Generalized Superimposed Training Scheme In Cell-free Massive MIMO Systems

Generalized Superimposed Training Scheme In Cell-free Massive MIMO Systems

Navneet Garg, *Member, IEEE* and Tharmalingam Ratnarajah, *Senior Member, IEEE*.

***Abstract*—Regular pilots in a massive multi-input multi-output (MIMO) system with large number of users suffer from the pilot contamination effect due to limited training time. In this paper, for a cell-free massive MIMO system, we have proposed a generalized superimposed pilot (GSP) scheme, 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 and distributed time processing approach is employed, and the corresponding rate components are analyzed. It is shown that with careful design of precoding matrix 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 have been verified via simulations. It is also inferred that centralized processing in cell free system improves the data detection performance than localized processing. Iterative data detection at the central node also improves the MSE of data estimates. The pilot contamination effect, is significantly reduced due to availability of larger number of pilots, as compared to regular pilots transmission.**

***Index Terms*—cell-free; channel estimation; massive MIMO; pilot training; superimposed pilot.**

I. Introduction

In the recent years, cell-free massive multi-input multi-output (CF-mMIMO) systems have gained considerable attention, since it is able to achieve all merits of traditional distributed large-scale MIMO and network MIMO systems with simple linear decoding schemes [1]–[5]. These merits include tremendous macro-diversity and coverage ratio, high spectral and energy efficiencies, low interference and path fading [6], [7]. In such systems, spatial data detection and 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 [8]. 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 $L \to \infty$ [9], 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 [10], structured PA based on geographical locations [11], graph coloring based PA [12]. Dynamic cooperation clustering has also been used to reuse the pilots in an efficient way in [8]. In [13], sparse channel matrix estimates are improved using deep neural networks for mmWave systems. In [14], [15], pilot power control is proposed to reduce the pilot contamination using convex approximation approach, whereas the rate-optimized power allocation is considered in [16], [17] via geometric programming. Energy efficiency maximization for power control is investigated for mmWave system in [18].

Note that the above works employ regular pilots (RPs) transmission, where the pilot and data symbols are sent separately in the coherence interval ($T = T_p + T_d$), where T_p and T_d are the number of time slots used for pilot and data transmission. The value of T_p decides the number of available pilots in the system, which in turn decide the strength of pilot contamination. In contrast to RPs, the transmission of superimposed pilots (SPs) has been investigated to have interesting features in traditional large-scale MIMO systems [6]. In SP scheme, the 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 $T > T_p$. 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 [19]. In spite of the performance degradation, the above works indicate that the SP scheme outperform the RP scheme in terms of the achievable sum rate. Interestingly, to improve the SP scheme's performance, a generalized SP (GSP) framework has been proposed in our previous work [9], where instead of sending the same number of data symbols as the number of pilots (T), the number of data symbols is reduced and optimized, and further precoded to avoid the possible pilot contamination in both the channel estimation and the data detection. The GSP 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. In the CF-mMIMO system, the work on SP schemes is limited [6], to the best of author's knowledge. Therefore, in such a system, it will be interesting to compare the performances of RPs, SPs and general SPs.

Related work on precoded superimposed training: In [20], [21], precoder and pilots are optimized under the condition $T \geq K + d$ and found to be proportional to the sub-discrete Fourier transform (DFT) matrix. In addition to that, authors

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.

in [22] provide orthogonal precoder and pilot design using the Cramer-Rao lower bound of the channel estimation error, whereas they compare joint and individual data & channel estimations via mutual information bounds in [23]. Precoder and pilots should be orthogonal, is proved in [24] for frequency selective channels. Similar orthogonal precoder-pilot design is obtained via an iterative algorithm in [25]. Five criterion for data identification is given in [26], including proportionality to DFT matrix; same magnitude of entries; orthogonal columns; orthogonal to pilots; and periodic auto-correlation function of columns of precoder. Authors in [27], [28] select square precoder matrix (e.g., Hadamard), and analyzes the channel estimation performance.

Related work on iterative detection: In [29]–[31] with conventional SP scheme, alternative least squares for joint channel and data estimates is used for improvement and to reduce the computation time, while utilizing maximum likelihood based data detection (Viterbi algorithm). In an orthogonal-frequency-division-multiplexing (OFDM) based amplify-and-forward (AF) cooperative system with partial data-dependent superimposed training [32], an iterative method is obtained based on the variational inference approach by formulating a free-energy function. In [33] for an OFDM system with AF relay, after getting channel estimates obtained via MMSE, iterative demodulation and decoding is performed using turbo detection and variational Bayesian approximation. A similar iterative process based on generalized approximate message passing is used for data detection in underwater acoustics communications [34], [35].

It can be noted in above works that either channel estimates or the data detection parts have been exercised to be improved individually. Together as a whole including optimal pilot design and iterative detection has not been well investigated in literature, especially for CF-mMIMO system. Towards that, in this paper, we analyze orthogonal precoder-pilot design, followed by iterative estimation.

Relation to our previous work [9]: For the superimposed symbol, the pilot allocation in this work is the generalization of [9], which had considered a specific suboptimal pilot scheme. Moreover, in the channel estimation, pilot reuse was not considered, that is, only $0 \leq K \leq T$ case was analyzed, whereas the present work also includes the case $K > T$. In addition to that, there were no-connection to local and centralized processing for CF-mMIMO system, nor [9] includes details about iterative and distributed-time processing.

A. Contributions

In this paper, a CF-mMIMO system is considered with multiple users, where the access points (APs) are connected via a central processing unit (CPU) [7]. For the proposed GSP scheme, we consider two scenarios, viz., centralized and distributed processing, and analyze the mean-squared error (MSE) of channel estimation and sum rates. Further, the estimates are improved using iterative data detection scheme with distributed time processing. Note that these scenarios correspond to the cooperation levels $L4$ (full-cooperation) and $L1$ (no-cooperation) of [7]. For the SP/GSP scheme, the performance of levels $L2$ and $L3$ (relaxed cooperation versions of $L4$) 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 general SPs, as compared to RPs, and around 54% improvement in sum rates over conventional SPs for the case of $T < K + d$, where d is the number of data symbols for transmission. 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 in the conventional SP scheme, the GSP scheme is proposed, where any number (from 1 to $T - 1$) of data symbols can be used for reliable communication. Moreover, the data symbols are precoded, making the channel estimation and data detection less susceptible to pilot contanimations.

2) Analysis for CF-mMIMO system: For both the centralized and distributed 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) Iterative estimation: To avoid large computations at the end of coherence time block, distributed time processing approach is presented, where the received symbols can be processed as they arrive, and the final detection will be decided when all symbols are received. Further, to improve the data detection for low complexity matched filter processing, a iterative algorithm is presented, which is demonstrated to converge in a few iterations.

4) Comparison and simulations: For the same transmit power constraint as in the GSP scheme, the estimates of channels and data for both the centralized and distributed scenarios are obtained. In simulations with the verified analytical expressions, the different performance indicators show the improved performance of the GSP scheme for CF-mMIMO system.

B. Organization

Section II describes the details the cell-free system and the GSP scheme. The analysis for centralized and distributed processing for both channel estimation and data detection are given in Section III. Section IV provides the iterative estimation. Regular pilot and conventional superimposed pilot schemes are compared in Section V, followed by simulations and conclusion in Section VI and VII, respectively.

C. 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 and Hermitian transpose of a matrix are denoted by $(\cdot)^*$, $(\cdot)^T$ and $(\cdot)^H$, respectively. $\mathcal{CN}(\mu, \mathbf{R})$ represents a circularly symmetric complex Gaussian random vector with mean μ and covariance

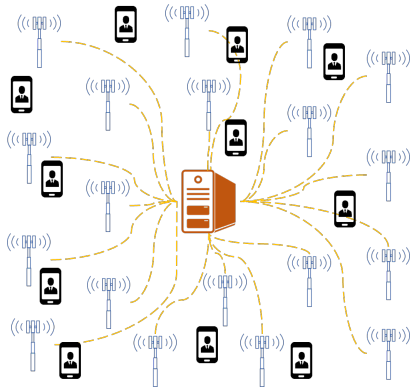

Figure 1. APs, users, and the CPU in the CF-mMIMO system.

matrix $\mathbf{R}$. The notations $\|\cdot\|_2$ and $\|\cdot\|_F$ denote the l_2 norm and Frobenious norm, respectively. $\mathcal{BD}(\mathbf{A}_i)$ denotes a block diagonal matrix with matrices $\mathbf{A}_i$ as its block diagonal components. The Kronecker delta $\delta_{kj} = 1$, when $k = j$, and is 0 otherwise. Notations $\text{mod}(a, b)$, $\lceil a \rceil$ and $1_{a\in\mathcal{A}}$ respectively denote the remainder of a/b, ceil value of a, and the indicator function.

II. System model and The GSP symbol

Consider L access points (APs) distributed in a given area as shown in Figure 1, serving K single antenna users. Without loss of generality, each AP is assumed to have M antennas. The received signal at the l^{th} BS 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 $l = 1, \ldots, L$ and $\mathbf{x}_k$ is the $T \times 1$ transmitted vector from the k^{th} user, for each $k = 1, \ldots, K$. The details of the transmitted vector is given in the next section. The quantity $\mathbf{h}_{kl}$ is the $M \times 1$ channel vector between k^{th} user and the l^{th} cell. We assume the Rayleigh fading channel model as

$$\mathbf{h}_{lk} \sim \mathcal{CN}\left(\mathbf{0}, \beta_{lk}\mathbf{I}_M\right), \tag{2}$$

where β_{lk} denotes the path loss of the corresponding channel link. The matrix $\mathbf{W}_l$ is the $M \times T$ additive Gaussian noise at the l^{th} BS with zero mean and σ^2 variance across its each entry. 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.

A. The GSP symbol

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

$$\mathbf{x}_k = \sqrt{P}\left(\mathbf{p}_k\sqrt{\lambda} + \mathbf{Z}_k\mathbf{s}_k\sqrt{1-\lambda}\right), \tag{3}$$

where P is the transmit power per user user per time slot; $\lambda \in [0, 1]$ 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 $\mathbf{s}_k$ denotes 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 $\mathbf{f}_i^H\mathbf{f}_j = T\delta_{ij}$. Then, each pilot can be selected as

$$\mathbf{p}_k = \mathbf{f}_{\underline{k}}, \forall k = 1, \ldots, K, \tag{4}$$

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

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

$$\frac{\mathbf{Z}_k^H\mathbf{p}_j}{T} = \begin{cases}\mathbf{0}, & (k,j) \in \bar{\mathcal{Z}},\\ \mathbf{e}_{(kj)}, & (k,j) \in \mathcal{Z},\end{cases} \tag{5}$$

where $\mathcal{Z} = \{1, \ldots, K\}^2 \setminus \bar{\mathcal{Z}}$; $\bar{\mathcal{Z}} = \left\{(k,j) : j \in \mathcal{C}_{\underline{k}}, \forall k = 1, \ldots, K\right\}$; $\mathbf{e}_{(kj)}$ is a vector of zeros and ones. Let $q_{kj} = \left\|\mathbf{e}_{(kj)}\right\|_2$. 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, and $k \notin \mathcal{Z}_k \subseteq \{i : \forall i \neq k\}$. 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

$$\frac{\mathbf{Z}_k^H\mathbf{Z}_l}{T} = \begin{cases}\mathbf{J}_{[kl]}, & k \neq l,\\ \mathbf{I}_d, & k = l,\end{cases} \tag{6}$$

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 defined as

$$\begin{aligned}\mathbf{R}_{k,ij}^{[\mathbf{Z}]} &= \mathbb{E}\frac{\mathbf{Z}_k^H\mathbf{x}_i}{T\sqrt{P}}\frac{\mathbf{x}_j^H\mathbf{Z}_k}{T\sqrt{P}}\\ &= \lambda\mathbf{e}_{ki}\mathbf{e}_{kj}^T + \frac{1-\lambda}{d}\delta_{ij}\mathbf{J}_{[ki]}\mathbf{J}_{[kj]}^T,\end{aligned} \tag{7}$$

and $c_{ki} = tr(\mathbf{R}_{k,ii}^{[\mathbf{Z}]})$.

III. Channel and data estimation in CF-mMIMO system

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 [7]. In the following, we first provide the analysis of localized processing based channel estimation and data detection. Thereafter, the CPU based centralized processing is analyzed.

A. Localized processing

To get the meaningful channel or data estimates, the received signal equation in (1) should satisfy the necessary condition that the number of equations (MT) must be at least equal to the number of variables, that is,

$$MK + Kd \leq MT, \tag{8}$$

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

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

$$\hat{\mathbf{h}}_{lk} = \arg\min_{\mathbf{h}_{lk}} \left\| \mathbf{Y}_l - \sqrt{P\lambda}\mathbf{h}_{lk}\mathbf{p}_k^H \right\|_F^2 \tag{9}$$

$$= \frac{\mathbf{Y}_l \mathbf{p}_k}{T\sqrt{P\lambda}} = \mathbf{h}_{lk} + \Delta_{lk}, \tag{10}$$

where the estimation error Δ_{lk} is given as

$$\Delta_{lk} = \frac{1}{T\sqrt{P\lambda}} \sum_{i \neq k} \mathbf{h}_{li}\mathbf{x}_i^H \mathbf{p}_k + \frac{\mathbf{W}_l \mathbf{p}_k}{T\sqrt{P\lambda}}$$

$$= \sum_{i \in \mathcal{C}_{\underline{k}} \setminus \{k\}} \mathbf{h}_{li} + \sqrt{\frac{1-\lambda}{\lambda}} \sum_{i \in \mathcal{Z}_k} \mathbf{h}_{li} s^*_{(ik)} + \mathbf{w}_{lk},$$

with $s^*_{(ik)} = \mathbf{s}_i^H \mathbf{e}_{(ik)}$ and $\mathbf{w}_{lk} = \frac{\mathbf{W}_l \mathbf{p}_k}{T\sqrt{P\lambda}} \sim \mathcal{CN}\left(\mathbf{0}, \frac{\sigma^2}{PT\lambda}\mathbf{I}_M\right)$. The factor $\frac{\sigma^2}{PT\lambda}$ denotes the inverse value of the pilot SNR over T time slots. Since the channel vectors are zero mean, the error Δ_{lk} has zero mean with the variance defined as $\mathbb{E}\|\Delta_{lk}\|_2^2$

$$\overset{(a)}{=} \sum_{i \in \mathcal{C}_{\underline{k}} \setminus \{k\}} \mathbb{E}\|\mathbf{h}_{li}\|_2^2 + \frac{1-\lambda}{\lambda} \sum_{i \in \mathcal{Z}_k} \mathbb{E}\|\mathbf{h}_{li}\|_2^2 \frac{q_{ik}}{d} + \mathbb{E}\|\mathbf{w}_{lk}\|_2^2$$

$$\overset{(b)}{=} \sum_{i \in \mathcal{C}_{\underline{k}} \setminus \{k\}} \beta_{li} M + \frac{1-\lambda}{\lambda} \sum_{i \in \mathcal{Z}_k} \frac{\beta_{li} M q_{ik}}{d} + \frac{\sigma^2 M}{PT\lambda} \overset{(c)}{=} M\alpha_{lk}$$

where in (a), the facts that $\mathbb{E}\left|\mathbf{s}_i^H \mathbf{e}_{(ik)}\right|^2 = \frac{q_{ik}}{d}$, and CSI and noise vectors are independent are used; in (b), $\mathbb{E}\|\mathbf{h}_{li}\|_2^2 = \beta_{li} M$; in (c), the variable α_{lk} is given as

$$\alpha_{lk} = \sum_{i \in \mathcal{C}_{\underline{k}} \setminus \{k\}} \beta_{li} + \frac{1-\lambda}{\lambda d} \sum_{i \in \mathcal{Z}_k} \beta_{li} q_{ik} + \frac{\sigma^2}{PT\lambda}. \tag{11}$$

The above factor shows that as the portion of data in the GSP symbol is decreased, the MSE decreases. The computational complexity is linear in terms of M and T.

2) 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

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

$$= \mathbf{Z}_k^H \mathbf{Y}_l^H \frac{\hat{\mathbf{h}}_{lk}}{\|\hat{\mathbf{h}}_{lk}\|_2^2} \cdot \frac{1}{T\sqrt{P(1-\lambda)}} \tag{12}$$

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

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

$$+ \frac{\mathbf{Z}_k^H \mathbf{W}^H \hat{\mathbf{h}}_{lk}}{\|\hat{\mathbf{h}}_{lk}\|_2^2} \cdot \frac{1}{T\sqrt{P(1-\lambda)}}, \tag{13}$$

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 1. *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*

$$P_{lk,S} = (\alpha_{lk} + \beta_{lk})^2 (M^2 + M), \tag{14}$$

$$P_{lk,SI} = \alpha_{lk}^2 M^2 \left(1 + \frac{1}{M} + \frac{\beta_{lk}}{M\alpha_{lk}}\right), \tag{15}$$

$$P_{lk,CI} = \frac{M}{1-\lambda} \sum_{i \neq k} c_{ki} \beta_{li} (\beta_{lk} + \alpha_{lk} + M\beta_{li}\zeta_{ik}), \tag{16}$$

$$P_{lk,N} = \frac{\sigma^2 M d}{TP(1-\lambda)} \left[\beta_{lk} + \alpha_{lk} + \frac{\sigma^2(M+1)}{T^2 P\lambda}\right], \tag{17}$$

where $\zeta_{jk} = 1_{j \in \mathcal{C}_{\underline{k}} \setminus \{k\}} + \frac{1-\lambda}{\lambda d} 1_{j \in \mathcal{Z}_k}$.

Proof: Proofs are given in the Appendix-A. ∎

It can be seen that the desired signal and SI power increases proportional to M^2. If the necessary condition in (8) is satisfied or $T \geq K + d$, $\zeta_{ik} = 0$ and CI power vanishes as compared to SI power. The M^2 factor in the noise power arises due to channel estimation errors. The corresponding rate expression can be written as $R_{lk} = \log_2(1 + SINR_{lk})$, where $SINR_{lk} = \frac{P_{lk,S}}{P_{lk,SI} + P_{lk,CI} + P_{lk,N}}$. Asymptotically, for $M \to \infty$, we have

$$\lim_{M \to \infty} SINR_k \to \frac{(\alpha_{lk} + \beta_{lk})^2}{\alpha_{lk}^2 + \frac{1}{1-\lambda}\sum_{i \neq k} c_{ki}\beta_{li}^2 \zeta_{ik} + \frac{\sigma^4 d}{T^3 P^2 (1-\lambda)\lambda}}, \tag{18}$$

where only the interference terms is present, that can be reduced with successive or iterative cancellation scheme, described in the next section.

B. Centralized processing

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

$$MLT \geq KML + Kd, \tag{19}$$

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

$$\underbrace{\begin{bmatrix}\mathbf{Y}_1\\ \vdots \\ \mathbf{Y}_L\end{bmatrix}}_{:=\mathbf{Y}} = \sum_k \underbrace{\begin{bmatrix}\mathbf{h}_{1k}\\ \vdots \\ \mathbf{h}_{Lk}\end{bmatrix}}_{:=\mathbf{h}_k} \mathbf{x}_k^H + \underbrace{\begin{bmatrix}\mathbf{W}_1\\ \vdots \\ \mathbf{W}_L\end{bmatrix}}_{:=\mathbf{W}}, \tag{20}$$

where $\mathbf{h}_k \sim \mathcal{CN}(\mathbf{0}, \mathbf{B}_k)$ with $\mathbf{B}_k{=}\mathcal{BD}(\beta_{1k}\mathbf{I}_M, \ldots, \beta_{Lk}\mathbf{I}_M)$. The corresponding channel estimates can be written as $\hat{\mathbf{h}}_k^H := \left[\hat{\mathbf{h}}_{1k}^H, \ldots, \hat{\mathbf{h}}_{Lk}^H\right] = \frac{\mathbf{Y}\mathbf{p}_k}{T\sqrt{P\lambda}} = \mathbf{h}_k^H + \Delta_k^H$, with $\Delta_k^H := \left[\Delta_{1k}^H, \ldots, \Delta_{Lk}^H\right]$ expressed as

$$\Delta_k = \sum_{i \in \mathcal{C}_{\underline{k}} \setminus \{k\}} \mathbf{h}_i + \sqrt{\frac{1-\lambda}{\lambda}} \sum_{i \in \mathcal{Z}_k} \mathbf{h}_i s_{(ik)}^* + \frac{\mathbf{W}\mathbf{p}_k}{T\sqrt{P\lambda}}. \tag{21}$$

This leads to the channel uncertainty covariance matrix as

$$\begin{aligned}\mathbf{A}_k &= \mathbb{E}\left[\Delta_k \Delta_k^H\right] \\ &= \sum_{i \in \mathcal{C}_{\underline{k}} \setminus \{k\}} \mathbf{B}_i + \frac{1-\lambda}{\lambda d} \sum_{i \in \mathcal{Z}_k} \mathbf{B}_i q_{ik} + \frac{\sigma^2 ML}{TP\lambda} \\ &= \mathcal{BD}\left(\mathbb{E}\left[\Delta_{1k}\Delta_{1k}^H\right], \ldots, \mathbb{E}\left[\Delta_{Lk}\Delta_{Lk}^H\right]\right) \\ &= \mathcal{BD}\left(\alpha_{1k}\mathbf{I}_M, \ldots, \alpha_{Lk}\mathbf{I}_M\right),\end{aligned}$$

which provides the MSE coefficient as

$$\begin{aligned}\alpha_k &:= \frac{\mathbb{E}\left\|\Delta_k\right\|^2}{ML} = \frac{1}{L}\sum_{l=1}^{L} \frac{\mathbb{E}\Delta_{lk}^H \Delta_{lk}}{M} = \frac{1}{L}\sum_{l=1}^{L} \alpha_{lk} \\ &= \sum_{i \in \mathcal{C}_{\underline{k}} \setminus \{k\}} \beta_i + \frac{1-\lambda}{\lambda d} \sum_{i \in \mathcal{Z}_k} \beta_i q_{ik} + \frac{\sigma^2}{TP\lambda},\end{aligned}$$

with $\beta_i = \sum_{l=1}^{L} \frac{\beta_{li}}{L}$.

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

$$\begin{aligned}\hat{\mathbf{s}}_k &= \arg\min_{\mathbf{s}_k} \left\|\mathbf{Y} - \sqrt{P(1-\lambda)}\hat{\mathbf{h}}_k \mathbf{s}_k^H \mathbf{Z}_k^H\right\|_F^2 \\ &= \mathbf{Z}_k^H \mathbf{Y}^H \frac{\hat{\mathbf{h}}_k}{\|\hat{\mathbf{h}}_k\|_2^2} \cdot \frac{1}{T\sqrt{P(1-\lambda)}} \qquad (22) \\ &= \mathbf{Z}_k^H \left[\sum_i \mathbf{x}_i \mathbf{h}_i^H\right] \frac{\hat{\mathbf{h}}_k}{\|\hat{\mathbf{h}}_k\|_2^2} \cdot \frac{1}{T\sqrt{P(1-\lambda)}} \qquad (23) \\ &= \mathbf{s}_k + \underbrace{\mathbf{s}_k\left(\frac{\mathbf{h}_k^H \hat{\mathbf{h}}_k}{\|\hat{\mathbf{h}}_k\|_2^2} - 1\right)}_{:=\mathbf{s}_{k,SI}} + \underbrace{\sum_{i \neq k} \frac{\mathbf{Z}_k^H \mathbf{x}_i}{T\sqrt{P}} \frac{\mathbf{h}_i^H \hat{\mathbf{h}}_k}{\|\hat{\mathbf{h}}_k\|_2^2} \frac{1}{\sqrt{1-\lambda}}}_{:=\mathbf{s}_{k,CI}} \\ &\quad + \frac{\mathbf{Z}_k^H \mathbf{W}^H \hat{\mathbf{h}}_k}{\|\hat{\mathbf{h}}_k\|_2^2} \cdot \frac{1}{T\sqrt{P(1-\lambda)}}, \qquad (24)\end{aligned}$$

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 (13), and the strength of these terms is derived in the following lemma.

Theorem 2. *For centralized processing at the CPU, the power of desired signal, SI, CI and noise for the k^{th} user can be obtained as*

$$P_{k,S} = M^2L^2\left(\alpha_k + \beta_k\right)^2 + ML\sum_{l=1}^{L}\frac{\left(\alpha_{lk} + \beta_{lk}\right)^2}{L}, \tag{25}$$

$$P_{k,SI} = \alpha_k^2 M^2 L^2 \left(1 + \frac{1}{L\alpha_k^2}\sum_{l=1}^{L}\frac{\beta_{lk}\alpha_{lk}}{ML} + \frac{1}{ML}\sum_{l=1}^{L}\frac{\alpha_{lk}^2}{L\alpha_k^2}\right), \tag{26}$$

$$P_{k,CI} = \frac{ML}{1-\lambda}\sum_{i \neq k} c_{ki} \sum_l \frac{\beta_{li}}{L}\left(\beta_{lk} + \alpha_{lk} + ML\frac{\beta_i^2}{\beta_{li}}\zeta_{ik}\right), \tag{27}$$

$$P_{k,N} = \frac{\sigma^2 MLd}{TP(1-\lambda)}\left[\beta_k + \alpha_k + \frac{\sigma^2(ML+1)}{T^2 P\lambda}\right], \tag{28}$$

Proof: Proofs are derived in the Appendix-B. ∎

As compared to the localized processing, all the terms are proportional to L, the number of APs. The rate expression for the k^{th} user can be given as $R_k = \log_2(1 + SINR_k)$, 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 $ML \to \infty$, we have

$$SINR_k \to \frac{\left(\alpha_k + \beta_k\right)^2}{\alpha_k^2 + \frac{1}{1-\lambda}\sum_{i \neq k} c_{ki}\beta_i^2\zeta_{ik} + \frac{\sigma^4 d}{T^3P^2\lambda(1-\lambda)}}. \tag{29}$$

It 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.

C. Cooperation overhead and distributed time processing

In this GSP scheme, each AP forwards the received signals $\mathbf{Y}_l$ to the CPU. Note that the received signal matrix $\mathbf{Y}_l$ spans for T time slots. Thus, instead of forwarding at the end of T slots. They can be forwarded from the first time slot. Let $\mathbf{Y}_l = [\mathbf{y}_l(1), \ldots, \mathbf{y}_l(T)]$, where $\mathbf{y}_l(t)$ be the t^{th} column of

$\mathbf{Y}_l$. Then, we can write the channel estimation equation at the end of time slot T as

$$\hat{\mathbf{h}}_{lk}(T) = \frac{\mathbf{Y}_l \mathbf{p}_k}{T\sqrt{P\lambda}} = \frac{1}{T\sqrt{P\lambda}} \sum_{i=1}^{T} \mathbf{y}_l(i) p_k(i), \tag{30}$$

where $p_k(i)$ is the i^{th} entry of the pilot vector $\mathbf{p}_k$. For distributed time processing at the CPU, we can update the estimate as

$$\begin{aligned}\hat{\mathbf{h}}_{lk}(t+1) &= \frac{1}{T\sqrt{P\lambda}} \sum_{i=1}^{t+1} \mathbf{y}_l(i) p_k(i) \\ &= \frac{\sum_{i=1}^{t} \mathbf{y}_l(i) p_k(i) + \mathbf{y}_l(t+1) p_k(t+1)}{T\sqrt{P\lambda}} \qquad (31) \\ &= \hat{\mathbf{h}}_{lk}(t) + \frac{\mathbf{y}_l(t+1) p_k(t+1)}{T\sqrt{P\lambda}}, \qquad (32)\end{aligned}$$

which significantly reduces the computational requirements at the end. Similarly, for the data estimates, we have

$$\begin{aligned}\hat{\mathbf{s}}_{lk}(T) &= \mathbf{Z}_k^H \mathbf{Y}_l^H \frac{\hat{\mathbf{h}}_{lk}(T)}{\|\hat{\mathbf{h}}_{lk}(T)\|_2^2} \cdot \frac{1}{T\sqrt{P(1-\lambda)}} \\ &= \sum_{i=1}^{T} \mathbf{z}_k(i) \mathbf{y}_l^H(i) \frac{\hat{\mathbf{h}}_{lk}(T)}{\|\hat{\mathbf{h}}_{lk}(T)\|_2^2} \cdot \frac{1}{T\sqrt{P(1-\lambda)}} \qquad (33) \\ &= \frac{\left[\mathbf{U}_{lk}(T-1) + \mathbf{z}_k(T)\mathbf{y}_l^H(T)\right] \hat{\mathbf{h}}_{lk}(T)}{\|\hat{\mathbf{h}}_{lk}(T)\|_2^2 T\sqrt{P(1-\lambda)}}, \qquad (34)\end{aligned}$$

where $\mathbf{Z}_k^H = \left[\mathbf{z}_k^H(1), \ldots, \mathbf{z}_k^H(T)\right]$; $\mathbf{U}_{lk}(t) = \sum_{i=1}^{t} \mathbf{z}_k(i)\mathbf{y}_l^H(i)$. The relation between the central and distributed data estimates can be given as $\hat{\mathbf{s}}_k = \sum_l \frac{\|\hat{\mathbf{h}}_{lk}\|_2^2}{\sum_l \|\hat{\mathbf{h}}_{lk}\|_2^2} \cdot \hat{\mathbf{s}}_{lk}$.

IV. Iterative estimation for the GSP scheme

From the above equations, it can be observed that channel and data estimates depend on each other. Therefore, an iterative channel and data estimation scheme can be employed to improve the bit error rate performance. Given the data and channel estimates $\left(\hat{\mathbf{s}}_k, \hat{\mathbf{h}}_{lk}\right)$, the channel estimates can be updated from (30) as

$$\hat{\mathbf{h}}_{lk} \leftarrow \frac{\tilde{\mathbf{Y}}_{lk} \mathbf{p}_k}{T\sqrt{P\lambda}}, \tag{35}$$

where $\tilde{\mathbf{Y}}_{lk} = \mathbf{Y}_l - \sum_{i \neq k} \hat{\mathbf{h}}_{li} \hat{\mathbf{x}}_i^H = [\tilde{\mathbf{y}}_{lk}(1), \ldots, \tilde{\mathbf{y}}_{lk}(T)]$. The corresponding update at the time slot t can be written as

$$\hat{\mathbf{h}}_{lk}(t) \leftarrow \hat{\mathbf{h}}_{lk}(t-1) + \frac{\tilde{\mathbf{y}}_{lk}(t) p_k(t)}{T\sqrt{P\lambda}}, \tag{36}$$

where based on $\tilde{\mathbf{Y}}_{lk}$, we have

$$\tilde{\mathbf{y}}_{lk}(t) = \mathbf{y}_l(t) - \sum_{i \neq k} \hat{\mathbf{h}}_{ki} \hat{x}_i^*(t), \tag{37}$$

$$\hat{x}_k(t) = \sqrt{P}\left(p_k(t)\sqrt{\lambda} + \mathbf{z}_k^H(t)\hat{\mathbf{s}}_k\sqrt{1-\lambda}\right). \tag{38}$$

Similarly, for $\tilde{\mathbf{Y}}_k = \mathbf{Y} - \sum_{i \neq k} \hat{\mathbf{h}}_i \hat{\mathbf{x}}_i^H = \left[\tilde{\mathbf{Y}}_{1k}^T, \ldots, \tilde{\mathbf{Y}}_{Lk}^T\right]^T$, the data estimates at CPU can be updated as $\hat{\mathbf{s}}_k \leftarrow \sum_l \frac{\|\hat{\mathbf{h}}_{lk}\|_2^2}{\sum_l \|\hat{\mathbf{h}}_{lk}\|_2^2} \hat{\mathbf{s}}_{lk}$, where $\hat{\mathbf{s}}_{lk} \leftarrow \mathbf{Z}_k^H \tilde{\mathbf{Y}}_{lk}^H \frac{\hat{\mathbf{h}}_{lk}}{\|\hat{\mathbf{h}}_{lk}\|_2^2} \cdot \frac{1}{T\sqrt{P(1-\lambda)}}$; the respective update in the time slot t can be given from (34) as

$$\begin{aligned}\hat{\mathbf{s}}_k(t) &\leftarrow \sum_l \frac{\|\hat{\mathbf{h}}_{lk}(t)\|_2^2}{\sum_l \|\hat{\mathbf{h}}_{lk}(t)\|_2^2} \hat{\mathbf{s}}_{lk}(t) \qquad (39) \\ &= \sum_l \frac{\left[\mathbf{U}_{lk}(t-1) + \mathbf{z}_k(t)\tilde{\mathbf{y}}_{lk}^H(t)\right] \hat{\mathbf{h}}_{lk}(t)}{T\sqrt{P(1-\lambda)} \sum_l \|\hat{\mathbf{h}}_{lk}(t)\|_2^2}. \qquad (40)\end{aligned}$$

The above distributed time processing and iterative estimation procedure at the CPU is summarized in the Algorithm 1, where $N_{\max}$ be the maximum number of iterations. This process divides the final iterative processing at the end of T time slots to over T time slots, which can result into faster processing of superimposed symbols. It can be seen from the literature [29]–[31] and the simulations that the iterative algorithm converges in few iterations subject to certain conditions, which are the necessary conditions mentioned in Table I. Due to less number of iterations, that is, counting $N_{\max}$ as a constant, the computational overhead is approximately the same as $\mathcal{O}(MKT)$.

Algorithm 1 Iterative algorithm distributed time processing at the CPU.

```
Initialize ĥ_lk(0) = 0, ŝ_lk(0) = 0, ∀l, k.
for t = 1, ..., T do
    receive observations from APs y_l(t), ∀l
    for iter = 1, ..., N_max do
        compute and update ỹ_lk(t) via (37) and (38).
        update channel estimates ĥ_lk(t) via (36).
        update data estimates ŝ_k(t) via (39).
    end for
end for
```

A. Optimization for λ and d

There are two factors that can be optimized for rate maximization. It can be seen from simulation results that the rate is a concave function with respect to λ, and linear with respect to d. Thus, for a specified d, there exist an optimum value of λ, that maximizes the rates. Based on the necessary conditions in (8) and (19), the rate optimization is given as follows.

For localized processing, the sum rate maximization problem for λ and d can be cast as

$$\arg \max_{\lambda \in [0,1], d \in \{1,\ldots,T-1\}} \sum_{lk} R_{lk}. \tag{41}$$

Similarly, for the centralized processing, the sum rate optimization can be written as

$$\arg \max_{\lambda \in [0,1], d \in \{1,\ldots,T-1\}} \sum_{k} R_{k}. \tag{42}$$

Note that the above problems for fixed d, are concave. Thus, they can be solved via a convex solver, such as CVX. Note that λ and d are system design variables. Therefore, the optimum value of d can be obtained by searching linearly in the set $\left\{1, \ldots, \left\lceil M\left(\frac{T}{K} - 1\right)\right\rceil\right\}$ for localized estimation or $\left\{1, \ldots, \left\lceil ML\left(\frac{T}{K} - 1\right)\right\rceil\right\}$ for centralized estimations, where the upper bound is obtained from the condition (8) and (19), respectively.

V. 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., $\mathbf{x}_k = \sqrt{P}\left(\mathbf{p}_k\sqrt{\lambda}+\sqrt{1-\lambda}\mathbf{F}\mathbf{s}_k\right)$, 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 (9), where the estimation error Δ_{lk} for this case is obtained as $\Delta_{lk} = \sum_{i\in\mathcal{C}_{\underline{k}}\setminus\{k\}}\mathbf{h}_{li} + \sqrt{T\frac{1-\lambda}{\lambda}}\sum_{i\neq k}\mathbf{h}_{li}\mathbf{s}_i^H\mathbf{p}_k + \frac{\mathbf{W}_l\mathbf{p}_k}{T\sqrt{P\lambda}}$. The error Δ_{lk} has zero mean with the variance factor given as

$$\alpha_{lk} = \sum_{i\in\mathcal{C}_{\underline{k}}\setminus\{k\}}\beta_{li} + T\frac{1-\lambda}{\lambda}\sum_{i\neq k}\beta_{li} + \frac{\sigma^2}{PT\lambda}, \tag{43}$$

where $\mathbb{E}\left|\mathbf{s}_i^H\mathbf{p}_k\right|^2 = tr\mathbf{p}_k^H\left(\mathbb{E}\mathbf{s}_k\mathbf{s}_k^H\right)\mathbf{p}_k = \frac{1}{T}\mathbf{p}_k^H\mathbf{p}_k = 1$. It can be seen that the data term in the channel estimation error is increased from $\frac{1}{d}$ in (11) to T, that is, the increment factor is Td, which is significantly large.

1) Localized processing: For reasonable estimates, the system should follow the necessary condition $MT \geq KM+KT$. The local data estimates via LS can be calculated as $\hat{\mathbf{s}}_{lk} = \left(\frac{\mathbf{Y}_l^H\hat{\mathbf{h}}_{lk}}{\|\hat{\mathbf{h}}_{lk}\|_2^2} - \sqrt{P\lambda}\mathbf{p}_k\right)\cdot\frac{1}{\sqrt{TP(1-\lambda)}}$.

2) Centralized processing: For the centralized channel estimation, the channel errors and the corresponding covariance matrices can similarly be defined as in section III-B1, i.e., $\alpha_k = \frac{1}{L}\sum_{l=1}^{L}\alpha_{lk}$ with α_{lk} defined above in (43). Towards the data detection, the data estimates via LS can be expressed as $\hat{\mathbf{s}}_k = \left(\frac{\mathbf{Y}^H\hat{\mathbf{h}}_k}{\|\hat{\mathbf{h}}_k\|_2^2} - \sqrt{P\lambda}\mathbf{p}_k\right)\cdot\frac{1}{\sqrt{TP(1-\lambda)}}$.

Similar to the GSP scheme, the above equation shows the limiting SINR for the large number of antennas. However, the SINR is influenced greatly by channel estimation error α_{lk} or α_k.

B. 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 contamination, we must have $T_p \geq K$. If $K > T_p$, 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

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

The channel estimates via LS can be given as $\hat{\mathbf{h}}_{lk} = \frac{\mathbf{Y}_{p,l}\mathbf{p}_k}{T_p\sqrt{\frac{\lambda TP}{T_p}}} = \mathbf{h}_{lk}+\Delta_{lk}$,where $\Delta_{lk} = \sum_{i\in\mathcal{C}_{\underline{k}}\setminus\{k\}}\mathbf{h}_{li} + \frac{\mathbf{W}_{p,l}\mathbf{p}_k}{T_p\sqrt{\frac{\lambda TP}{T_p}}}$. The error Δ_{lk} has zero mean with the variance given as

$$\mathbb{E}\|\Delta_{lk}\|_2^2 = \sum_{i\in\mathcal{C}_{\underline{k}}\setminus\{k\}}\mathbb{E}\|\mathbf{h}_{li}\|_2^2 + \mathbb{E}\|\frac{\mathbf{W}_{p,l}\mathbf{p}_k}{T_p\sqrt{\frac{\lambda TP}{T_p}}}\|_2^2 \tag{44}$$

$$= \underbrace{\left(\sum_{i\in\mathcal{C}_{\underline{k}}\setminus\{k\}}\beta_{li} + \frac{\sigma^2}{PT\lambda}\right)}_{\alpha_{lk}}M. \tag{45}$$

To get the channel estimates at the CPU, there are two options. Each AP can forward either the $M\times T_p$ received signal matrix $\mathbf{Y}_{p,l}, \forall l$, or directly 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 III-B1, we can write the channel error vector and its variance factor as $\Delta_k = \sum_{i\in\mathcal{C}_{\underline{k}}\setminus\{k\}}\mathbf{h}_i + \frac{\mathbf{W}_p\mathbf{p}_k}{T_p\sqrt{\frac{\lambda TP}{T_p}}}$ and $\alpha_k = \sum_{i\in\mathcal{C}_{\underline{k}}\setminus\{k\}}\sum_{l=1}^{L}\frac{\beta_{li}}{L} + \frac{\sigma^2}{PT\lambda}$.

2) Localized processing: 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

$$\mathbf{x}_k = \sqrt{PT(1-\lambda)}\mathbf{V}_k\mathbf{s}_k, \tag{46}$$

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

3) Centralized processing: The received signal at the CPU can be written as in (20), where we have MLT_d equations and dK variables. 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 $\frac{MLT_d}{K}$. That is, the CPU based estimation can increase the data rate significantly. The data estimates using least squares can be obtained as

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

×	×	Conditions
GSP	L	$d \leq M\left(\frac{T}{K}-1\right)$
	C	$d \leq M\left(\frac{LT}{K}-1\right)$
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 I
COMPARISON OF NECESSARY CONDITIONS FOR DIFFERENT TRANSMISSION SCHEMES.

It can be noted that this limiting case is not affected by the transmitted power. It is rather influenced by the number of 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 I shows the different condition for meaningful channel and data estimation for different schemes.

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 [7]

Cooperation levels L1-L4 are adopted from [7]. Level 4 is fully centralized processing case, where all the observations from APs are forwarded to CPU without any local processing. Levels 3 and 2 are gradually relaxed versions of L4. In L3, local channel and data estimates are computed and sent to the CPU, whereas in L2, only data estimates are forwarded. Lastly, for L1, neither any estimate nor observation is sent to CPU; each AP decodes one user's channel and data estimates, treating the others' transmission as noise.

1) Level 3: local processing and large scale fading decoding: From the level 4 processing, the data estimates can be written in terms of local channel estimates as

$$\begin{aligned}\hat{\mathbf{s}}_k &= \mathbf{Z}_k^H \mathbf{Y}^H \frac{\hat{\mathbf{h}}_k}{\|\hat{\mathbf{h}}_k\|_2^2} \cdot \frac{1}{T\sqrt{P(1-\lambda)}} \\ &= \sum_{l=1}^{L} \frac{\|\hat{\mathbf{h}}_{lk}\|_2^2}{\sum_{l=1}^{L}\|\hat{\mathbf{h}}_{lk}\|_2^2} \cdot \mathbf{Z}_k^H \mathbf{Y}_l^H \frac{\hat{\mathbf{h}}_{lk}}{\|\hat{\mathbf{h}}_{lk}\|_2^2} \cdot \frac{1}{T\sqrt{P(1-\lambda)}} \\ &= \sum_{l=1}^{L} w_{lk} \cdot \hat{\mathbf{s}}_{lk},\end{aligned}$$

where $\hat{\mathbf{s}}_{lk} = \mathbf{Z}_k^H \mathbf{Y}_l^H \frac{\hat{\mathbf{h}}_{lk}}{\|\hat{\mathbf{h}}_{lk}\|_2^2} \cdot \frac{1}{T\sqrt{P(1-\lambda)}}$ denotes locally processed data estimate. The above equation denotes that the central estimate is a linear sum of local estimates with weights $w_{lk} = \frac{\|\hat{\mathbf{h}}_{lk}\|_2^2}{\sum_{l=1}^{L}\|\hat{\mathbf{h}}_{lk}\|_2^2}$. It means that the data estimates can be locally calculated and forwarded to the CPU for the final estimation. 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 with $w_{lk} = \frac{1}{L(\beta_k+\alpha_k)}$, where it is assumed that CC knows the locations of the users and APs to compute β_k and α_k. In this case, each l^{th} AP can send the value $\hat{\mathbf{s}}_{lk}\frac{\|\hat{\mathbf{h}}_{lk}\|_2^2}{M} = \mathbf{Z}_k^H \mathbf{Y}_l^H \frac{\hat{\mathbf{h}}_{lk}}{M}$, instead of $\hat{\mathbf{s}}_{lk}$ as

$$\hat{\mathbf{s}}_k^{[L2]} = \frac{1}{ML(\beta_k+\alpha_k)} \sum_{l=1}^{L} \hat{\mathbf{s}}_{lk}\|\hat{\mathbf{h}}_{lk}\|_2^2 = \hat{\mathbf{s}}_k \frac{\|\hat{\mathbf{h}}_k\|_2^2}{ML(\beta_k+\alpha_k)},$$

which tends to $\hat{\mathbf{s}}_k$ as $ML \to \infty$. In this case, SINR also remains unaltered, since a scalar is multiplied in both desired signal term and interference-noise terms.

VI. SIMULATION RESULTS

A. Simulation settings

We consider $L = 32$ APs with each AP having $M = 2$ antennas and $T = 6$. Simulations are averaged over 10^4 runs. Locations of APs are selected via uniform distribution within a 2D area $[-500, 500]^2\, m^2$. The minimum distance between APs is set to be $50m$. An urban environment setup is chosen with 3GPP microcell pathloss model at 2GHz frequency as

$$\beta_{lk}(\text{in dB}) = -30.5 - 36.7\log_{10} D_{kl} + F_{lk}, \qquad (47)$$

where D_{lk} is the distance between the l^{th} AP and the k^{th} user; and $F_{lk} \sim \mathcal{N}\left(0, 4^2\right)$ represents shadowing [7]. For the RPs, the training time $T_p = \lceil \min(K, 0.25T) \rceil$ is set to be 25% of coherence time slots. Transmit power is set to be $P_0 = 100$ mW, $\rho = 0.9$, and the noise power is $\sigma^2 = -126$ dB.

Note that simulations for the GSP scheme have been verified with the theoretical results. However, for the simplicity of the presentation, the analytical results are omitted in the following figures. The following results present the scenario of pilot contamination and data detection error rates.

B. Channel estimation

Figure 2 plots the MSE for the channel estimates with respect to the pilot power fraction (λ) for three pilot schemes with different K and d. It can be observed that MSE changes in RP with λ are negligible, since for $K \geq 2$, we have $T_p = 2$, which causes pilot contamination, that is, the pilot reuse (when $T_p < K$) affects the strength of estimates than the SNR. RP yields less MSE than SP schemes at very low value of pilot power (around $\lambda = 0.1$), whereas for higher λ, RP has around 10 times higher MSE than SP and GSP schemes. It can be seen that the MSE of conventional SP scheme is also (approximately 92%, 30% and 13% at $K = 4, 5, 8$ respectively) higher than that of the GSP scheme. However, as the number of users are increased in the system,

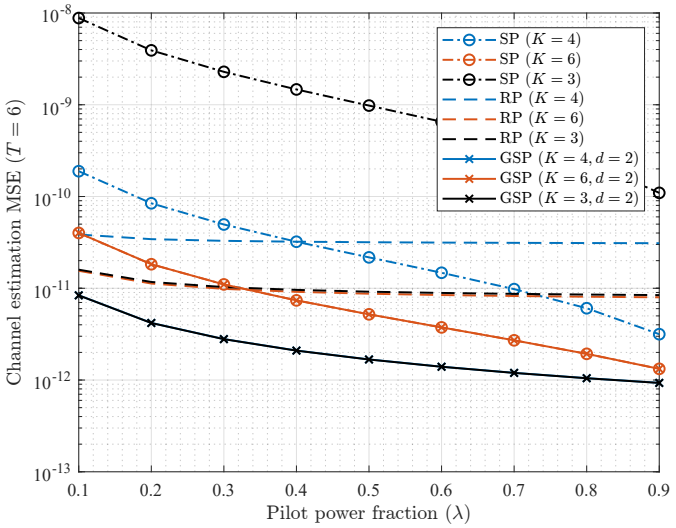

Figure 2. Averaged MSE $\frac{1}{K}\sum_k \alpha_k = \frac{1}{LK}\sum_{l,k} \alpha_{lk}$ of the channel estimation versus the pilot power allocation factor, when $T < K + d$ for the GSP scheme, and $d = T_d$ for the RP scheme.

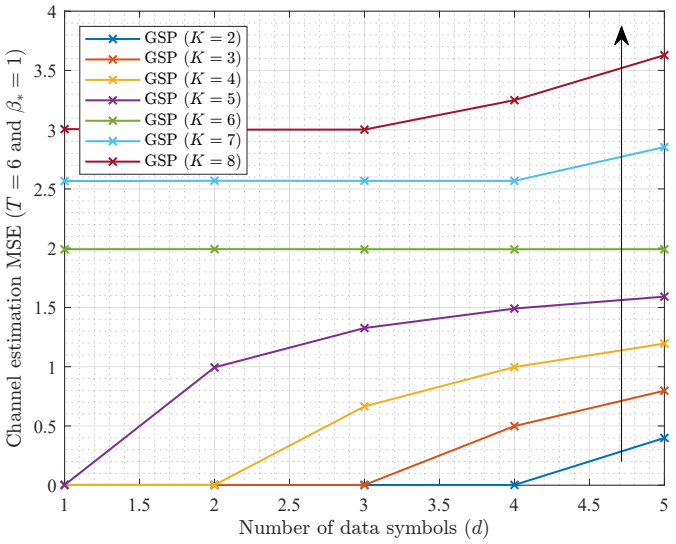

Figure 3. Averaged MSE $\frac{1}{K}\sum_k \alpha_k$ of the channel estimation for the GSP scheme versus the number of data symbols (d) at $\lambda = 0.5$.

the gap between them disappears due to larger strength of pilot contamination.

For the GSP scheme, the number of data symbols also affects the channel estimation. Thus, Figure 3 plots the MSE of channel estimates versus the number of data symbols (d) with $\lambda = 0.5$ for the case of $\beta_{lk} = 1, \forall l, k$. It can be noted that the MSE increases as d is increased. MSE is close to zero (of the order of 10^{-12}) for the cases of $T \geq K + d$. For $T < K + d$, MSE increases per symbol increase in d. For large number of users, MSE increases to larger values. For per user increase with $d = 6$ (between $K = 2, \ldots, T$), approximately 0.4 MSE value increase can be seen.

C. Interference powers and sum rates

For the localized processing, Figure 4 plot the average powers of SI $\frac{1}{LK}\sum_{l,k} P_{lk,SI}$, and CI $\frac{1}{LK}\sum_{l,k} P_{lk,CI}$, on left and

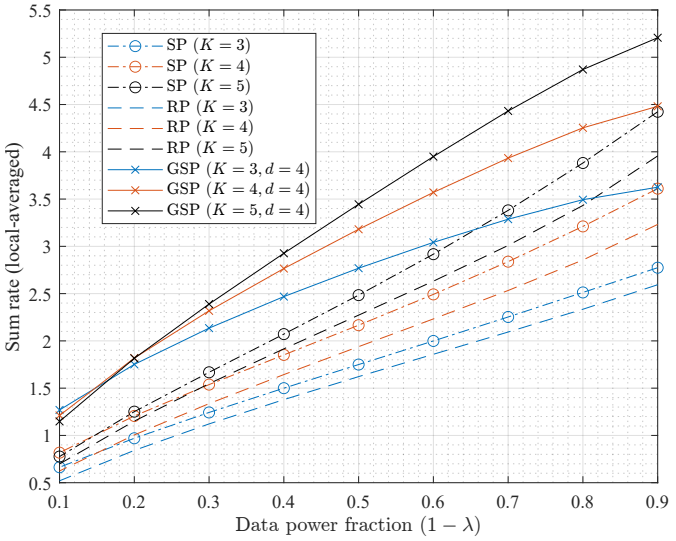

Figure 5. For the localized processing, the average sum rate $\frac{1}{L}\sum_{l,k} R_{lk}$ versus the data power fraction $(1 - \lambda)$ for $T = 6$.

right axes respectively, and the average sum rates $\frac{1}{L}\sum_{l,k} R_{lk}$ are plotted in Figure 5. Note that these performance measures are averaged over the number of APs in the system. In these figures, the RP performs worse, since $T_p < K$ provides poor channel estimates; and as the number of users, K, is increased, the performance gets worse.

- From Figure 4, as the data power fraction is increased, the self interference increases, which is intuitive as the SI term for the k^{th} user arises due to its own channel imperfections. Since the GSP provides better channel estimates, the SI power of the same is lower than that of the conventional SP scheme. The gap between the GSP and the SP scheme gets lower, as the number of users or the number of data symbols is increased.
- Figure 4 shows the convex behavior for the CI power when $T < K + d$ with $K = 5$, 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 scheme follows the similar behavior similar as for the SI power.
- Figure 5 reflects the concave behavior of the averaged sum rates with respect to λ. It shows that there exists an optimum value of λ that maximizes the sum rates. 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 $K = 3$ and $\lambda = 0.5$, around 58% of rate improvement can be observed with respect to the SP scheme.

For centralized processing, the power of SI & CI and the sum rate is plotted in Figures 6 and 7 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

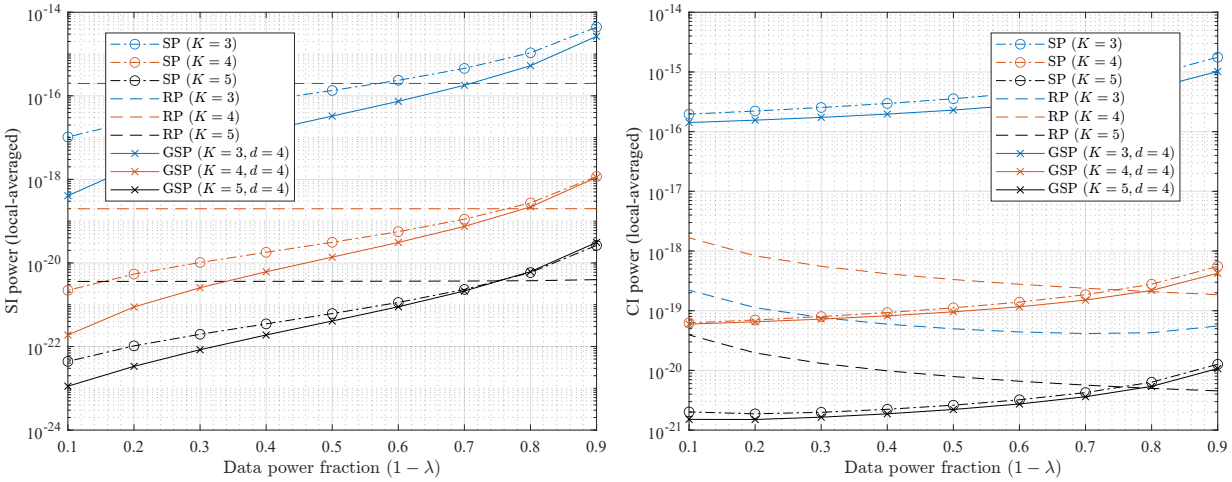

Figure 4. For the localized processing, the powers of SI (left) and the CI (right) versus the data power fraction $(1-\lambda)$ for $T=6$.

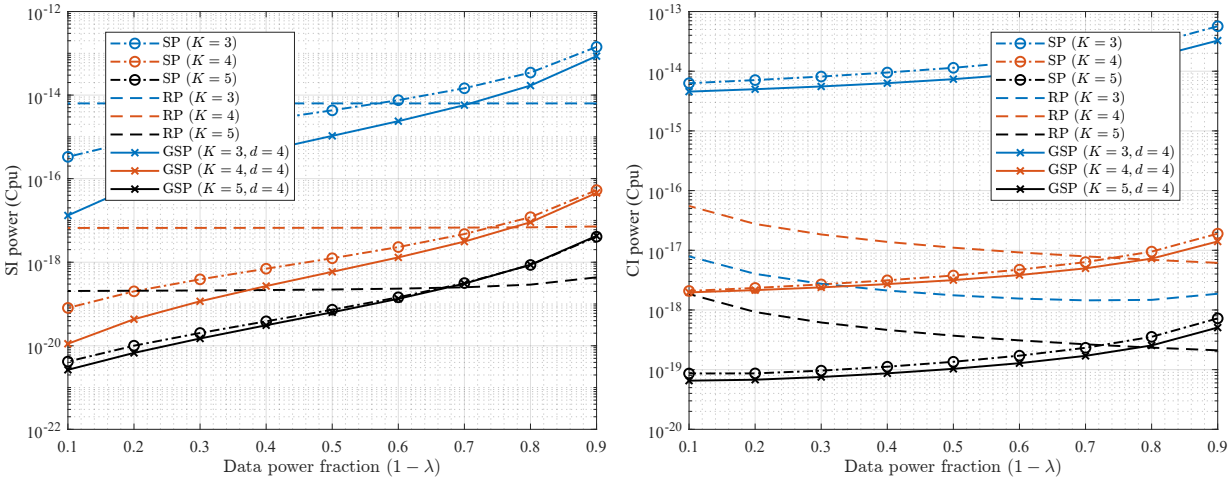

Figure 6. For the centralized processing, the powers of SI (left) and the CI (right) versus the data power fraction $(1-\lambda)$ for $T=6$.

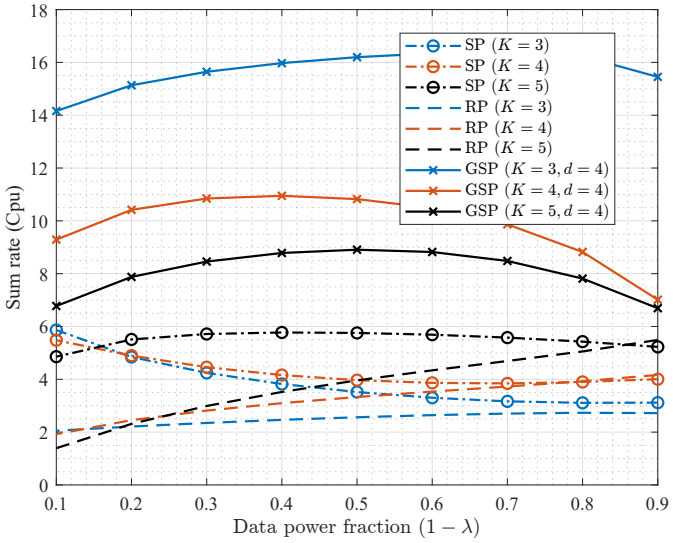

Figure 7. For the centralized processing, the average sum rate $\frac{1}{L}\sum_k R_k$ versus the data power fraction $(1-\lambda)$ for $T=6$.

localized processing. Regarding the trend with respect to K and d for different schemes, similar points can be drawn as in Figures 4-5. 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 4, 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 (λ) is required for CPU based processing. In terms of rate, for $\lambda = 0.5$ and $K = 3, 4, 5$, improvements of respectively 360%, 172% and 54% are present for the GSP scheme over the SP scheme.

D. MSE of data estimates with iterative cancellation

Figure 8 plots the MSE of data estimates at CPU, averaged across all users at different steps of iterative procedure described earlier. It can be seen that the iterative procedure improves the bit rates as iteration progresses. However, due to higher amount of self and cross interference powers incurred in low complexity processing (instead of using inverse via zero forcing), the MSE values converge in few iterations. It can also be seen that as with respect to the power of data transmission,

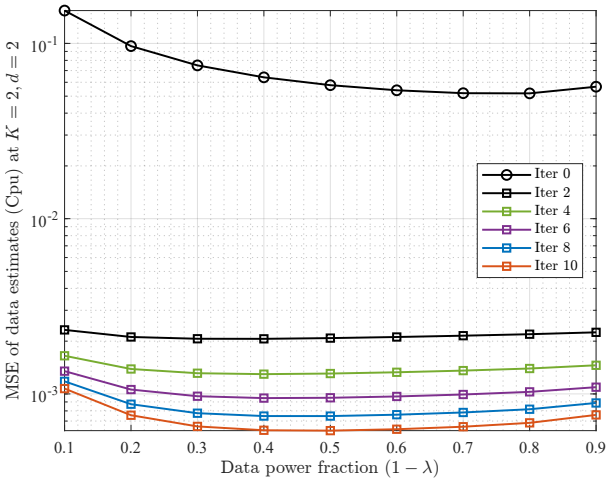

Figure 8. For the centralized processing, the average MSE of data estimates versus the data power fraction $(1-\lambda)$ at different iterations.

the MSE shows convex behavior, which occurs due to worse channel estimation at $(1-\lambda) \to 1$.

VII. Conclusion

In this paper, we have proposed generalized superimposed training scheme with low complexity processing instead of employing zero forcing. We have analyzed the channel estimation MSE and sum rates components for localized and centralized scenarios, and obtained the necessary condition 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 have shown the superior performance of proposed superimposed scheme both in terms of channel estimation and data detection. It has also demonstrated reduction in interference powers, when centralized processing is used.

Appendix

A. Localized processing with GSP symbols

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

$$P_{lk,S} = \mathbb{E}\left\| \|\hat{\mathbf{h}}_{lk}\|_2^2 \mathbf{s}_k \right\|_2^2 = \mathbb{E}\|\hat{\mathbf{h}}_{lk}\|_2^4 \cdot \mathbb{E}\|\mathbf{s}_k\|_2^2 \tag{48}$$

$$= \mathbb{E}\|\mathbf{h}_{lk} + \Delta_{lk}\|_2^4 = \left(\alpha_{lk} + \beta_{lk}\right)^2 \left(M^2 + M\right). \tag{49}$$

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

$$P_{lk,SI} = \mathbb{E}\left\| \mathbf{s}_k \left(\mathbf{h}_{lk}^H \hat{\mathbf{h}}_{lk} - \hat{\mathbf{h}}_{lk}^H \hat{\mathbf{h}}_{lk} \right) \right\|_2^2$$

$$\overset{(a)}{=} \mathbb{E}\left\| -\Delta_k^H \left(\mathbf{h}_{lk} + \Delta_{lk} \right) \mathbf{s}_k \right\|_2^2 \tag{50}$$

$$\overset{(b)}{=} \mathbb{E}\left| \Delta_{lk}^H \left(\mathbf{h}_{lk} + \Delta_{lk} \right) \right|^2 \overset{(c)}{=} \mathbb{E}\left| \Delta_{lk}^H \mathbf{h}_{lk} \right|^2 + \mathbb{E}\left| \Delta_{lk}^H \Delta_{lk} \right|^2$$

$$\overset{(d)}{=} \beta_{lk}\alpha_{lk} M + \left(M^2 + M\right) \alpha_{lk}^2 \tag{51}$$

$$= \alpha_{lk}^2 M^2 \left(1 + \frac{1}{M} + \frac{\beta_{lk}}{M\alpha_{lk}} \right), \tag{52}$$

where in (a), $\hat{\mathbf{h}}_{lk}^H = \mathbf{h}_{lk} + \Delta_{lk}$ is substituted; in (b) and (c), the facts that the CSI error Δ_{lk} and $\mathbf{h}_{lk}$ are independent and zero mean, are used; in (d), $\mathbb{E}\left|\Delta_{lk}^H \mathbf{h}_{lk}\right|^2 = \beta_{lk}\mathbb{E}\Delta_k^H \Delta_k = \beta_{lk}\alpha_{lk} M$, and the second moment value $\alpha_{lk}^2\left(M^2+M\right)$ of Chi-squared distribution $\Delta_{lk}^H \Delta_{lk} \sim \frac{\alpha_{lk}}{2}\chi_{2M}^2(0)$ is used.

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

$$P_{lk,CI} = \frac{1}{1-\lambda}\mathbb{E}\left\| \sum_{i\neq k} \frac{\mathbf{Z}_k^H \mathbf{x}_i}{T\sqrt{P}} \mathbf{h}_{li}^H \hat{\mathbf{h}}_{lk} \right\|_2^2$$

$$= \frac{1}{1-\lambda} \sum_{i\neq k}\sum_{j\neq k} tr\mathbb{E}\left[\frac{\mathbf{Z}_k^H \mathbf{x}_i}{T\sqrt{P}} \frac{\mathbf{x}_j^H \mathbf{Z}_k}{T\sqrt{P}} \hat{\mathbf{h}}_{lk}^H \mathbf{h}_{lj} \mathbf{h}_{li}^H \hat{\mathbf{h}}_{lk} \right] \tag{53}$$

$$\overset{(a)}{=} \frac{1}{1-\lambda} \sum_{i\neq k} \mathbb{E}\left\| \frac{\mathbf{Z}_k^H \mathbf{x}_i}{T\sqrt{P}} \right\|_2^2 \mathbb{E}\left| \mathbf{h}_{li}^H \hat{\mathbf{h}}_{lk} \right|^2$$

$$\overset{(b)}{=} \frac{1}{1-\lambda} \sum_{i\neq k} c_{ki} \left(\beta_{li}\beta_{lk} M + \beta_{li}\alpha_{lk} M + \zeta_{ik} M^2 \beta_{li}^2 \right) \tag{54}$$

$$= \frac{M}{1-\lambda} \sum_{i\neq k} c_{ki}\beta_{li} \left(\beta_{lk} + \alpha_{lk} + \zeta_{ik} M \beta_{li} \right), \tag{55}$$

where in (a), the fact that $\mathbf{h}_{lj}$ and $\mathbf{h}_{li}$ are zero mean and uncorrelated vectors for all $i \neq j$ is used; in (b), the value from (7) is used, and $\mathbb{E}\left|\mathbf{h}_{li}^H \hat{\mathbf{h}}_{lk}\right|^2 = \mathbb{E}\left|\mathbf{h}_{li}^H \mathbf{h}_{lk}\right|^2 + \mathbb{E}\left|\mathbf{h}_{li}^H \Delta_{lk}\right|^2$ with $\mathbb{E}\left|\mathbf{h}_{li}^H \mathbf{h}_{lk}\right|^2 = \beta_{li}\beta_{lk} M$ is used. For (d), the second term ζ_{ljk} is obtained for $j \neq k$ as $\mathbb{E}\left|\mathbf{h}_{lj}^H \Delta_{lk}\right|^2 =$

$$\mathbb{E}\Big| \sum_{i\in\mathcal{C}_{\underline{k}}\setminus\{k\}} \mathbf{h}_{lj}^H \mathbf{h}_{li} + \sqrt{\frac{1-\lambda}{\lambda}} \sum_{i\in\mathcal{Z}_k} \mathbf{h}_{lj}^H \mathbf{h}_{li} s_{(ik)}^* + \frac{\mathbf{h}_{lj}^H \mathbf{W}\mathbf{p}_k}{T\sqrt{P\lambda}} \Big|^2$$

$$= \sum_{i\in\mathcal{C}_{\underline{k}}\setminus\{k\}} \mathbb{E}\left|\mathbf{h}_{lj}^H \mathbf{h}_{li}\right|^2 + \frac{1-\lambda}{\lambda d} \sum_{i\in\mathcal{Z}_k} q_{ik}\mathbb{E}\left|\mathbf{h}_{lj}^H \mathbf{h}_{li}\right|^2 + \mathbb{E}\left| \frac{\mathbf{h}_{lj}^H \mathbf{W}\mathbf{p}_k}{T\sqrt{P\lambda}} \right|^2$$

$$= \sum_{i\in\mathcal{C}_{\underline{k}}\setminus\{k\}} \beta_{li}\beta_{lj} M + \frac{1-\lambda}{\lambda d} \sum_{i\in\mathcal{Z}_k} q_{ik}\beta_{li}\beta_{lj} M + \beta_{lj}\frac{\sigma^2 M}{TP\lambda} + M^2\beta_{lj}^2 \left(1_{j\in\mathcal{C}_{\underline{k}}\setminus\{k\}} + \frac{1-\lambda}{\lambda d} 1_{j\in\mathcal{Z}_k} \right)$$

$$= \beta_{lj}\alpha_{lk} M + M^2 \beta_{lj}^2 \zeta_{jk},$$

where $\zeta_{jk} = 1_{j\in\mathcal{C}_{\underline{k}}\setminus\{k\}} + \frac{1-\lambda}{\lambda d} 1_{j\in\mathcal{Z}_k}$; the indicator function denote that ones of the disjoint sets, $\mathcal{C}_{\underline{k}} \setminus \{k\}$ and $\mathcal{Z}_k$, may contain the index j. For $j = i$, the value can be given as $\mathbb{E}\left|\mathbf{h}_{lj}^H \mathbf{h}_{lj}\right|^2 = \left(M^2 + M\right)\beta_{lj}^2$, where $M^2\beta_{lj}^2$ is used in the indicator function.

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

$$
\begin{aligned}
P_{lk,N}\times T^2P(1-\lambda) &= \mathbb{E}\left\|\mathbf{Z}_k^H\mathbf{W}_l^H\hat{\mathbf{h}}_{lk}\right\|_2^2 = \mathbb{E}\|\mathbf{Z}_k^H\mathbf{W}_l^H \\
&\times\Big(\sum_{i\in\mathcal{C}_{\underline{k}}}\mathbf{h}_{li}+\frac{\mathbf{W}_l\mathbf{p}_k}{T\sqrt{P\lambda}}+\sqrt{\frac{1-\lambda}{\lambda}}\sum_{i\in\mathcal{Z}_k}\mathbf{h}_{li}s^*_{(ik)}\Big)\|_2^2 \\
&= \sum_{i\in\mathcal{C}_{\underline{k}}}\mathbb{E}\left\|\mathbf{Z}_k^H\mathbf{W}_l^H\mathbf{h}_{li}\right\|_2^2+\frac{1-\lambda}{\lambda d}\sum_{i\in\mathcal{Z}_k}q_{ik}\mathbb{E}\left\|\mathbf{Z}_k^H\mathbf{W}_l^H\mathbf{h}_{li}\right\|_2^2 \\
&\quad+\mathbb{E}\left\|\frac{\mathbf{Z}_k^H\mathbf{W}_l^H\mathbf{W}_l\mathbf{p}_k}{T\sqrt{P\lambda}}\right\|_2^2 \\
&= \sum_{i\in\mathcal{C}_{\underline{k}}}\beta_{li}\sigma^2MdT+\frac{1-\lambda}{\lambda}\sum_{i\in\mathcal{Z}_k}q_{ik}\beta_{li}\sigma^2MT \\
&\quad+\frac{1}{T^2P\lambda}\sigma^4MTd\left(T+M+1\right) \\
&= \sigma^2MTd\Bigg[\beta_{lk}+\sum_{i\in\mathcal{C}_{\underline{k}}\setminus\{k\}}\beta_{li}+\frac{1-\lambda}{\lambda d}\sum_{i\in\mathcal{Z}_k}q_{ik}\beta_{li} \\
&\quad+\frac{\sigma^2}{TP\lambda}\left(1+\frac{M+1}{T}\right)\Bigg] \\
&= \sigma^2MTd\left[\beta_{lk}+\alpha_{lk}+\frac{\sigma^2(M+1)}{T^2P\lambda}\right], \qquad (56)
\end{aligned}
$$

where the first and second term is simplified as

$$
\begin{aligned}
&\mathbb{E}\left\|\mathbf{Z}_k^H\mathbf{W}_l^H\mathbf{h}_{lk}\right\|_2^2 = \beta_{lk}tr\left[\mathbf{Z}_k^H\mathbb{E}\left\{\mathbf{W}_l^H\mathbf{W}_l\right\}\mathbf{Z}_k\right] \\
&= tr(\mathbf{Z}_k^H\mathbf{Z}_k)\sigma^2M\beta_{lk} = \beta_{lk}\sigma^2MdT,
\end{aligned}
$$

and the third term is given for (say) $\mathbf{W}_\star=[\bar{\mathbf{w}}_1,\ldots,\bar{\mathbf{w}}_T]$ and $\left[\mathbf{W}_\star^H\mathbf{W}_\star\right]_{i,j}=\bar{\mathbf{w}}_i^H\bar{\mathbf{w}}_j$ as

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

with $\mathcal{D}\left(\mathbf{p}_*\mathbf{p}_*^H\right)=\mathbf{I}_T$.

B. Centralized processing with GSP symbols

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

$$
\begin{aligned}
&\mathbb{E}\left\|\|\hat{\mathbf{h}}_k\|_2^2\mathbf{s}_k\right\|_2^2 = \mathbb{E}\|\hat{\mathbf{h}}_k\|_2^4\cdot\mathbb{E}\left\|\mathbf{s}_k\right\|_2^2 = \mathbb{E}\|\mathbf{h}_k+\Delta_k\|_2^4 \\
&= \sum_{l=1}^L\sum_{m=1}^L\mathbb{E}\|\mathbf{h}_{lk}+\Delta_{lk}\|_2^2\|\mathbf{h}_{mk}+\Delta_{mk}\|_2^2 \qquad (57) \\
&= \sum_{l=1}^L\sum_{m=1,m\neq l}^L\mathbb{E}\|\mathbf{h}_{lk}+\Delta_{lk}\|_2^2\cdot\mathbb{E}\|\mathbf{h}_{mk}+\Delta_{mk}\|_2^2 \\
&\quad+\sum_{l=m=1}^L\mathbb{E}\|\mathbf{h}_{lk}+\Delta_{lk}\|_2^4 \qquad (58) \\
&= \sum_{l=1}^L\sum_{m=1,m\neq l}^L M^2\left(\alpha_{lk}+\beta_{lk}\right)\left(\alpha_{mk}+\beta_{mk}\right) \\
&\quad+\sum_{l=1}^L\left(\alpha_{lk}+\beta_{lk}\right)^2\left(M^2+M\right) \qquad (59) \\
&= \sum_{l=1}^L\sum_{m=1}^L M^2\left(\alpha_{lk}+\beta_{lk}\right)\left(\alpha_{mk}+\beta_{mk}\right) \\
&\quad+\sum_{l=1}^L\left(\alpha_{lk}+\beta_{lk}\right)^2M \qquad (60) \\
&= M^2L^2\left(\alpha_k+\beta_k\right)^2+ML\times\frac{1}{L}\sum_{l=1}^L\left(\alpha_{lk}+\beta_{lk}\right)^2. \qquad (61)
\end{aligned}
$$

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

$$
\begin{aligned}
P_{k,SI} &= \mathbb{E}\left\|\mathbf{s}_k\left(\mathbf{h}_k^H\hat{\mathbf{h}}_k-\hat{\mathbf{h}}_k^H\hat{\mathbf{h}}_k\right)\right\|_2^2 \\
&\overset{(a)}{=}\mathbb{E}\left\|-\Delta_k^H\left(\mathbf{h}_k+\Delta_k\right)\mathbf{s}_k\right\|_2^2\overset{(b)}{=}\mathbb{E}\left|\Delta_k^H\left(\mathbf{h}_k+\Delta_k\right)\right|^2 \\
&\overset{(c)}{=}\mathbb{E}\left|\Delta_k^H\mathbf{h}_k\right|^2+\mathbb{E}\left|\Delta_k^H\Delta_k\right|^2 \qquad (62) \\
&\overset{(d)}{=}\sum_{l=1}^L\beta_{lk}\alpha_{lk}M+M^2L^2\alpha_k^2+M\sum_{l=1}^L\alpha_{lk}^2 \qquad (63) \\
&= \alpha_k^2M^2L^2\left(\frac{1}{L}\sum_{l=1}^L\frac{\beta_{lk}\alpha_{lk}}{ML\alpha_k^2}+1+\frac{1}{ML}\sum_{l=1}^L\frac{\alpha_{lk}^2}{L\alpha_k^2}\right), \qquad (64)
\end{aligned}
$$

where in (a), $\hat{\mathbf{h}}_k^H=\mathbf{h}_k+\Delta_k$ is substituted; in (b) and (c), the fact that the zero-mean vectors Δ_k and $\mathbf{h}_k$ are independent, is used; in (d), $\mathbb{E}\left|\Delta_k^H\mathbf{h}_k\right|^2=tr\mathbb{E}\mathbf{B}_k\Delta_k\Delta_k^H=tr\left(\mathbf{B}_k\mathbf{A}_k\right)=\sum_{l=1}^L\beta_{lk}\alpha_{lk}M$, and the second moment value $\alpha_{lk}\left(M^2+M\right)$ of Chi-squared distribution $\Delta_{lk}^H\Delta_{lk}\sim\frac{\alpha_{lk}}{2}\chi^2_{2M}(0)$ is used, that is, similar to (61), we get $\mathbb{E}\left|\Delta_k^H\Delta_k\right|^2=M^2\left(\sum_{l=1}^L\alpha_{lk}\right)^2+M\sum_{l=1}^L\alpha_{lk}^2$ with $\alpha_k=\frac{1}{L}\sum_l\alpha_{lk}$.

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

$$P_{k,CI} = \frac{1}{1-\lambda}\mathbb{E}\left\|\sum_{i\neq k}\frac{\mathbf{Z}_k^H\mathbf{x}_i}{T\sqrt{P}}\mathbf{h}_i^H\hat{\mathbf{h}}_k\right\|_2^2$$
$$= \frac{1}{1-\lambda}\sum_{i\neq k}\sum_{j\neq k} tr\mathbb{E}\left[\frac{\mathbf{Z}_k^H\mathbf{x}_i}{T\sqrt{P}}\frac{\mathbf{x}_j^H\mathbf{Z}_k}{T\sqrt{P}}\hat{\mathbf{h}}_k^H\mathbf{h}_j\mathbf{h}_i^H\hat{\mathbf{h}}_k\right] \quad (65)$$
$$\overset{(a)}{=} \frac{1}{1-\lambda}\sum_{i\neq k}\mathbb{E}\left\|\frac{\mathbf{Z}_k^H\mathbf{x}_i}{T\sqrt{P}}\right\|_2^2\mathbb{E}\left|\mathbf{h}_i^H\hat{\mathbf{h}}_k\right|^2 \quad (66)$$
$$\overset{(b)}{=} \frac{M}{1-\lambda}\sum_{i\neq k}c_{ki}\sum_l\beta_{li}\left(\beta_{lk}+\alpha_{lk}+ML\frac{\beta_i^2}{\beta_{li}}\zeta_{ik}\right), \quad (67)$$

where in (a), the fact that $\mathbf{h}_j$ and $\mathbf{h}_i$ are zero mean and uncorrelated vectors is used; in (b), the value from (7) is used, and $\mathbb{E}\left|\mathbf{h}_i^H\hat{\mathbf{h}}_k\right|^2 = \mathbb{E}\left|\mathbf{h}_i^H\mathbf{h}_k\right|^2 + \mathbb{E}\left|\mathbf{h}_i^H\Delta_k\right|^2$ with $\mathbb{E}\left|\mathbf{h}_i^H\mathbf{h}_k\right|^2 = \sum_l\beta_{li}\beta_{lk}M$ is used. The term ζ_{ik} is obtained for $j\neq k$ as $\mathbb{E}\left|\mathbf{h}_j^H\Delta_k\right|^2 =$

$$\mathbb{E}\left|\sum_{i\in\mathcal{C}_{\underline{k}}\setminus\{k\}}\mathbf{h}_j^H\mathbf{h}_i + \sqrt{\frac{1-\lambda}{\lambda}}\sum_{i\in\mathcal{Z}_k}\mathbf{h}_j^H\mathbf{h}_i s_{(ik)}^* + \frac{\mathbf{h}_j^H\mathbf{W}\mathbf{p}_k}{T\sqrt{P\lambda}}\right|^2$$
$$= \sum_{i\in\mathcal{C}_{\underline{k}}\setminus\{k\}}\mathbb{E}\left|\mathbf{h}_j^H\mathbf{h}_i\right|^2 + \frac{(1-\lambda)}{\lambda d}\sum_{i\in\mathcal{Z}_k}q_{ik}\mathbb{E}\left|\mathbf{h}_j^H\mathbf{h}_i\right|^2 + \mathbb{E}\left|\frac{\mathbf{h}_j^H\mathbf{W}\mathbf{p}_k}{T\sqrt{P\lambda}}\right|^2$$
$$= \sum_{i\in\mathcal{C}_{\underline{k}}\setminus\{k\}}\sum_l\beta_{li}\beta_{lj}M + \frac{1-\lambda}{\lambda d}\sum_{i\in\mathcal{Z}_k}q_{ik}\sum_l\beta_{li}\beta_{lj}M + \sum_l\beta_{lj}\frac{\sigma^2M}{TP\lambda} + M^2L^2\beta_j^2\left(1_{j\in\mathcal{C}_{\underline{k}}\setminus\{k\}} + \frac{1-\lambda}{\lambda d}1_{j\in\mathcal{Z}_k}\right)$$
$$= \sum_l\beta_{lj}\alpha_{lk}M + M^2L^2\beta_j^2\zeta_{jk},$$

where $\zeta_{jk} = 1_{j\in\mathcal{C}_{\underline{k}}\setminus\{k\}} + \frac{1-\lambda}{\lambda d}1_{j\in\mathcal{Z}_k}$; the indicator function denote that ones of the disjoint sets, $\mathcal{C}_{\underline{k}}\setminus\{k\}$ and $\mathcal{Z}_k$, may contain the index j. Thus, for $j=i$, the value can be given as (similar to (61)) $\mathbb{E}\left|\mathbf{h}_j^H\mathbf{h}_j\right|^2 = M^2L^2\beta_j^2 + M\sum_l\beta_{lj}^2$, where the first value term $M^2L^2\beta_j^2$ is mentioned with the indicator function in the previous equation.

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

$$P_{k,N}\times T^2P(1-\lambda) = \mathbb{E}\left\|\mathbf{Z}_k^H\mathbf{W}^H\hat{\mathbf{h}}_k\right\|_2^2 = \mathbb{E}\|\mathbf{Z}_k^H\mathbf{W}^H \times\Big(\sum_{i\in\mathcal{C}_{\underline{k}}}\mathbf{h}_i + \sqrt{\frac{1-\lambda}{\lambda}}\sum_{i\in\mathcal{Z}_k}\mathbf{h}_i s_{(ik)}^* + \frac{\mathbf{W}\mathbf{p}_k}{T\sqrt{P\lambda}}\Big)\|_2^2$$
$$= \sum_{i\in\mathcal{C}_{\underline{k}}}\mathbb{E}\left\|\mathbf{Z}_k^H\mathbf{W}^H\mathbf{h}_i\right\|_2^2 + \frac{1-\lambda}{\lambda d}\sum_{i\in\mathcal{Z}_k}q_{ik}\mathbb{E}\left\|\mathbf{Z}_k^H\mathbf{W}^H\mathbf{h}_i\right\|_2^2 + \mathbb{E}\left\|\frac{\mathbf{Z}_k^H\mathbf{W}^H\mathbf{W}\mathbf{p}_k}{T\sqrt{P\lambda}}\right\|_2^2 \quad (68)$$
$$= \sum_{i\in\mathcal{C}_{\underline{k}}}\sum_{l=1}^L\beta_{ik}\sigma^2MdT + \frac{1-\lambda}{\lambda}\sum_{i\in\mathcal{Z}_k}q_{ik}\sum_{l=1}^L\beta_{ik}\sigma^2MT + \frac{1}{T^2P\lambda}\sigma^4MLTd(T+ML+1) \quad (69)$$
$$= \sigma^2MLTd\Bigg[\sum_{l=1}^L\frac{\beta_{lk}}{L} + \sum_{i\in\mathcal{C}_{\underline{k}}\setminus\{k\}}\sum_{l=1}^L\frac{\beta_{li}}{L} + \frac{1-\lambda}{\lambda d}\sum_{i\in\mathcal{Z}_k}q_{ik}\sum_{l=1}^L\frac{\beta_{li}}{L} + \frac{\sigma^2}{TP\lambda}\left(1+\frac{ML+1}{T}\right)\Bigg]$$
$$= \sigma^2MLTd\left[\beta_k+\alpha_k+\frac{\sigma^2(ML+1)}{T^2P\lambda}\right], \quad (70)$$

where the first and second term is simplified as

$$\mathbb{E}\left\|\mathbf{Z}_k^H\mathbf{W}^H\mathbf{h}_k\right\|_2^2 = tr\left[\mathbf{Z}_k^H\mathbb{E}\left\{\mathbf{W}^H\mathbf{B}_k\mathbf{W}\right\}\mathbf{Z}_k\right]$$
$$= tr(\mathbf{Z}_k^H\mathbf{Z}_k)tr(\mathbf{B}_k)\sigma^2 = \sum_{l=1}^L\beta_{lk}\sigma^2MdT,$$

and the third term can be obtained in a similar manner as in (56), $\mathbb{E}\left\|\mathbf{Z}_k^H\mathbf{W}^H\mathbf{W}\mathbf{p}_k\right\|_2^2 = \sigma^4ML\cdot Td(T+ML+1)$.

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 Trends in Signal Processing*, vol. 14, no. 3-4, pp. 162–472, 2021. [Online]. Available: http://dx.doi.org/10.1561/2000000109

[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 Transactions on Signal Processing*, vol. 63, no. 18, pp. 4784–4799, 2015.

[5] H. He, J. Xue, T. Ratnarajah, F. A. Khan, and C. B. Papadias, "Modeling and analysis of cloud radio access networks using matÃ©rn hard-core point processes," *IEEE Transactions on Wireless Communications*, vol. 15, no. 6, pp. 4074–4087, 2016.

[6] 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.

[7] E. Bjornson 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.

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

[9] 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.

[10] 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.

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

[12] 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.

[13] 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.

[14] 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.

[15] J. Zhang, Y. Wei, E. Bjornson, 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.

[16] 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.

[17] 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.

[18] 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.

[19] 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.

[20] S. Ohno and G. Giannakis, "Optimal training and redundant precoding for block transmissions with application to wireless OFDM," *IEEE Transactions on Communications*, vol. 50, no. 12, pp. 2113–2123, Dec 2002.

[21] 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.

[22] 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.

[23] ——, "On the effect of receiver estimation error upon channel mutual information," *IEEE Transactions on Signal Processing*, vol. 54, no. 2, pp. 459–472, Feb 2006.

[24] N. N. Tran, D. H. Pham, H. D. Tuan, and H. H. Nguyen, "Orthogonal affine precoding and decoding for channel estimation and source detection in MIMO frequency-selective fading channels," *IEEE Transactions on Signal Processing*, vol. 57, no. 3, pp. 1151–1162, March 2009.

[25] C.-T. Chiang and C. C. Fung, "Robust training sequence design for spatially correlated MIMO channel estimation," *IEEE Transactions on Vehicular Technology*, vol. 60, no. 7, pp. 2882–2894, Sep. 2011.

[26] K.-C. Chan, C.-P. Li, C.-Y. Hung, and W.-J. Huang, "A precoding scheme for eliminating data identification problem in single carrier system using data-dependent superimposed training," *IEEE Access*, vol. 7, pp. 45 930–45 939, 2019.

[27] J. C. Estrada-Jimenez and M. J. Fernandez-Getino Garcia, "Partial-data superimposed training with data precoding for OFDM systems," *IEEE Transactions on Broadcasting*, vol. 65, no. 2, pp. 234–244, June 2019.

[28] 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.

[29] X. Meng, J. K. Tugnait, and S. He, "Iterative joint channel estimation and data detection using superimposed training: Algorithms and performance analysis," *IEEE Transactions on Vehicular Technology*, vol. 56, no. 4, pp. 1873–1880, July 2007.

[30] S. He, J. K. Tugnait, and X. Meng, "On superimposed training for MIMO channel estimation and symbol detection," *IEEE Transactions on Signal Processing*, vol. 55, no. 6, pp. 3007–3021, June 2007.

[31] X. Dai, H. Zhang, and D. Li, "Linearly time-varying channel estimation for MIMO/OFDM systems using superimposed training," *IEEE Transactions on Communications*, vol. 58, no. 2, pp. 681–693, February 2010.

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

[33] S. M. S. Sadough and Z. Chamideh, "An improved variational inference approach to iterative OFDM receiver design for superimposed training-based af relay networks," *IEEE Transactions on Vehicular Technology*, vol. 67, no. 3, pp. 2243–2253, March 2018.

[34] G. Yang, T. Liu, H. Ding, Q. Yan, and X. Wang, "Joint channel estimation and generalized approximate messaging passing-based equalization for underwater acoustic communications," *IEEE Access*, vol. 9, pp. 56 757–56 764, 2021.

[35] G. Yang, Q. Guo, H. Ding, Q. Yan, and D. D. Huang, "Joint message-passing-based bidirectional channel estimation and equalization with superimposed training for underwater acoustic communications," *IEEE Journal of Oceanic Engineering*, vol. 46, no. 4, pp. 1463–1476, Oct 2021.