P and λ), number of time slots (T), and the number of antennas (M).

Since $tr(\mathbf{C}_{lk}^{[\hat{\mathbf{h}}]}) \propto M$, it can be seen that the desired signal and SI power increases proportional to M^2 from Lemma (3). The CI power is proportional to M. The M^2 factor in the noise power arises due to the same noise in both channel estimation and data detection. The resulting rate expression can be written as

$$R_{lk} = \log_2\left(1 + SINR_{lk}\right), \tag{29}$$

where $SINR_{lk} = \frac{P_{lk,S}}{P_{lk,SI}+P_{lk,CI}+P_{lk,N}}$. Asymptotically, for $M \to \infty$, we have

$$\lim_{M\to\infty} SINR_{lk} \to \frac{\frac{f_4(\bar{\mathbf{h}}_{lk}+\bar{\Delta}_{lk},\mathbf{C}_{lk,lk}^{[\hat{\mathbf{h}}]})}{M^2}}{\frac{f_4(\bar{\Delta}_{lk},\mathbf{C}_{lk,lk}^{[\Delta]})}{M^2} + \frac{\sigma^4 d}{T^3P^2\lambda(1-\lambda)}}. \tag{30}$$

Similarly, when $P \to \infty$ with the necessary condition (16) satisfied, we obtain $\mathbf{C}_{lk,lk}^{[\Delta]} \to \mathbf{0}$. It can be observed from (1) that when $M < K$ or (16) is not satisfied, the local processing is inefficient for detection due to increased interference from users and thus, centralized processing is essential, which is presented below.

B. Centralized processing

For meaningful channel and data estimates at the CPU, the system should satisfy the necessary condition, that is,

$$MLT \geq KM + Kd, \tag{31}$$

where MLT are the number of observations, and the KM and Kd correspond to channel and data estimates of K users. E.g, for $K = 3$ and $T = 6$, $d \leq M\left(\frac{LT}{K} - 1\right) = M\left(2L - 1\right)$. It shows that CPU based processing can increase the allowable number of data symbols per user or the number of users in the system by a factor of $\frac{\frac{LT}{K}-1}{\frac{T}{K}-1} \approx L$. If $ML = K$, $d \leq T - \frac{K}{L}$.

1) Channel estimation : In this level, all L APs forward their received signals to the CPU, which performs both the channel estimation and the data detection. At the CPU, the combined received signal can be written as

$$\underbrace{\begin{bmatrix}\mathbf{Y}_1\\ \vdots \\ \mathbf{Y}_L\end{bmatrix}}_{:=\mathbf{Y}} = \sum_k \underbrace{\begin{bmatrix}\mathbf{h}_{1k}\\ \vdots \\ \mathbf{h}_{Lk}\end{bmatrix}}_{:=\mathbf{h}_k} \mathbf{x}_k^H + \underbrace{\begin{bmatrix}\mathbf{W}_1\\ \vdots \\ \mathbf{W}_L\end{bmatrix}}_{:=\mathbf{W}}. \tag{32}$$

The corresponding channel estimates can be written as $\hat{\mathbf{h}}_k^H := \left[\hat{\mathbf{h}}_{1k}^H, \ldots, \hat{\mathbf{h}}_{Lk}^H\right] = \mathbf{h}_k^H + \Delta_k^H$, with $\Delta_k^H := \left[\Delta_{1k}^H, \ldots, \Delta_{Lk}^H\right]$. The correlation between two error vectors can be given as

$$\mathbf{R}_{k,j}^{[\Delta]} = \mathbb{E}\Delta_k\Delta_j^H = \bar{\Delta}_k\bar{\Delta}_j^H + \delta_{kj}\mathbf{C}_k^{[\Delta]}, \tag{33}$$

where $\mathbf{C}_k^{[\Delta]}$ is a $L \times L$ block matrix with $\mathbf{C}_{lk,mk}^{[\Delta]}$ as its $(l,m)^{th}$ block. Similarly, the estimated channel correlation can be defined as

$$\begin{aligned}\mathbf{R}_{k,j}^{[\hat{\mathbf{h}}]} &= \mathbb{E}\hat{\mathbf{h}}_k\hat{\mathbf{h}}_j^H \\ &= \left(\bar{\mathbf{h}}_k + \bar{\Delta}_k\right)\left(\bar{\mathbf{h}}_j + \bar{\Delta}_j\right)^H + \delta_{kj}\mathbf{C}_k^{[\hat{\mathbf{h}}]},\end{aligned}$$

where $\mathbf{C}_k^{[\hat{\mathbf{h}}]} = \mathbf{C}_k^{[\mathbf{h}]} + \mathbf{C}_k^{[\Delta]}$ and $\mathbf{C}_k^{[\mathbf{h}]} = \mathcal{D}\left(\mathbf{C}_{1k}^{[\mathbf{h}]}, \ldots, \mathbf{C}_{Lk}^{[\mathbf{h}]}\right)$ is a block diagonal matrix.

2) Data detection : Towards the data detection, the data estimates via LS can be expressed as

$$\hat{\mathbf{s}}_k = \arg\min_{\mathbf{s}_k} \left\|\mathbf{Y} - \sqrt{P(1-\lambda)}\hat{\mathbf{h}}_k\mathbf{s}_k^H\mathbf{Z}_k^H\right\|_F^2 \tag{34a}$$

$$= \mathbf{Z}_k^H\mathbf{Y}^H\frac{\hat{\mathbf{h}}_k}{\|\hat{\mathbf{h}}_k\|_2^2}\cdot\frac{1}{T\sqrt{P(1-\lambda)}} \tag{34b}$$

$$= \mathbf{s}_k + \underbrace{\mathbf{s}_k\left(\frac{\mathbf{h}_k^H\hat{\mathbf{h}}_k}{\|\hat{\mathbf{h}}_k\|_2^2} - 1\right)}_{:=\mathbf{s}_{k,SI}} + \underbrace{\sum_{j\neq k}\frac{\mathbf{Z}_k^H\mathbf{x}_j}{T\sqrt{P}}\frac{\mathbf{h}_j^H\hat{\mathbf{h}}_k}{\|\hat{\mathbf{h}}_k\|_2^2}\frac{1}{\sqrt{1-\lambda}}}_{:=\mathbf{s}_{k,CI}}$$

$$+ \frac{\mathbf{Z}_k^H\mathbf{W}^H\hat{\mathbf{h}}_k}{\|\hat{\mathbf{h}}_k\|_2^2}\cdot\frac{1}{T\sqrt{P(1-\lambda)}}, \tag{34c}$$

wherein the desired signal term ($\mathbf{s}_k$), SI term ($\mathbf{s}_{k,SI}$), CI term ($\mathbf{s}_{k,CI}$) and noise vector have similar affects as in (24d), and the strength of these terms is derived in the following lemma.

Theorem 6. *For centralized processing at the CPU, the power of desired signal, SI, CI and noise for the k^{th} user can be obtained as*

$$P_{k,S} = f_4\left(\bar{\mathbf{h}}_k + \bar{\Delta}_k, \mathbf{C}_k^{[\hat{\mathbf{h}}]}\right), \tag{35}$$

$$P_{k,SI} = \sum_{l,m} tr\left(\mathbf{R}_{lk,mk}^{[\mathbf{h}]}\mathbf{R}_{mk,lk}^{[\Delta]}\right) + f_4\left(\bar{\Delta}_k, \mathbf{C}_k^{[\Delta]}\right) + 2\Re\sum_{l=1}^{L}\bar{\Delta}_{lk}^H\bar{\mathbf{h}}_{lk}\sum_{m=1}^{L} tr\left(\mathbf{R}_{mk,mk}^{[\Delta]}\right), \tag{36}$$

$$P_{k,CI} = \frac{M}{1-\lambda}\sum_{j_1\neq k}\sum_{j_2\neq k} tr\mathbf{R}_{k,j_1j_2}^{[\mathbf{Z}]} \times \sum_{l,m}\frac{tr\left(\mathbf{R}_{lj_2,mj_1}^{[\mathbf{h}]}\mathbf{R}_{mk,lk}^{[\hat{\mathbf{h}}]}\right)}{M}, \tag{37}$$

$$P_{k,N} = \frac{\sigma^2 MLd}{TP(1-\lambda)}\left[\sum_l \frac{tr(\mathbf{R}_{lk,lk}^{[\hat{\mathbf{h}}]})}{ML} + \frac{\sigma^2(ML+1)}{T^2P\lambda}\right]. \tag{38}$$

Proof: Proof is given in Appendix-E. ∎

In addition to the insights drawn for Theorem 5, the dependence of different powers on the number of APs ($\propto L^2$) can be observed, showing the aggregation of signal, interference and noise components at the centralized CPU. The rate expression for the k^{th} user can be given as $R_k = \log_2\left(1 + SINR_k\right)$, where $SINR_k = \frac{P_{k,S}}{P_{k,SI}+P_{k,CI}+P_{k,N}}$. To get further insights with respect to L, for $L \to \infty$, we obtain

$$\lim_{L\to\infty} SINR_k = \frac{\frac{f_4\left(\bar{\mathbf{h}}_k+\bar{\Delta}_k,\mathbf{C}_k^{[\hat{\mathbf{h}}]}\right)}{M^2L^2}}{\frac{P_{k,SI}+P_{k,CI}}{M^2L^2} + \frac{\sigma^4 d}{T^3P^2\lambda(1-\lambda)}}, \tag{39}$$

where only the dominant component of noise power is given. Further, enforcing $M \to \infty$ provides

$$SINR_k \to \frac{\frac{f_4\left(\bar{\mathbf{h}}_k+\bar{\Delta}_k,\mathbf{C}_k^{[\hat{\mathbf{h}}]}\right)}{M^2L^2}}{\frac{f_4\left(\bar{\Delta}_k,\mathbf{C}_k^{[\Delta]}\right)}{M^2L^2} + \frac{\sigma^4 d}{T^3P^2\lambda(1-\lambda)}}, \tag{40}$$

where the cross interference vanishes for $M \to \infty$. It also shows that the main factor limiting the SINR is the channel estimation errors, which is controlled using the power

allocation factor λ. In other words, the factor λ can be optimized to get better data estimates and the SINRs.

Note that the CPU estimates of data can also be written in terms of local APs' estimates as

$$\hat{\mathbf{s}}_k = \sum_l w_{lk}\hat{\mathbf{s}}_{lk}, \tag{41}$$

where $w_{lk} = \frac{\hat{\mathbf{h}}_{lk}^H \hat{\mathbf{h}}_{lk}}{\sum_l \hat{\mathbf{h}}_{lk}^H \hat{\mathbf{h}}_{lk}}$.

IV. Comparison of SP and RP schemes

In this section, we compare with the conventional SP and RP schemes and with other cooperation schemes. Thus, in the following, first SP and RP schemes are provided in brief for the present system model.

A. Conventional SP scheme

In the conventional superimposed schemes the data and pilots are added together, i.e., $\mathbf{x}_k = \sqrt{P}\left(\mathbf{p}_k\sqrt{\lambda} + \sqrt{(1-\lambda)}\mathbf{F}\mathbf{s}_k\right)$ with $\mathbf{Z}_k = \mathbf{F}$, where the size of $\mathbf{s}_k$ is matched to the size of $\mathbf{p}_k$, that is, $d = T$. This leads to unnecessary data contamination and severe channel estimation errors. However, for the purpose of comparison, the analysis is as follows. First, the transmit power constraint can be verified as $\mathbb{E}\left\{\mathbf{x}_k^H\mathbf{x}_k\right\} = P\left(T\lambda + T(1-\lambda)\right) = PT$, where $\mathbb{E}\left\{\mathbf{s}_k\mathbf{s}_k^H\right\} = \frac{1}{T}\mathbf{I}_T$. Next, from the received signal equation in (1), the local channel estimates can similarly be computed as in (17b), where the estimation error Δ_{lk} for this case is obtained as

$$\Delta_{lk} = \sum_{j\in\mathcal{C}_{\underline{k}}\setminus\{k\}} \mathbf{h}_{lj} + \sqrt{\frac{1-\lambda}{\lambda}}\sum_{j\neq k}\mathbf{h}_{lj}\mathbf{s}_j^H\mathbf{F}^H\mathbf{p}_k + \frac{\mathbf{W}_l\mathbf{p}_k}{T\sqrt{P\lambda}}.$$

It can be seen that the data term in the channel estimation error is increased from $\frac{1}{d}$ in $\mathbf{C}_{lk,mk}^{[\Delta]}$ from (20) to T, that is, the increment factor is Td, which is significantly large. For local processing, the necessary condition $MT \geq KM + KT$ should be satisfied, and $MLT \geq KM + KT$ for centralized processing. The rest of equations for data estimation can be followed similarly as in the GSP scheme as

$$\hat{\mathbf{s}}_{lk} = \left(\frac{\mathbf{Y}_l^H\hat{\mathbf{h}}_{lk}}{\|\hat{\mathbf{h}}_{lk}\|_2^2} - \sqrt{P\lambda}\mathbf{p}_k\right)\frac{1}{\sqrt{TP(1-\lambda)}}. \tag{42}$$

B. The RP scheme

In the conventional non-superimposed schemes, the whole coherence time is divided into two parts, viz., training phase and data estimation phase. Let $T = T_p + T_d$, where T_p and T_d denote the durations of the training phase and data estimations phases, respectively. For K users in the system, to avoid the pilot reuse, we must have $T_p \geq K$. If $T_p < K$, this causes pilot reuse in the system deteriorating the channel and consecutively the data estimation due to pilot contamination. In simulations, we shall review both cases with and without pilot contamination.

1) Channel estimation: In this phase, each user transmits a $T_p \times 1$ vector $\mathbf{x}_k = \mathbf{p}_k\sqrt{\frac{\lambda TP}{T_p}}$ satisfying the power constraint $\mathbb{E}\|\mathbf{x}_k\|_2^2 = \lambda TP$, where $\mathbf{p}_k^H\mathbf{p}_j = T_p\delta_{kj}, \forall k, j$. The received signal at the l^{th} AP can be written as

$$\mathbf{Y}_{p,l} = \sum_k \mathbf{h}_{lk}\mathbf{p}_k^H\sqrt{\frac{\lambda TP}{T_p}} + \mathbf{W}_{p,l}. \tag{43}$$

The channel estimates via LS can be given as

$$\hat{\mathbf{h}}_{lk} = \frac{\mathbf{Y}_{p,l}\mathbf{p}_k}{T_p\sqrt{\frac{\lambda TP}{T_p}}} = \mathbf{h}_{lk} + \Delta_{lk}, \tag{44}$$

where $\Delta_{lk} = \sum_{j\in\mathcal{C}_{\underline{k}}\setminus\{k\}}\mathbf{h}_{lj} + \frac{\mathbf{W}_{p,l}\mathbf{p}_k}{T_p\sqrt{\frac{\lambda TP}{T_p}}}$. To get the channel estimates at the CPU, there are two options. Each AP can either forward the $M \times T_p$ received signal matrix $\mathbf{Y}_{p,l}, \forall l$, or directly forward the $M \times 1$ estimated channel vectors $\hat{\mathbf{h}}_{lk}, \forall l$ for K users. If $T_p = K$, then both approaches are equivalent, else if $T_p > K$, forwarding channel estimates is a better choice. Thus, at the CPU, from the section III-B1, we can write the channel error vector as $\Delta_k^H = \left[\Delta_{1k}^H, \ldots, \Delta_{Lk}^H\right]$.

2) Data estimation: In the data estimation phase, without loss of generality (to fairly compare with the superimposed schemes), let each user transmit d data streams over T_d time slots. Then, the transmitted signal from each k^{th} user is given in form of a $T_d \times 1$ vector as

$$\mathbf{x}_k = \sqrt{PT(1-\lambda)}\mathbf{V}_k\mathbf{s}_k, \tag{45}$$

where $\mathbf{V}_k$ is a $T_d \times d$ precoding matrix such that $\|\mathbf{V}_k\|_F^2 = d$, and $\mathbb{E}\left\{\mathbf{s}_k\mathbf{s}_k^H\right\} = \frac{1}{d}\mathbf{I}_d$. The transmit power constraint can be verified as $\mathbb{E}\left\{\mathbf{x}_k^H\mathbf{x}_k\right\} = PT(1-\lambda)\cdot tr\mathbb{E}\mathbf{V}_k\mathbf{s}_k\mathbf{s}_k^H\mathbf{V}_k^H = PT(1-\lambda)\frac{\|\mathbf{V}_k\|_F^2}{d} = PT(1-\lambda)$.

The received signal at the l^{th} AP can be written similar to (1), where we have MT_d equations and dK variables. For meaningful estimation, $MT_d \geq dK$ or $d \leq \frac{MT_d}{K}$, i.e., per user data streams should be less than the value $\frac{MT_d}{K}$. Least squares estimates can be given as

$$\hat{\mathbf{s}}_{lk} = \mathbf{V}_k^H\mathbf{Y}_l^H\frac{\hat{\mathbf{h}}_{lk}}{\hat{\mathbf{h}}_{lk}^H\hat{\mathbf{h}}_{lk}}\cdot\frac{1}{\sqrt{PT(1-\lambda)}} \tag{46a}$$

$$= \sum_i \mathbf{V}_k^H\mathbf{V}_i\mathbf{s}_i\frac{\mathbf{h}_{li}^H\hat{\mathbf{h}}_{lk}}{\hat{\mathbf{h}}_{lk}^H\hat{\mathbf{h}}_{lk}} + \mathbf{V}_k^H\mathbf{W}_l^H\frac{\hat{\mathbf{h}}_{lk}}{\hat{\mathbf{h}}_{lk}^H\hat{\mathbf{h}}_{lk}}\cdot\frac{1}{\sqrt{PT(1-\lambda)}} \tag{46b}$$

$$= \mathbf{s}_k + \underbrace{\left(\frac{\mathbf{h}_{lk}^H\hat{\mathbf{h}}_{lk}}{\hat{\mathbf{h}}_{lk}^H\hat{\mathbf{h}}_{lk}}\mathbf{V}_k^H\mathbf{V}_k - \mathbf{I}\right)\mathbf{s}_k}_{\text{self-interference}} + \underbrace{\sum_{i\neq k}\mathbf{V}_k^H\mathbf{V}_i\mathbf{s}_i\frac{\mathbf{h}_{li}^H\hat{\mathbf{h}}_{lk}}{\hat{\mathbf{h}}_{lk}^H\hat{\mathbf{h}}_{lk}}}_{\text{cross-interference}}$$

$$+ \mathbf{V}_k^H\mathbf{W}_l^H\frac{\hat{\mathbf{h}}_{lk}}{\hat{\mathbf{h}}_{lk}^H\hat{\mathbf{h}}_{lk}}\frac{1}{\sqrt{PT(1-\lambda)}} \tag{46c}$$

The CPU estimate can also be written in terms of local estimates as $\hat{\mathbf{s}}_k = \sum_l \frac{\hat{\mathbf{h}}_{lk}^H\hat{\mathbf{h}}_{lk}}{\sum_l \hat{\mathbf{h}}_{lk}^H\hat{\mathbf{h}}_{lk}}\hat{\mathbf{s}}_{lk}$. For meaningful data estimates, the system should satisfy the condition, $MLT_d \geq dK$ or $d \leq \frac{MLT_d}{K}$, i.e., per user data streams should be less than this number $\frac{MLT_d}{K}$.

It can be noted that this limiting case is not affected by the transmitted power. It is rather influenced by the number of

Scheme	L/C	Conditions
GSP	L	$d \leq M\left(\frac{T}{K}-1\right)$, $M \geq K$, $T \geq K+d$
	C	$d \leq M\left(\frac{LT}{K}-1\right)$, $ML \geq K$, $T \geq K+d$
SP	L	$K \leq \frac{MT}{M+T}$, $d=T$
	C	$K \leq \frac{MLT}{M+T}$, $d=T$
RP	L	$T_p > K$, $d \leq M\left(\frac{T}{K}-\frac{T_p}{K}\right)$
	C	$T_p > K$, $d \leq M\left(\frac{LT}{K}-\frac{LT_p}{K}\right)$

Table II
COMPARISON OF NECESSARY CONDITIONS FOR DIFFERENT TRANSMISSION SCHEMES WITH LOCALIZED (L) AND CENTRALIZED (C) PROCESSING.

data streams for transmission. When the number of streams is equal to the number of available slots, SINR reduces to a constant, which is due to the channel estimation errors.

C. Comparison of GSP with RP and SP schemes

These schemes can be compared in terms of sum rate and the estimation delay for localized and centralized processing. For localized processing, the delay for RP schemes is lower as compared to SP and GSP, in which data decoding requires to wait for T time slots. Therefore, SP/GSP schemes are useful for fast fading channels with small coherence time or can be utilized at the starting portion of the coherence time frame. Table II shows the different conditions for meaningful channel and data estimation for different schemes. The main conditions limiting the performance of RP and SP schemes are $T_p > K$ and $d = T$ respectively. In the GSP scheme for centralized processing, the large number of users can be accommodated given the constraint $K \leq ML$ with no-pilot reuse; however, $T \geq K + d$ is essential for reliable data detection.

The information rate corresponds to number of data symbols successfully communicated over the wireless channel. In terms of sum rate, GSP provides superior performance to both SP and RP, since GSP communicates more number of reliable data symbols. Regarding the computational overhead, all schemes bear the similar overhead.

D. Comparison with other cooperation scenarios from [5]

1) Level 3: local processing and large scale fading decoding: From the level 4 processing, the data estimates can be written in terms of a linear sum of local channel estimates. It means that the data estimates can be locally calculated and forwarded to the CPU for the final estimation, rather than forwarding the channel estimates. In this case, the SINR expression remains the same, since the coefficients of linear sum are the same.

2) Level 2: Local processing and centralized decoding: To further relax the cooperation requirements, the linear combining can be relaxed simple averaging. In this case, each l^{th} AP can send the value $\hat{\mathbf{s}}_{lk}\frac{\|\hat{\mathbf{h}}_{lk}\|_2^2}{M}$, instead of $\hat{\mathbf{s}}_{lk}$, which aggregates at the CPU as

$$\frac{1}{M}\sum_{l=1}^{L}\hat{\mathbf{s}}_{lk}\|\hat{\mathbf{h}}_{lk}\|_2^2 = \hat{\mathbf{s}}_k\frac{\|\hat{\mathbf{h}}_k\|_2^2}{M},$$

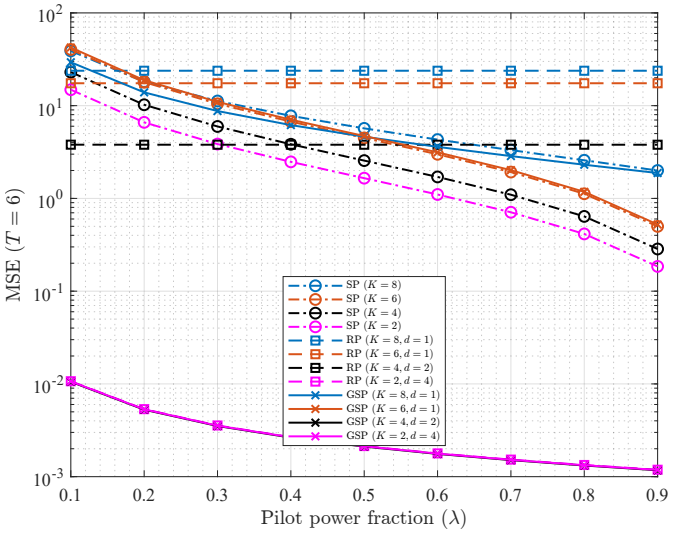

Figure 3. Averaged channel estimation MSE $\frac{1}{KL}\sum_{lk}\mathbb{E}\|\Delta_{lk}\|_2^2$ versus the pilot power allocation factor λ for different cases.

where $\frac{\|\hat{\mathbf{h}}_k\|_2^2}{M}$ is a constant and it does not affect the resulting SINR and sum rate.

Therefore, it can be concluded that for SP/GSP schemes, cooperation levels (L2-L4) have equal performance, which is better than non-cooperation level (L1).

V. SIMULATION RESULTS

A. Simulation settings

We choose $M = 2$, $L = 4$, $T = 6$ $N_{IRS} = 4$, $Q = 16$, and vary K and d in order to evaluate different possible cases as

- $K = 8$ and $d = 1$: GSP/SP contamination case ($K > T$)
- $K = 6$ and $d = 1$: full pilots case ($T < K + d$ and $T = K$)
- $K = 4$ and $d = 2$: CC processing case ($M < K \leq ML$) or RP contamination case ($T_p < K$)
- $K = 2$ and $d = 4$: local processing case ($M \geq K$)

Simulations are averaged over 12500 runs. For the RPs, the training time $T_p = \lceil 0.25T \rceil = 2$ is set to be 25% of coherence time slots. Transmit power is set to be $P_0 = 25$ dB, $\rho = 0.9$, and the noise power is $\sigma^2 = 1$. Note that the theoretical results for the GSP scheme have been verified with simulations. However, for the simplicity of the presentation, analytical results are omitted in the following figures.

B. Channel estimation

Figure 3 plots the MSE for the channel estimates with respect to the pilot power fraction (λ) for three pilot schemes with different cases of K and d. It can be observed that MSE changes in RP with λ are negligible, since $K \geq 2 = T_p$ causes pilot contamination, that is, the pilot reuse (when $T_p < K$) affects the strength of estimates than the SNR. It can be seen that the MSE of conventional SP scheme is also higher than that of the GSP scheme. However, as the number of users are increased in the system, the gap between them disappears due to larger strength of pilot contamination. For $K = 2$ and

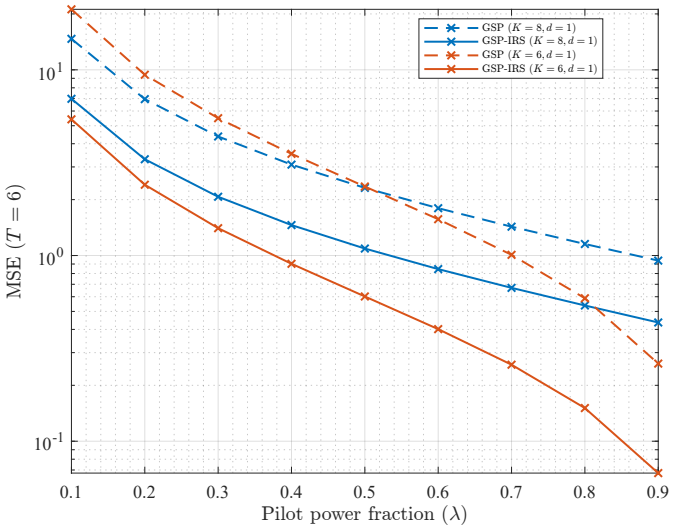

Figure 4. Averaged channel estimation MSE $\frac{1}{KL}\sum_{lk}\mathbb{E}\|\Delta_{lk}\|_2^2$ versus the pilot power allocation factor λ for SP scheme with and without IRSs.

$K = 4$, the GSP can be seen to provide much less MSE than that for the SP scheme.

Figure 4 plot the channel estimation MSE for the cases of presence and absence of IRSs in the system for pilot reuse cases. It can be observed that the intelligent surfaces can improve the MSE significantly. At $\lambda = 0.5$, MSE is reduced by 74% for $K = 8$ and 53% for $K = 6$.

C. Interference powers and sum rates

For the localized processing, Figures 5 (a)-(c) plot the average powers of SI $\frac{1}{LK}\sum_{l,k}P_{lk,SI}$, CI $\frac{1}{LK}\sum_{l,k}P_{lk,CI}$, and the sum rates $\frac{1}{L}\sum_{l,k}R_{lk}$. Note that these performance measures are averaged over the number of APs in the system. In these figures, the RP scheme performs worse at lower values of $1-\lambda$, since $T_p < K$ provides poor channel estimates; and as the number of users (K) is increased, the performance gets worse.

- From Figure 5 (a), as the data power fraction is increased, the self interference increases, which is intuitive since the SI term for the k^{th} user occurs due to its own channel imperfections. Since the GSP provides better channel estimates for $K = 2$ and $K = 4$, the SI power is much lower than that of the conventional SP scheme. The gap between the GSP and the SP schemes gets lower, as the number of users or the number of data symbols is increased.
- Figure 5 (b) shows the convex behavior for the CI power, that is, there exists an optimal value of λ, which can minimize the strength of CI. This convex behavior arises due to the data symbol terms in the channel estimation error. The more the data symbols in the channel estimates, the more the CI power variations with λ. The gap between the GSP and SP schemes decreases for an increase in K, similar to the SI power. For $K = 2$ and $K = 4$, the CI values for the GSP scheme are approximately the same, which is due to the low-complexity least squares based processing for data symbols (rather than zero-forcing).
- Figure 5 (c) reflects the concave behavior of the averaged sum rates with respect to λ, which shows that there is an optimum value of λ for sum rate maximization. This optimum point shifts to the right as the number of users are increases, which is due to the fact that the existence of more users provides more data symbols in the system, and thus, to get higher rates, the power fraction for the data symbols should be increased.

For centralized processing, the power of SI & CI and the sum rate is plotted in Figures 6 (a)-(c) respectively. It can be seen that due to the accumulated signals at the CPU from APs, the powers of SI and CI is higher than that in case of localized processing. Regarding the trend with respect to K and d for different schemes, similar inferences can be drawn as above for the Figures 5 (a)-(c). In addition to that, it can be seen that for CI power, the λ-value for minimum CI shifts to left as compared to Figure 5 (b), which is due to the reason that CI term consists of interfering terms, and at the CPU, more observations are present to estimate better. Thus, to minimize CI power, less power $(1-\lambda)$ is required for CPU based processing.

D. Bit error rates

Figure 7 plots the bit error rate of QPSK symbols at CPU and local estimates, averaged across all users. It can be seen that the CPU processing improves the bit rates. The error rate is a convex function of data transmission power, that is, there exists a value of λ that can minimize the bit error rate.

VI. Conclusion

In this paper, for an IRS-adied CF-mMIMO system, we have proposed and analyzed generalized superimposed training scheme with low complexity processing. For the channel estimation MSE and sum rates components for localized and centralized scenarios, we obtained the necessary conditions to avoid pilot contamination and successful data detection. Further, we have discussed with other cooperative scenarios and compared with regular pilot scheme and conventional superimposed scheme. Simulation results show the superior performance of proposed superimposed scheme both in terms of channel estimation and data detection. It also demonstrate significant reduction in interference components, when centralized processing is used.

As a part of future work, we will investigate iterative procedure to further improve the bit error rates and sum rates. For distantly located users, power allocation can also be optimized, along with OFDM based channel estimation incorporating interpolation.

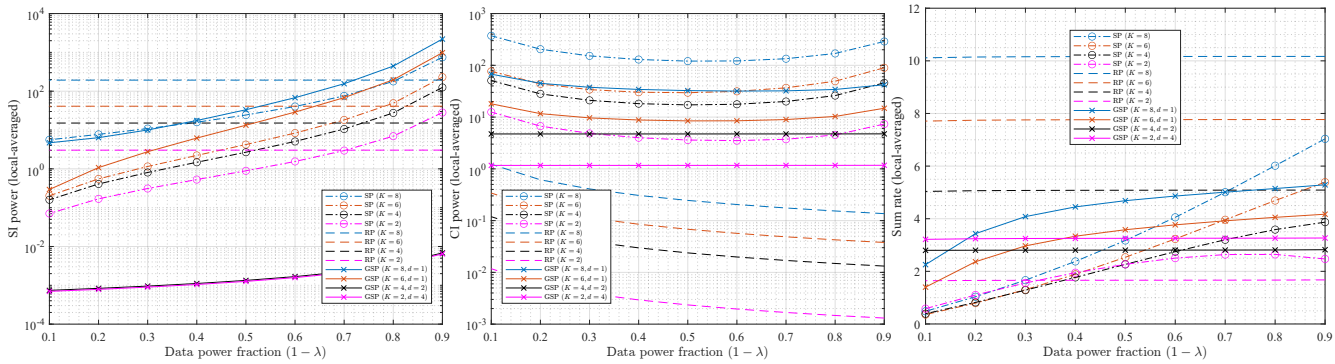

Figure 5. For the localized processing, the powers of SI, CI and sum rate versus the data power fraction $(1-\lambda)$ for SP (dashed-dot lines), RP (dashed lines) and GSP (solid lines) schemes with users $K=2$ (blue), 4 (red), 6 (black), 8 (magenta).

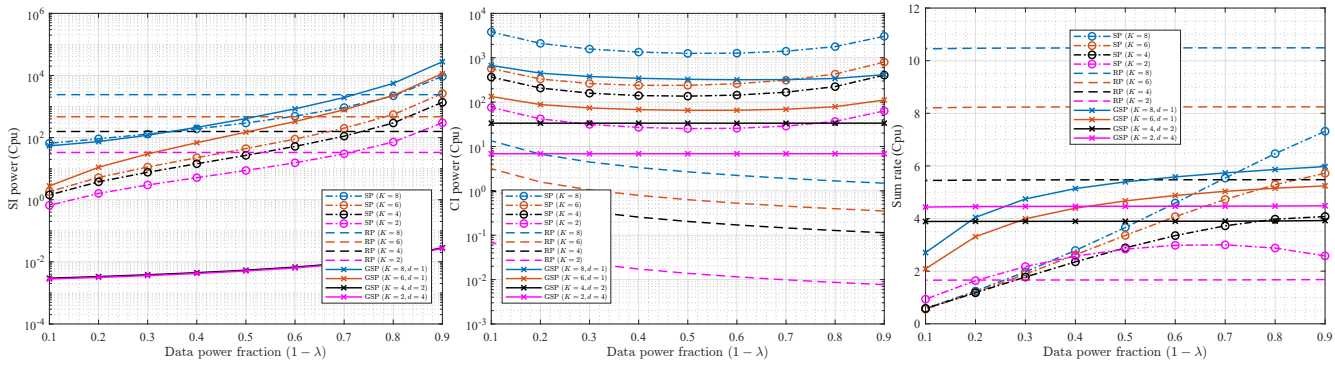

Figure 6. For the centralized processing, the powers of SI, CI and sum rate versus the data power fraction $(1-\lambda)$ for SP (dashed-dot lines), RP (dashed lines) and GSP (solid lines) schemes with users $K=2$ (blue), 4 (red), 6 (black), 8 (magenta).

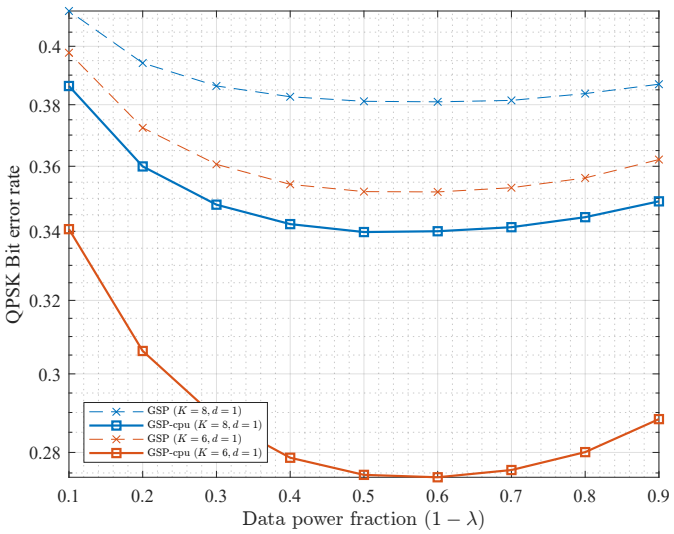

Figure 7. For the local and centralized processing, the average QPSK bit error rates of data estimates versus the data power fraction $(1-\lambda)$.

A. Cross covariance lemma

$$\mathbb{E}\left\{\Delta_{lk}\Delta_{mj}^H\right\} \overset{(a)}{=} \sum_{j_1\in\mathcal{C}_{\underline{k}}\setminus\{k\},j_2\in\mathcal{C}_{\underline{j}}\setminus\{j\}} \mathbb{E}\mathbf{h}_{lj_1}\mathbf{h}_{mj_2}^H + \frac{1-\lambda}{\lambda d}\sum_{j_1\in\mathcal{Z}_k}\mathbb{E}\mathbf{h}_{lj_1}\sum_{j_2\in\mathcal{Z}_j}\mathbf{h}_{mj_2}^H\delta_{j_1j_2}\delta_{kj}+\delta_{lm}\delta_{kj}\mathbb{E}\mathbf{w}_{lk}\mathbf{w}_{mj}^H$$
$$= \sum \left(\bar{\mathbf{h}}_{lj_1}\bar{\mathbf{h}}_{mj_2}^H+\delta_{lm}\delta_{j_1j_2}\mathbf{C}_{lj_1}^{[\mathbf{h}]}\right)$$

where $\mathbb{E}s_{(jk)}^*=0$ is used in (a); the expression of $\mathbf{C}_{lk,mk}^{[\Delta]}$ is given in (21).

B. Fourth norm of the Rician channel

Proof: If $\mathbf{x}\sim\mathcal{CN}(0,\mathbf{I})$, we derive

$$tr\mathbb{E}\mathbf{A}\mathbf{x}\mathbf{x}^H\mathbf{B}\mathbf{x}\mathbf{x}^H = \sum_{ijkl}\mathbb{E}A_{ij}x_jx_k^*B_{kl}x_lx_i^* \tag{47a}$$
$$= \sum_{i=j=k=l}+\sum_{i=j,k=l,i\neq k}+\sum_{i=l,j=k,i\neq j}\mathbb{E}A_{ij}x_jx_k^*B_{kl}x_lx_i^*$$
$$= \sum_i A_{ii}B_{ii}2+\sum_{i\neq k}A_{ii}B_{kk}+\sum_{i\neq j}A_{ij}B_{ji} \tag{47b}$$
$$= \sum_{i,k}A_{ii}B_{kk}+\sum_{i,j}A_{ij}B_{ji} \tag{47c}$$
$$= tr(\mathbf{A})tr(\mathbf{B})+tr(\mathbf{AB}), \tag{47d}$$

where $|x_i|^2\sim\frac{1}{2}\chi_2^2(0)$ and $\mathbb{E}|x_i|^4=2$. For $\mathbf{x}\sim\mathcal{CN}(\mathbf{0},\mathbf{C})$, we get

$$\mathbb{E}\|\mathbf{x}\|^4 = tr\mathbb{E}\mathbf{x}\mathbf{x}^H\mathbf{x}\mathbf{x}^H \tag{48a}$$
$$= tr\mathbb{E}\mathbf{C}\mathbf{C}^{-1/2}\mathbf{x}\mathbf{x}^H\mathbf{C}^{-1/2}\mathbf{C}\mathbf{C}^{-1/2}\mathbf{x}\mathbf{x}^H\mathbf{C}^{-1/2} \tag{48b}$$
$$= tr\mathbb{E}\mathbf{C}\tilde{\mathbf{x}}\tilde{\mathbf{x}}^H\mathbf{C}\tilde{\mathbf{x}}\tilde{\mathbf{x}}^H = tr(\mathbf{C})^2+tr(\mathbf{C}^2), \tag{48c}$$

where $\tilde{\mathbf{x}}=\mathbf{C}^{-1/2}\mathbf{x}\sim\mathcal{CN}(0,\mathbf{I})$.

For $\mathbf{h}=\bar{\mathbf{h}}+\tilde{\mathbf{h}}$ with $\bar{\mathbf{h}}$ being a constant and $\tilde{\mathbf{h}}\sim\mathcal{CN}(\mathbf{0},\mathbf{C})$, the forth order moment can be obtained as

$$\mathbb{E}\left|\mathbf{h}^H\mathbf{h}\right|^2=\mathbb{E}\left|\bar{\mathbf{h}}^H\bar{\mathbf{h}}+\tilde{\mathbf{h}}^H\bar{\mathbf{h}}+\bar{\mathbf{h}}^H\tilde{\mathbf{h}}+\tilde{\mathbf{h}}^H\tilde{\mathbf{h}}\right|^2 \tag{49a}$$

$$=\left|\bar{\mathbf{h}}^H\bar{\mathbf{h}}\right|^2+\mathbb{E}\left|\tilde{\mathbf{h}}^H\bar{\mathbf{h}}\right|^2+\mathbb{E}\left|\bar{\mathbf{h}}^H\tilde{\mathbf{h}}\right|^2+\mathbb{E}\left|\tilde{\mathbf{h}}^H\tilde{\mathbf{h}}\right|^2$$
$$+\cancel{2\mathbb{E}\Re\tilde{\mathbf{h}}^H\bar{\mathbf{h}}\bar{\mathbf{h}}^H\tilde{\mathbf{h}}}+\cancel{2\mathbb{E}\Re\bar{\mathbf{h}}^H\tilde{\mathbf{h}}\bar{\mathbf{h}}^H\tilde{\mathbf{h}}}+2\mathbb{E}\bar{\mathbf{h}}^H\bar{\mathbf{h}}\tilde{\mathbf{h}}^H\tilde{\mathbf{h}}$$
$$+\cancel{2\mathbb{E}\Re\tilde{\mathbf{h}}^H\bar{\mathbf{h}}\tilde{\mathbf{h}}^H\tilde{\mathbf{h}}}+\cancel{2\mathbb{E}\Re\bar{\mathbf{h}}^H\tilde{\mathbf{h}}\tilde{\mathbf{h}}^H\tilde{\mathbf{h}}}+\cancel{2\mathbb{E}\Re\tilde{\mathbf{h}}^H\tilde{\mathbf{h}}\tilde{\mathbf{h}}^H\tilde{\mathbf{h}}} \tag{49b}$$

$$=\left|\bar{\mathbf{h}}^H\bar{\mathbf{h}}\right|^2+2\bar{\mathbf{h}}^H\mathbb{E}\tilde{\mathbf{h}}\tilde{\mathbf{h}}^H\bar{\mathbf{h}}+\mathbb{E}\left|\tilde{\mathbf{h}}^H\tilde{\mathbf{h}}\right|^2+2\bar{\mathbf{h}}^H\bar{\mathbf{h}}tr\mathbf{C} \tag{49c}$$

$$=\left|\bar{\mathbf{h}}^H\bar{\mathbf{h}}\right|^2+2\bar{\mathbf{h}}^H\mathbf{C}\bar{\mathbf{h}}+2\bar{\mathbf{h}}^H\bar{\mathbf{h}}tr\mathbf{C}+tr(\mathbf{C})^2+tr(\mathbf{C}^2)$$

$$=\left(\bar{\mathbf{h}}^H\bar{\mathbf{h}}+tr\mathbf{C}\right)^2+2\bar{\mathbf{h}}^H\mathbf{C}\bar{\mathbf{h}}+tr(\mathbf{C}^2) \tag{49d}$$

$$=\left(\bar{\mathbf{h}}^H\bar{\mathbf{h}}+tr\mathbf{C}\right)^2+2tr\mathbf{C}\left(\bar{\mathbf{h}}\bar{\mathbf{h}}^H+\mathbf{C}\right)-tr(\mathbf{C}^2) \tag{49e}$$

$$=tr(\mathbf{R})^2+2tr(\mathbf{CR})-tr(\mathbf{C}^2)\coloneqq f_4(\bar{\mathbf{h}},\mathbf{C}), \tag{49f}$$

where $\mathbf{R}=\bar{\mathbf{x}}\bar{\mathbf{x}}^H+\mathbf{C}$ and the terms are canceled due to zero mean of $\tilde{\mathbf{h}}$. ■

C. IRS optimization lemma

$$\arg\min_{|\phi_{iq}|=1,\forall i,q}\sum_{lk}tr\mathbf{C}_{lk,lk}^{[\Delta]} \tag{50}$$

$$\overset{(a)}{=}\arg\min_{|\phi_{iq}|=1,\forall i,q}\sum_{lk}\sum_{i\in\mathcal{Z}_k,lk}tr\bar{\mathbf{h}}_{li}\bar{\mathbf{h}}_{li}^H \tag{51}$$

$$\overset{(b)}{=}\arg\min_{|\phi_{iq}|=1,\forall i,q}\boldsymbol{\phi}^H\mathbf{U}_\phi\boldsymbol{\phi}+2\Re\mathbf{c}_\phi^H\boldsymbol{\phi} \tag{52}$$

$$\overset{(d)}{=}-\mathbf{U}_\phi^{-1}\mathbf{c}_\phi\odot\frac{1}{\left|\mathbf{U}_\phi^{-1}\mathbf{c}_\phi\right|}, \tag{53}$$

where in (a), $\mathbf{C}_{lk,lk}^{[\Delta]}$ is substituted and simplified since $\mathbf{C}_{li}^{[\mathbf{h}]}$ does not contain $\boldsymbol{\phi}$; in (b), $\bar{\mathbf{h}}_{li}$ value is used and simplified for $\mathbf{U}_\phi=\sum_{lk,j\in\mathcal{Z}_k}\bar{\mathbf{U}}_{lj}^H\bar{\mathbf{U}}_{lj}$ and $\mathbf{c}_\phi^H=\sum_{lk,j\in\mathcal{Z}_k}\sqrt{\frac{\beta_{g,lj}G_{lj}}{G_{lj}^+}}\bar{\mathbf{g}}_{lj}^H\bar{\mathbf{U}}_{lj}$ with $j\in\mathcal{Z}_k$, $l=1,\ldots,L$ and $k=1,\ldots,K$; (d) is obtained by solving via differentiation.

D. Localized processing with GSP symbols

1) Signal power : The signal power can be derived as

$$P_{lk,S}=\mathbb{E}\left\|\|\hat{\mathbf{h}}_{lk}\|_2^2\mathbf{s}_k\right\|_2^2=\mathbb{E}\|\hat{\mathbf{h}}_{lk}\|_2^4\cdot\mathbb{E}\left\|\mathbf{s}_k\right\|_2^2 \tag{54a}$$

$$=f_4(\bar{\mathbf{h}}_{lk}+\bar{\Delta}_{lk},\mathbf{C}_{lk,lk}^{[\hat{\mathbf{h}}]}). \tag{54b}$$

2) Power of self-interference : The power of self-interference component can be computed as

$$P_{lk,SI}=\mathbb{E}\left\|\mathbf{s}_k\left(\mathbf{h}_{lk}^H\hat{\mathbf{h}}_{lk}-\hat{\mathbf{h}}_{lk}^H\hat{\mathbf{h}}_{lk}\right)\right\|_2^2$$
$$\overset{(a)}{=}\mathbb{E}\left\|-\Delta_k^H\left(\mathbf{h}_{lk}+\Delta_{lk}\right)\mathbf{s}_k\right\|_2^2$$
$$\overset{(b)}{=}\mathbb{E}\left|\Delta_{lk}^H\left(\mathbf{h}_{lk}+\Delta_{lk}\right)\right|^2$$
$$\overset{(c)}{=}\mathbb{E}\left|\Delta_{lk}^H\mathbf{h}_{lk}\right|^2+2\Re\mathbb{E}\Delta_{lk}^H\mathbf{h}_{lk}\Delta_{lk}^H\Delta_{lk}+\mathbb{E}\left|\Delta_{lk}^H\Delta_{lk}\right|^2$$
$$\overset{(d)}{=}tr\left(\mathbf{R}_{lk,lk}^{[\mathbf{h}]}\mathbf{R}_{lk,lk}^{[\Delta]}\right)+2\Re\bar{\Delta}_{lk}^H\bar{\mathbf{h}}_{lk}tr\mathbf{R}_{lk,lk}^{[\Delta]}+f_4(\bar{\Delta}_{lk},\mathbf{C}_{lk,lk}^{[\Delta]}),$$

where in (a), $\hat{\mathbf{h}}_{lk}^H=\mathbf{h}_{lk}+\Delta_{lk}$ is substituted; in (b), $tr\mathbb{E}\mathbf{s}_k\mathbf{s}_k^H=1$; in (c), the square is expanded; in (d), $\mathbb{E}\left|\Delta_{lk}^H\mathbf{h}_{lk}\right|^2=\mathbb{E}\Delta_{lk}^H\mathbf{R}_{lk}^{[\mathbf{h}]}\Delta_{lk}=tr\left(\mathbf{R}_{lk,lk}^{[\mathbf{h}]}\mathbf{R}_{lk,lk}^{[\Delta]}\right)$.

3) Power of cross-interference: The power of CI term can be derived as

$$P_{lk,CI}=\frac{1}{1-\lambda}\mathbb{E}\left\|\sum_{i\neq k}\frac{\mathbf{Z}_k^H\mathbf{x}_i}{T\sqrt{P}}\mathbf{h}_{li}^H\hat{\mathbf{h}}_{lk}\right\|_2^2$$
$$=\frac{1}{1-\lambda}\sum_{i\neq k}\sum_{j\neq k}tr\mathbb{E}\left[\frac{\mathbf{Z}_k^H\mathbf{x}_i}{T\sqrt{P}}\frac{\mathbf{x}_j^H\mathbf{Z}_k}{T\sqrt{P}}\right]\mathbb{E}\left\{\hat{\mathbf{h}}_{lk}^H\mathbf{h}_{lj}\mathbf{h}_{li}^H\hat{\mathbf{h}}_{lk}\right\}$$
$$=\frac{M}{1-\lambda}\sum_{i\neq k}\sum_{j\neq k}tr\mathbf{R}_{k,ij}^{[\mathbf{Z}]}\times\frac{1}{M}tr\left[\mathbf{R}_{lj,li}^{[\mathbf{h}]}\mathbf{R}_{lk,lk}^{[\hat{\mathbf{h}}]}\right],$$

where $\mathbb{E}\hat{\mathbf{h}}_{lk}^H\mathbf{h}_{lj}\mathbf{h}_{li}^H\hat{\mathbf{h}}_{lk}=tr\left(\mathbb{E}\mathbf{h}_{lj}\mathbf{h}_{li}^H\cdot\mathbb{E}\hat{\mathbf{h}}_{lk}\hat{\mathbf{h}}_{lk}^H\right)=tr\left[\mathbf{R}_{lj,li}^{[\mathbf{h}]}\mathbf{R}_{lklk}^{[\hat{\mathbf{h}}]}\right]$.

4) Noise power: The noise power can be calculated as

$$P_{lk,N}\times T^2P(1-\lambda)=\mathbb{E}\left\|\mathbf{Z}_k^H\mathbf{W}_l^H\hat{\mathbf{h}}_{lk}\right\|_2^2$$
$$=\mathbb{E}\|\mathbf{Z}_k^H\mathbf{W}_l^H\left(\mathring{\Delta}_{lk}+\frac{\mathbf{W}_l\mathbf{p}_k}{T\sqrt{P\lambda}}\right)\|_2^2$$
$$=\mathbb{E}\|\mathbf{Z}_k^H\mathbf{W}_l^H\mathring{\Delta}_{lk}\|_2^2+\mathbb{E}\left\|\frac{\mathbf{Z}_k^H\mathbf{W}_l^H\mathbf{W}_l\mathbf{p}_k}{T\sqrt{P\lambda}}\right\|_2^2$$
$$=dT\sigma^2tr(\mathbf{R}_{lk,lk}^{[\mathring{\Delta}]})+\frac{\sigma^4MTd\left(T+M+1\right)}{T^2P\lambda}$$
$$=\sigma^2MTd\left[\frac{tr(\mathbf{R}_{lk,lk}^{[\mathring{\Delta}]})}{M}+\frac{\sigma^2}{PT\lambda}+\frac{\sigma^2\left(M+1\right)}{T^2P\lambda}\right]$$
$$=\sigma^2MTd\left[\frac{tr(\mathbf{R}_{lk,lk}^{[\hat{\mathbf{h}}]})}{M}+\frac{\sigma^2(M+1)}{T^2P\lambda}\right], \tag{57}$$

where for $\mathbf{R}_{lk,lk}^{[\mathring{\Delta}]}=\mathbb{E}\mathring{\Delta}_{lk}\mathring{\Delta}_{lk}^H$, the first term is simplified as

$$\mathbb{E}\left\|\mathbf{Z}_k^H\mathbf{W}_l^H\mathring{\Delta}_{lk}\right\|_2^2=tr\left[\mathbf{Z}_k^H\mathbb{E}\left\{\mathbf{W}_l^H\mathbf{R}_{lk}^{[\mathring{\Delta}]}\mathbf{W}_l\right\}\mathbf{Z}_k\right]$$
$$=tr(\mathbf{Z}_k^H\mathbf{Z}_k)\sigma^2tr(\mathbf{R}_{lk,lk}^{[\mathring{\Delta}]}),$$

followed by the the second term's simplification for (say) $\mathbf{W}_l = [\bar{\mathbf{w}}_1, \ldots, \bar{\mathbf{w}}_T]$ and $\left[\mathbf{W}_l^H \mathbf{W}_l\right]_{i,j} = \bar{\mathbf{w}}_i^H \bar{\mathbf{w}}_j$ as

$$
\begin{aligned}
&\mathbb{E}\left\|\mathbf{Z}_k^H \mathbf{W}_l^H \mathbf{W}_l \mathbf{p}_k\right\|_2^2 = tr\mathbb{E}\left[\mathbf{W}_l^H \mathbf{W}_l \mathbf{Z}_k \mathbf{Z}_k^H \mathbf{W}_l^H \mathbf{W}_l \mathbf{p}_k \mathbf{p}_k^H\right] \\
&= \sum_{i,j,k,l} \mathbb{E}\left[\left[\mathbf{W}_l^H \mathbf{W}_l\right]_{i,j}\left[\mathbf{Z}_k \mathbf{Z}_k^H\right]_{j,k}\left[\mathbf{W}_l^H \mathbf{W}_l\right]_{k,l}\left[\mathbf{p}_k \mathbf{p}_k^H\right]_{l,i}\right] \\
&= \sum_{i,j,k,l} \mathbb{E}\left[\bar{\mathbf{w}}_i^H \bar{\mathbf{w}}_j \left[\mathbf{Z}_k \mathbf{Z}_k^H\right]_{j,k} \bar{\mathbf{w}}_k^H \bar{\mathbf{w}}_l \left[\mathbf{p}_k \mathbf{p}_k^H\right]_{l,i}\right] \\
&= \sum_{i=j,k=l,i\neq k} \mathbb{E}\left[\bar{\mathbf{w}}_i^H \bar{\mathbf{w}}_i \left[\mathbf{Z}_k \mathbf{Z}_k^H\right]_{i,l} \bar{\mathbf{w}}_l^H \bar{\mathbf{w}}_l \left[\mathbf{p}_k \mathbf{p}_k^H\right]_{l,i}\right] \\
&\quad + \sum_{j=k,i=l,l\neq k} \mathbb{E}\left[\bar{\mathbf{w}}_i^H \bar{\mathbf{w}}_j \bar{\mathbf{w}}_j^H \bar{\mathbf{w}}_i \left[\mathbf{Z}_k \mathbf{Z}_k^H\right]_{j,j}\left[\mathbf{p}_k \mathbf{p}_k^H\right]_{i,i}\right] \\
&\quad + \sum_{i=j=k=l} \mathbb{E}\left[\bar{\mathbf{w}}_i^H \bar{\mathbf{w}}_i \bar{\mathbf{w}}_i^H \bar{\mathbf{w}}_i \left[\mathbf{Z}_k \mathbf{Z}_k^H\right]_{i,i}\left[\mathbf{p}_k \mathbf{p}_k^H\right]_{i,i}\right] \\
&= \sigma^4 M^2 \sum_{i,l} \left[\mathbf{Z}_k \mathbf{Z}_k^H\right]_{i,l}\left[\mathbf{p}_k \mathbf{p}_k^H\right]_{l,i} \\
&\quad + \sigma^4 M \sum_{i,j} \left[\mathbf{Z}_k \mathbf{Z}_k^H\right]_{j,j}\left[\mathbf{p}_k \mathbf{p}_k^H\right]_{i,i} \\
&\quad + \sigma^4 (M^2+M) \sum_{i} \left[\mathbf{Z}_k \mathbf{Z}_k^H\right]_{i,i}\left[\mathbf{p}_k \mathbf{p}_k^H\right]_{i,i} \\
&= \sigma^4 M^2 tr\left(\mathbf{Z}_k \mathbf{Z}_k^H \mathbf{p}_k \mathbf{p}_k^H\right) + \sigma^4 M tr\left[\mathbf{Z}_k \mathbf{Z}_k^H\right] tr\left[\mathbf{p}_k \mathbf{p}_k^H\right] \\
&\quad + \sigma^4 (M^2+M) tr\left[\mathbf{Z}_k \mathbf{Z}_k^H \underbrace{\mathcal{D}\left(\mathbf{p}_k \mathbf{p}_k^H\right)}_{=\mathbf{I}_T}\right] \\
&= \sigma^4 M \cdot Td \cdot T + \sigma^4 (M^2+M) \cdot Td \\
&= \sigma^4 M \cdot Td\,(T+M+1),
\end{aligned}
$$

and $\frac{tr(\mathbf{R}_{lk}^{[\hat{\mathbf{h}}]})}{M} = \frac{tr(\mathbf{R}_{lk}^{[\mathring{\Delta}]})}{M} + \frac{\sigma^2}{PT\lambda}$.

E. Centralized processing with GSP symbols

1) Signal power: The signal power can be derived as

$$
\begin{aligned}
&\mathbb{E}\left\|\|\hat{\mathbf{h}}_k\|_2^2 \mathbf{s}_k\right\|_2^2 = \mathbb{E}\|\hat{\mathbf{h}}_k\|_2^4 \cdot \mathbb{E}\left\|\mathbf{s}_k\right\|_2^2 = \mathbb{E}\|\hat{\mathbf{h}}_k\|_2^4 \\
&= f_4\left(\bar{\mathbf{h}}_k + \bar{\Delta}_k, \mathbf{C}_k^{[\hat{\mathbf{h}}]}\right).
\end{aligned}
$$

2) Power of self-interference term: The power of self-interference component can be computed as

$$
\begin{aligned}
&P_{k,SI} = \mathbb{E}\left\|\mathbf{s}_k\left(\mathbf{h}_k^H \hat{\mathbf{h}}_k - \hat{\mathbf{h}}_k^H \hat{\mathbf{h}}_k\right)\right\|_2^2 \overset{(a)}{=} \mathbb{E}\left|\Delta_k^H\left(\mathbf{h}_k + \Delta_k\right)\right|^2 \\
&= \sum_{l=1}^{L}\sum_{m=1}^{L} \mathbb{E}\Delta_{lk}^H\left(\mathbf{h}_{lk} + \Delta_{lk}\right)\left(\mathbf{h}_{mk} + \Delta_{mk}\right)^H \Delta_{mk} \\
&= \sum_{l=1}^{L}\sum_{m=1}^{L} \mathbb{E}\Delta_{lk}^H\left(\mathbf{h}_{lk}\mathbf{h}_{mk}^H + 2\Re\Delta_{lk}\mathbf{h}_{mk}^H + \Delta_{lk}\Delta_{mk}^H\right)\Delta_{mk} \\
&\overset{(b)}{=} \sum_{l,m} tr\left(\mathbf{R}_{lk,mk}^{[\mathbf{h}]}\mathbf{R}_{mk,lk}^{[\Delta]}\right) + 2\Re\sum_{l=1}^{L}\bar{\Delta}_{lk}^H\bar{\mathbf{h}}_{lk}\sum_{m=1}^{L} tr\left(\mathbf{R}_{mk,mk}^{[\Delta]}\right) \\
&+ f_4\left(\bar{\Delta}_k, \mathbf{C}_k^{[\Delta]}\right),
\end{aligned}
$$

where in (a), $\hat{\mathbf{h}}_k^H = \mathbf{h}_k + \Delta_k$ and $tr\mathbb{E}\mathbf{s}_k\mathbf{s}_k^H = 1$ is used; in (b), the following values are used;

$$
\begin{aligned}
\mathbb{E}\Delta_{lk}^H\mathbf{h}_{lk}\mathbf{h}_{mk}^H\Delta_{mk} &= tr\left(\mathbf{R}_{lk,mk}^{[\mathbf{h}]}\mathbf{R}_{mk,lk}^{[\Delta]}\right) \\
\mathbb{E}\Delta_{lk}^H\mathbf{h}_{lk}\Delta_{mk}^H\Delta_{mk} &= \bar{\Delta}_{lk}^H\bar{\mathbf{h}}_{lk} tr\left(\mathbf{R}_{mk,mk}^{[\Delta]}\right).
\end{aligned}
$$

3) Power of cross-interference term: The power of CI term can be derived as

$$
\begin{aligned}
&P_{k,CI} = \frac{1}{1-\lambda}\mathbb{E}\left\|\sum_{i\neq k}\frac{\mathbf{Z}_k^H\mathbf{x}_i}{T\sqrt{P}}\mathbf{h}_i^H\hat{\mathbf{h}}_k\right\|_2^2 \\
&= \frac{1}{1-\lambda}\sum_{i\neq k}\sum_{j\neq k} tr\mathbb{E}\left[\frac{\mathbf{Z}_k^H\mathbf{x}_i}{T\sqrt{P}}\frac{\mathbf{x}_j^H\mathbf{Z}_k}{T\sqrt{P}}\right]\mathbb{E}\left\{\hat{\mathbf{h}}_k^H\mathbf{h}_j\mathbf{h}_i^H\hat{\mathbf{h}}_k\right\} \\
&\overset{(a)}{=} \frac{ML^2}{1-\lambda}\sum_{i\neq k}\sum_{j\neq k} tr\mathbf{R}_{k,ij}^{[\mathbf{Z}]} \times \frac{1}{ML^2}\sum_{l,m} tr\left[\mathbf{R}_{lj,mi}^{[\mathbf{h}]}\mathbf{R}_{mk,lk}^{[\hat{\mathbf{h}}]}\right],
\end{aligned}
$$

where $\mathbb{E}\hat{\mathbf{h}}_k^H\mathbf{h}_j\mathbf{h}_i^H\hat{\mathbf{h}}_k = \sum_{l,m} tr\left(\mathbb{E}\mathbf{h}_{lj}\mathbf{h}_{mi}^H \cdot \mathbb{E}\hat{\mathbf{h}}_{mk}\hat{\mathbf{h}}_{lk}^H\right) = \sum_{l,m} tr\left[\mathbf{R}_{lj,mi}^{[\mathbf{h}]}\mathbf{R}_{mk,lk}^{[\hat{\mathbf{h}}]}\right]$.

4) Noise power: The noise power can be calculated as

$$
\begin{aligned}
&P_{k,N} \times T^2 P(1-\lambda) = \mathbb{E}\left\|\mathbf{Z}_k^H\mathbf{W}^H\hat{\mathbf{h}}_k\right\|_2^2 \\
&= \mathbb{E}\|\mathbf{Z}_k^H\mathbf{W}^H\left(\mathring{\Delta}_k + \frac{\mathbf{W}\mathbf{p}_k}{T\sqrt{P\lambda}}\right)\|_2^2 \qquad (58) \\
&= \mathbb{E}\|\mathbf{Z}_k^H\mathbf{W}_l^H\mathring{\Delta}_k\|_2^2 + \mathbb{E}\left\|\frac{\mathbf{Z}_k^H\mathbf{W}^H\mathbf{W}\mathbf{p}_k}{T\sqrt{P\lambda}}\right\|_2^2 \qquad (59) \\
&= dT\sigma^2\sum_l tr(\mathbf{R}_{lk,lk}^{[\mathring{\Delta}]}) + \frac{1}{T^2P\lambda}\sigma^4 MLTd\,(T+ML+1) \\
&\sigma^2 MLTd\left[\sum_l\frac{tr(\mathbf{R}_{lk,lk}^{[\mathring{\Delta}]})}{ML} + \frac{\sigma^2}{PT\lambda} + \frac{\sigma^2\,(ML+1)}{T^2P\lambda}\right] \qquad (60) \\
&= \sigma^2 MLTd\left[\sum_l\frac{tr(\mathbf{R}_{lk,lk}^{[\hat{\mathbf{h}}]})}{ML} + \frac{\sigma^2(ML+1)}{T^2P\lambda}\right], \qquad (61)
\end{aligned}
$$

where for $\mathbf{R}_{lk}^{[\mathring{\Delta}]} = \mathbb{E}\mathring{\Delta}_{lk}\mathring{\Delta}_{lk}^H$, the steps of simplification are similar to the Appendix-D4.

REFERENCES

[1] J. Zhang, S. Chen, Y. Lin, J. Zheng, B. Ai, and L. Hanzo, "Cell-free massive mimo: A new next-generation paradigm," *IEEE Access*, vol. 7, pp. 99 878–99 888, 2019.

[2] S. Buzzi and C. D'Andrea, "Cell-free massive mimo: User-centric approach," *IEEE Wireless Communications Letters*, vol. 6, no. 6, pp. 706–709, 2017.

[3] O. T. Demir, E. Bjornson, and L. Sanguinetti, "Foundations of user-centric cell-free massive mimo," *Foundations and TrendsÂ® in Signal Processing*, vol. 14, no. 3-4, pp. 162–472, 2021. [Online]. Available: http://dx.doi.org/10.1561/2000000109

[4] Y. Zhang, X. Qiao, L. Yang, and H. Zhu, "Superimposed pilots are beneficial for mitigating pilot contamination in cell-free massive mimo," *IEEE Communications Letters*, vol. 25, no. 1, pp. 279–283, 2021.

[5] E. BjÃ¶rnson and L. Sanguinetti, "Making cell-free massive mimo competitive with mmse processing and centralized implementation," *IEEE Transactions on Wireless Communications*, vol. 19, no. 1, pp. 77–90, 2020.

[6] ——, "Scalable cell-free massive mimo systems," *IEEE Transactions on Communications*, vol. 68, no. 7, pp. 4247–4261, 2020.

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

[8] H. Liu, J. Zhang, X. Zhang, A. Kurniawan, T. Juhana, and B. Ai, "Tabu-search-based pilot assignment for cell-free massive mimo systems," *IEEE Transactions on Vehicular Technology*, vol. 69, no. 2, pp. 2286–2290, 2020.

[9] Y. Zhang, H. Cao, P. Zhong, C. Qi, and L. Yang, "Location-based greedy pilot assignment for cell-free massive mimo systems," in *2018 IEEE 4th International Conference on Computer and Communications (ICCC)*, 2018, pp. 392–396.

[10] H. Liu, J. Zhang, S. Jin, and B. Ai, "Graph coloring based pilot assignment for cell-free massive mimo systems," *IEEE Transactions on Vehicular Technology*, vol. 69, no. 8, pp. 9180–9184, 2020.

[11] Y. Jin, J. Zhang, S. Jin, and B. Ai, "Channel estimation for cell-free mmwave massive mimo through deep learning," *IEEE Transactions on Vehicular Technology*, vol. 68, no. 10, pp. 10 325–10 329, 2019.

[12] T. C. Mai, H. Q. Ngo, M. Egan, and T. Q. Duong, "Pilot power control for cell-free massive mimo," *IEEE Transactions on Vehicular Technology*, vol. 67, no. 11, pp. 11 264–11 268, 2018.

[13] J. Zhang, Y. Wei, E. BjÃ¶rnson, Y. Han, and S. Jin, "Performance analysis and power control of cell-free massive mimo systems with hardware impairments," *IEEE Access*, vol. 6, pp. 55 302–55 314, 2018.

[14] M. Zhou, Y. Zhang, X. Qiao, and L. Yang, "Spatially correlated rayleigh fading for cell-free massive mimo systems," *IEEE Access*, vol. 8, pp. 42 154–42 168, 2020.

[15] S. Chakraborty, O. T. Demir, E. Bjornson, and P. Giselsson, "Efficient downlink power allocation algorithms for cell-free massive mimo systems," *IEEE Open Journal of the Communications Society*, vol. 2, pp. 168–186, 2021.

[16] M. Alonzo, S. Buzzi, A. Zappone, and C. D'Elia, "Energy-efficient power control in cell-free and user-centric massive mimo at millimeter wave," *IEEE Transactions on Green Communications and Networking*, vol. 3, no. 3, pp. 651–663, 2019.

[17] O. Ozdogan, E. BjÃ¶rnson, and J. Zhang, "Performance of cell-free massive mimo with rician fading and phase shifts," *IEEE Transactions on Wireless Communications*, vol. 18, no. 11, pp. 5299–5315, 2019.

[18] 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 *2018 52nd Asilomar Conference on Signals, Systems, and Computers*, 2018, pp. 980–984.

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

[20] J. C. Estrada-JimÃ©nez, B. G. GuzmÃ¡n, M. J. FernÃ¡ndez-Getino GarcÃa, and V. P. G. JimÃ©nez, "Superimposed training-based channel estimation for miso optical-ofdm vlc," *IEEE Transactions on Vehicular Technology*, vol. 68, no. 6, pp. 6161–6166, June 2019.

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

[22] 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 Transactions on Signal Processing*, vol. 60, no. 1, pp. 274–284, Jan 2012.

[23] 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 Transactions on Vehicular Technology*, vol. 65, no. 3, pp. 1492–1505, March 2016.

[24] K.-C. Chan, W.-C. Huang, C.-P. Li, and H.-J. Li, "Elimination of data identification problem for data-dependent superimposed training," *IEEE Transactions on Signal Processing*, vol. 63, no. 6, pp. 1595–1604, March 2015.

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

[26] N. Garg, A. K. Jagannatham, G. Sharma, and T. Ratnarajah, "Precoder feedback schemes for robust interference alignment with bounded csi uncertainty," *IEEE Transactions on Signal and Information Processing over Networks*, vol. 6, pp. 407–425, 2020.

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

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

[29] A. Vosoughi and A. Scaglione, "Everything you always wanted to know about training: guidelines derived using the affine precoding framework and the crb," *IEEE Transactions on Signal Processing*, vol. 54, no. 3, pp. 940–954, March 2006.

[30] M. Ghogho, T. Whitworth, A. Swami, and D. McLernon, "Full-rank and rank-deficient precoding schemes for single-carrier block transmissions," *IEEE Transactions on Signal Processing*, vol. 57, no. 11, pp. 4433–4442, Nov 2009.

[31] J. C. Estrada-JimÃ©nez and M. J. FernÃ¡ndez-Getino GarcÃa, "Partial-data superimposed training with data precoding for ofdm systems," *IEEE Transactions on Broadcasting*, vol. 65, no. 2, pp. 234–244, June 2019.

[32] Q.-U.-A. Nadeem, H. Alwazani, A. Kammoun, A. Chaaban, M. Debbah, and M.-S. Alouini, "Intelligent reflecting surface-assisted multi-user miso communication: Channel estimation and beamforming design," *IEEE Open Journal of the Communications Society*, vol. 1, pp. 661–680, 2020.

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

[34] E. Bjornson, O. Ozdogan, and E. G. Larsson, "Intelligent reflecting surface versus decode-and-forward: How large surfaces are needed to beat relaying?" *IEEE Wireless Communications Letters*, vol. 9, no. 2, pp. 244–248, 2020.

[35] S. Zeng, H. Zhang, B. Di, Y. Tan, Z. Han, H. V. Poor, and L. Song, "Reconfigurable intelligent surfaces in 6g: Reflective, transmissive, or both?" *IEEE Communications Letters*, vol. 25, no. 6, pp. 2063–2067, June 2021.

[36] H. Å iljak, N. Ashraf, M. T. Barros, D. P. Martins, B. Butler, A. Farhang, N. Marchetti, and S. Balasubramaniam, "Evolving intelligent reflector surface toward 6g for public health: Application in airborne virus detection," *IEEE Network*, vol. 35, no. 5, pp. 306–312, Sep. 2021.

[37] Q. Liu, S. Sun, B. Rong, and M. Kadoch, "Intelligent reflective surface based 6g communications for sustainable energy infrastructure," *IEEE Wireless Communications*, vol. 28, no. 6, pp. 49–55, December 2021.

[38] D. L. Galappaththige, D. Kudathanthirige, and G. Amarasuriya, "Performance analysis of distributed intelligent reflective surface aided communications," in *GLOBECOM 2020 - 2020 IEEE Global Communications Conference*, Dec 2020, pp. 1–6.

[39] T. Hou, Y. Liu, Z. Song, X. Sun, Y. Chen, and L. Hanzo, "Mimo assisted networks relying on intelligent reflective surfaces: A stochastic geometry based analysis," *IEEE Transactions on Vehicular Technology*, vol. 71, no. 1, pp. 571–582, Jan 2022.

[40] H. Xie, J. Xu, and Y.-F. Liu, "Max-min fairness in irs-aided multi-cell miso systems with joint transmit and reflective beamforming," *IEEE Transactions on Wireless Communications*, vol. 20, no. 2, pp. 1379–1393, Feb 2021.

[41] J.-S. Jung, C.-Y. Park, J.-H. Oh, and H.-K. Song, "Intelligent reflecting surface for spectral efficiency maximization in the multi-user miso communication systems," *IEEE Access*, vol. 9, pp. 134 695–134 702, 2021.

[42] H. Zhang, B. Di, Z. Han, H. V. Poor, and L. Song, "Reconfigurable intelligent surface assisted multi-user communications: How many reflective elements do we need?" *IEEE Wireless Communications Letters*, vol. 10, no. 5, pp. 1098–1102, May 2021.

[43] 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 *GLOBECOM 2020 - 2020 IEEE Global Communications Conference*, 2020, pp. 1–6.

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

[45] 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.

[46] Y. Zhang, B. Di, H. Zhang, J. Lin, C. Xu, D. Zhang, Y. Li, and L. Song, "Beyond cell-free mimo: Energy efficient reconfigurable intelligent surface aided cell-free mimo communications," *IEEE Transactions on Cognitive Communications and Networking*, vol. 7, no. 2, pp. 412–426, June 2021.

[47] N. Garg, M. Sellathurai, and T. Ratnarajah, "Low complexity joint omp methods for fdd channel estimation in massive mimo systems," in *IEEE 21st International Workshop on Signal Processing Advances in Wireless Communications (SPAWC)*, May 2020, pp. 1–5.