# FD-MIMO via Pilot-Data Superposition: Tensor-Based DOA Estimation and System Performance

Zhou Zhou, Lingjia Liu, and Jianzhong Zhang

***Abstract*—Increased pilot overhead is one of the major issues for 5G full-dimension MIMO (FD-MIMO) systems. In this paper, we introduce the FD-MIMO system using pilot-data superposition to reduce the uplink pilot overhead and investigate the impact of superimposed pilots on the overall network performance. The pilot-data superposition can relatively increase the resources allocated to the uplink data transmission. However, in time division duplex (TDD) systems, it also negatively impacts the downlink throughput since the relatively reduced pilot power can affect the estimation of the channel state information (CSI) which is used for the downlink precoding. To improve the CSI estimation, the intrinsic tensor feature of the FD-MIMO channel is exploited in our method. The introduced CSI estimation algorithm is designed through the expectation-maximization (EM) framework via tensor as the processing data structure. Furthermore, Cramér-Rao lower bound (CRLB) is utilized as a metric in our evaluation. The overall system achievable rate which is a weighted sum of the uplink and downlink throughput is investigated to reveal the fundamental trade-off between the uplink and downlink transmission in the TDD-based FD-MIMO system. In our simulation, the results demonstrate the superior performance of the introduced strategy as opposed to the conventional orthogonal pilot-data approach. Meanwhile, the pilot power allocation, as well as different transmitting and receiving strategies are investigated to offer various trade-off points between the uplink and downlink transmission.**

***Index Terms*—FD-MIMO, DOA estimation, tensor decomposition, EM algorithm and MIMO precoding.**

## I. Introduction

MASSIVE-MIMO [1] is a key enabling technology in 5G system. With a large number of antennas configured at the base station (BS), the system can perform high spectral and energy efficiency [2]. To circumvent the form factor limitation at the BS, full-dimension MIMO (FD-MIMO) [3], [4] is introduced in 3GPP as one of the realization of massive-MIMO in sub-6GHz spectrum bands. In FD-MIMO, active antenna elements are placed in a 2D panel to enable 3D beamforming. To achieve the gains promised by FD-MIMO, channel state information at the transmitter (CSIT) is crucial for conducting the corresponding precoding/beamforming operation at downlink transmission.

For time division duplex (TDD) systems, CSIT for the downlink transmission can be obtained through the uplink channel estimation according to channel reciprocity. In conventional systems, such as 3GPP LTE/LTE-Advanced, uplink pilots and uplink data are assigned to separate resource elements. However, as the number of users increases, the uplink pilot overhead becomes unaffordable, which limits the achievability of 5G FD-MIMO systems [3]. To overcome this issue, the pilot-data superposition, in which the pilots are superimposed with the transmitted data, is introduced for FD-MIMO systems. In this way, although the resources for data transmission are relatively increased, additional interference for uplink channel estimation is involved in. In other words, the relatively reduced pilot power inevitably sacrifices the accuracy of CSIT estimation, where the resulting data demodulation is conducted through error-corrupted channel spaces. Therefore, it is unknown whether this introduced superposition scheme can increase uplink throughput or not.

For the pilot-data superposition system, it is meaningful to look at the overall system achievable rate which is defined as

$$R = \kappa_{DL} R^{DL} + \kappa_{UL} R^{UL} \tag{1}$$

where $R^{DL}$ denotes the downlink transmission rate; $R^{UL}$ is the uplink rate which is summed up by the rate of pure uplink data and data with superposition pilots interference; $\kappa_{DL}$ and $\kappa_{UL}$ respectively represent the weight of uplink and downlink transmission priority. Accordingly, we are able to characterize the trade-off between uplink and downlink throughput using different priorities. To be specific, $\kappa_{DL}$ and $\kappa_{UL}$ can be chosen such that they are proportional to the number of allocated resource elements. Previously, [5], [6] conclude that the achievable transmission rate of pilot-data superposition systems is superior to the system with orthogonal pilots-data when pilot contamination exists. Meanwhile, [7] shows the pilot-data superposition system can offer substantial gains of energy and spectrum efficiency. For the system having a large number of communication links, [8] demonstrates that the incorporated superimposed pilot scheme performs better than conventional strategies when the channel coherent time is limited.

Manuscript received February 6, 2019; revised June 18, 2019; accepted August 11, 2019. Date of publication August 29, 2019; date of current version September 20, 2019. The work of Z. Zhou and L. Liu was supported by U.S. National Science Foundation under Grant ECCS-1811497. The guest editor coordinating the review of this article and approving it for publication was Prof. Feifei Gao. *(Corresponding author: Lingjia Liu.)*

Z. Zhou and L. Liu are with the Bradley Department of Electrical and Computer Engineering, Virginia Tech., Blacksburg, VA 24061 USA (e-mail: zhou89@vt.edu; ljliu@ieee.org).

J. Zhang is with the Standards and Mobility Innovation Lab., Samsung Research America, Richardson, TX 75082 USA (e-mail: jianzhong.z@samsung.com).

Digital Object Identifier 10.1109/JSTSP.2019.2938488

Note that the existing works of pilot-data superposition systems mainly focus on general massive MIMO systems with ideal assumptions such as Rayleigh fading channel and infinite number of antennas. However, these assumptions are not always true for practical systems. Furthermore, there is no study on applying the pilot-data superposition scheme to FD-MIMO systems which have been standardized in 3GPP LTE-Advanced Pro. This motivates us to investigate our discussion on the FD-MIMO system. To this end, we primarily consider the algorithm of CSIT acquisition since it is a function of the data-pilot power ratio which implicitly determines the overall system rate. Traditionally, the CSIT is obtained by inferring the entries of underlying MIMO channel matrix. Alternatively, the channel can be characterized by a few numbers of parameters, such as direction of arrivals (DOAs), direction of departures (DODs) and delay spread, etc. [9]. Therefore, the channel estimation can be equivalently conducted by retrieving these channel parameters, especially through the estimation of DOAs [10].

In [11], the CSI in FD-MIMO systems is estimated through subspace methods [12], [13]. Specifically, it characterized the downlink rate of angular beamforming as a function of the DOA estimation error. Furthermore, in [14], the estimation method is extended to the joint estimation of DOAs and delays of wideband channels. In [15], a variational Bayesian-based method is utilized to infer the DOAs with outliers in the observation. Very recently, [16] introduced an estimation and feedback framework for the estimation of DOAs/DODs pairs in dual-polarized MIMO systems. In [17], [18], a spatial 2-D fast Fourier transform-based DOA estimation method has been introduced to unify the uplink and downlink channel estimation. Meanwhile, [19] proposed the DOA estimation in dual-polarized massive MIMO systems using tensor decomposition. However, all of the above works are designated to single-user MIMO systems with orthogonal pilot-data structures. For the uplink multi-user DOA estimation with pilot-data superposition, the following two challenges need to be addressed

- The introduced algorithm needs to be targeted at low signal-to-interference-plus-noise ratio (SINR) regime, because of the strong interference, such as self-data interference, intra-cell pilot-data interference, inter-cell pilot-data interference, and pilot contamination.
- The algorithm is required to be able to pair estimated angles, since the pilot signals coming from all users are mixed in the uplink, and each path has both elevation and azimuth angles.

To efficiently address these issues, we introduce a DOA estimation algorithm by solving the maximum likelihood estimation using expectation-maximization (EM) algorithm [20]. We note that the spatial transmission channel of FD-MIMO systems can be naturally expressed as multi-way arrays (a tensor) [21], where the modes of the tensor correspond to the antenna arrays at BS and MSs. In general, it is not necessary to use this tensor structure to design the DOA estimation algorithm, such as [11], [14], [22] where the channel vectorization is first conducted at the received signal to construct a two-dimensional signal model. However, from our previous work [23], [24], it shows this channel vectorization can lead to high computational complexity in the DOA estimation [23], [24]. Therefore, we continue to leverage the underlying channel tensor structure in the algorithm design. As a result, our developed algorithm iteratively refines the DOA estimation by a tensor decomposition at each M-step of the EM algorithm. This tensor decomposition procedure converts the multiple-dimensional likelihood estimation to several single dimensional parameter searchings, which can significantly improve the efficiency and performance [25]. Furthermore, the derived EM algorithm can be interpreted as the following interference cancellation procedure. In the E-step of the EM algorithm, the interference is estimated via a linear minimum mean square error estimation (LMMSE) according to the estimated DOAs from the previous iteration. In the M-step, the DOAs are updated by using the observation subtracted from the estimated interference at the E-step. Compared with the one-stage tensor decomposition method introduced in [23], [24], the algorithm framework introduced in this paper is more general and robust to the interference.

Meanwhile, the Cramér-Rao lower bound (CRLB) as a function of the power allocation is derived, serving as a benchmark to evaluate the DOA estimation performance. Simulation results suggest that our introduced channel estimation algorithm can perform close to the CRLB when the pilot-interference power is low. Moreover, to further evaluate our introduced algorithm, various downlink multiuser FD-MIMO precoding strategies are introduced. According to (1), the uplink-downlink achievable rate trade-off is characterized through the power allocation on superimposed pilots, uplink-downlink resources split ratio and DOAs estimation error. The conventional orthogonal pilot-data transmission is incorporated into our framework as a special case where the power of superposition data is zero. Furthermore, in our numerical results, the optimal power allocations of the superimposed pilot for different FD-MIMO precoding strategies are obtained. Overall, our contributions in this paper are summarized as follows:

- A tensor decomposition based DOA estimation algorithm is introduced in the multi-user FD-MIMO system with the pilot-data superposition.
- Theoretical analyses related to our introduced algorithm are characterized by the uniqueness of the M-step decomposition, CRLB of the DOA estimation and algorithm complexity.
- Extensive performance evaluations are conducted to the uplink and downlink throughput trade-off, which offers guidance to the system design.

The paper structure is organized as follows: Section II introduces the system model of pilot-data superposition of multi-user FD-MIMO systems. The EM algorithm based channel estimation is derived in Section III. The Cramér-Rao lower bound and the uniqueness of the tensor decomposition in the M-step is discussed in Section IV-A. In Section V, we characterize the overall system achievable rate to reveal the uplink-downlink trade-off in the FD-MIMO system. Section VI contains the simulation evaluation results. Finally, Section VII concludes the paper.

*Notations:* $\boldsymbol{a}$ stands for a column vector. $\boldsymbol{A}$ is denoted as a matrix. $\boldsymbol{A}^T$, $\boldsymbol{A}^H$ and $\boldsymbol{A}^*$ stand for the transpose, hermitian and

conjugate respectively. The mathcal version $\boldsymbol{\mathcal{A}}$ represents the tensor. $\circ$ stands for the outer product between two vectors, $\otimes$ is the Kronecker product of two matrices, $\odot$ stands for the Khatri-Rao product. $\times_n$ stands for the multi-mode product on the $n$th mode of a tensor, $\boldsymbol{A}_{(n)}$ represents the mode-n unfolding of the tensor $\boldsymbol{\mathcal{A}}$. For more details about tensor, please refer to [21].

## II. System Model

We consider a multi-cell multi-user FD-MIMO system with TDD operation. In each cell, we consider one BS serves $J$ MSs simultaneously. The BS is configured with FD-MIMO, while the MSs are with uniform linear arrays. Moreover, we denote the number of neighboring BSs in each cell as $G$.

### A. Channel Model

The uplink channel from the $j$th MS at the $g$th BS to the $i$th BS can be naturally written as a third order tensor, i.e., $\boldsymbol{\mathcal{H}}_{jg,i} \in \mathbb{C}^{N_{BSa}\times N_{BSe}\times N_{MS}}$, where $N_{BSa}$, $N_{BSe}$ and $N_m$ stand for the number of antennas in BS azimuth direction, BS elevation direction and MS respectively. At the $k$th sub-carrier of one OFDM symbol, the channel tensor can be expressed as,

$$\boldsymbol{\mathcal{H}}_{jg,i}(k) \triangleq \sum_{l=0}^{L_{\mathrm{jg},i}-1} \sqrt{\Lambda_{jg,i}}\boldsymbol{\mathcal{C}}_{jg,i}(l)e^{\frac{-j2\pi k\tau_{jg,i,l}f_s}{N_c}} \tag{2}$$

where $L_{jg,i}$ stands for the number of paths, $\sqrt{\Lambda_{jg,i}}$ represents the pathloss, $f_s$ is the sample rate, $N_c$ is the number of sub-carriers and $\tau_{jg,i,l}$ is the time delay;

$$\boldsymbol{\mathcal{C}}_{jg,i}(l) \triangleq \sum_{p=0}^{P_{jg,i,l}-1} \alpha_{jg,i,l}^{(p)}\boldsymbol{a}\left(u_{jg,i,l}^{(p)}\right)\circ\boldsymbol{a}\left(v_{jg,i,l}^{(p)}\right)\circ\boldsymbol{a}\left(\omega_{jg,i,l}^{(p)}\right) \tag{3}$$

in which $P_{jg,i,l}$ stands for the number of sub-paths derived from the $l$th path, $\alpha_{jg,i,l}^{(p)}$ is the associated path gain, and the steering vectors $\boldsymbol{a}$ are defined as

$$\boldsymbol{a}(u_{jg,i,l}^{(p)}) \triangleq \left[1, e^{-j\pi u_{jg,i,l}^{(p)}}, \ldots, e^{-j\pi(N_{BSa}-1)u_{jg,i,l}^{(p)}}\right]^T \tag{4}$$

$$\boldsymbol{a}(v_{jg,i,l}^{(p)}) \triangleq \left[1, e^{-j\pi v_{jg,i,l}^{(p)}}, \ldots, e^{-j\pi(N_{BSe}-1)v_{jg,i,l}^{(p)}}\right]^T \tag{5}$$

$$\boldsymbol{a}(\omega_{jg,i,l}^{(p)}) \triangleq \left[1, e^{-j\pi\omega_{jg,i,l}^{(p)}}, \ldots, e^{-j\pi(N_m-1)\omega_{jg,i,l}^{(p)}}\right]^T \tag{6}$$

where $u$ and $v$ are the virtual angular representations of the direction of arrival (DOA) at the azimuth and elevation direction respectively [26], and $\omega$ is the virtual direction of departure (DOD). The definition of the virtual angle is $u_{jg,i,l}^{(p)} = 2\Delta/\lambda\sin(\theta_{jg,i,l}^{(p)})$, in which $\theta$ is the actual DOA, $\Delta$ is the antenna spacing and $\lambda$ is the wavelength. Meanwhile, for ease of later discussion, we define the matrix $\boldsymbol{A}(\boldsymbol{u}_{jg,i})$ as

$$\boldsymbol{A}(\boldsymbol{u}_{jg,i}) = \left[\boldsymbol{a}\left(u_{jg,i,0}^{(0)}\right), \ldots, \boldsymbol{a}\left(u_{jg,i,L_{jg,i}-1}^{(P_{jg,i,l}-1)}\right)\right]$$

Similarly, $\boldsymbol{A}(\boldsymbol{v}_{jg,i})$ and $\boldsymbol{A}(\boldsymbol{\omega}_{jg,i})$ are defined in the same way.

### B. Received Signal Model

During the uplink channel estimation stage, the MSs send data symbols $\boldsymbol{X} \in \mathbb{C}^{N_m\times T}$ superimposed by pilot symbols $\boldsymbol{S} \in \mathbb{C}^{N_m\times T}$ to the BS, where $T$ stands for the number of symbols in each block. Therefore, at the $i$th BS and the $k$th sub-carrier, the received signal is expressed as

$$\boldsymbol{\mathcal{Z}}_i(k) = \sum_{g=0}^{G-1}\sum_{j=0}^{J-1}\boldsymbol{\mathcal{H}}_{jg,i}(k)\times_3(\boldsymbol{X}_{jg}(k)+\boldsymbol{S}_{jg}(k)) + \boldsymbol{\mathcal{W}}_i(k) \tag{7}$$

where $\boldsymbol{\mathcal{Z}}_i(k) \in \mathbb{C}^{N_{Ba}\times N_{Be}\times T}$ and $\boldsymbol{\mathcal{W}}_i(k) \in \mathbb{C}^{N_{Ba}\times N_{Be}\times T}$ represents the additive noise. Furthermore, we assume the distributions of $\boldsymbol{X}$ and $\boldsymbol{S}$ are independent to each other. As we can see, both the channel $\boldsymbol{\mathcal{H}}_{jg,i}(k)$ and the received signal $\boldsymbol{\mathcal{Z}}_i(k)$ are able to be represented by tensor. This offers us with an motivation on using tensor based processing in the algorithm design. Besides, MSs usually transmit the pilot and data symbols with precoding. Therefore, $\boldsymbol{X}$ and $\boldsymbol{S}$ are expressed as

$$\boldsymbol{X}_{jg}(k) \triangleq \boldsymbol{P}^*(\boldsymbol{\omega}_{jg,g})\bar{\boldsymbol{X}}_{jg}(k) \tag{8}$$

$$\boldsymbol{S}_{jg}(k) \triangleq \boldsymbol{P}^*(\boldsymbol{\omega}_{jg,g})\bar{\boldsymbol{S}}_{jg}(k) \tag{9}$$

where $\boldsymbol{P}^*(\boldsymbol{\omega}_{jg,g}) \in \mathbb{C}^{N_m\times M}$ represents any angular information based precoding scheme and $M$ stands for the number of transmission streams. For ease of discussion, in the precoding design of each MS, we assume MSs have obtained angular channel knowledge perfectly. In this way, the precoding can be conducted by conjugate beamforming, $\boldsymbol{P}(\boldsymbol{\omega}_{jg,g}) = \boldsymbol{A}(\boldsymbol{\omega}_{jg,g})$, or zero forcing, $\boldsymbol{P}(\boldsymbol{\omega}_{jg,g}) = \boldsymbol{A}^{-1}(\boldsymbol{\omega}_{jg,g})$. In order to characterize the pilot power allocation, we further assume $\bar{\boldsymbol{S}}_{jg}(k) \in \mathbb{C}^{M\times T}$ and $\bar{\boldsymbol{X}}_{jg}(k) \in \mathbb{C}^{M\times T}$ are with the energy constraints $\|\bar{\boldsymbol{s}}_{ni}(k)^H\|_2 = \sqrt{\gamma}$ and $\|\bar{\boldsymbol{x}}_{ni}(k)^H\|_2 = \sqrt{1-\gamma}$, in which $\bar{\boldsymbol{s}}_{ni}(k)^H$ and $\bar{\boldsymbol{x}}_{ni}(k)^H$ respectively represent a row-vector of $\bar{\boldsymbol{S}}_{ni}(k)$ and $\bar{\boldsymbol{X}}_{ni}(k)$.

Now, we consider the CSI estimation from the received tensor $\boldsymbol{\mathcal{Z}}_i(k)$. For the $n$th MS on the $i$th cell, by multiplying $\bar{\boldsymbol{S}}_{ni}^H$ on the received tensor $\boldsymbol{\mathcal{Z}}_i(k)$, we have

$$\boldsymbol{\mathcal{Y}}_{ni}(k) = \boldsymbol{\mathcal{Z}}_i(k)\times_3\bar{\boldsymbol{S}}_{ni}^H + \boldsymbol{\mathcal{W}}_i'(k) \tag{10}$$

which can be further expressed as

$$\begin{aligned}\boldsymbol{\mathcal{Y}}_{ni}(k) = {}& \bar{\boldsymbol{\mathcal{H}}}_{ni,i}(k)\times_3(\sqrt{\gamma(1-\gamma)}\bar{\boldsymbol{\Omega}}_{ni}(k)+\gamma\boldsymbol{I}) \\ &+ \sum_{j\neq n}\bar{\boldsymbol{\mathcal{H}}}_{ji,i}(k)\times_3\left(\sqrt{\gamma(1-\gamma)}\bar{\boldsymbol{\Omega}}_{ji}(k)+\rho_1\gamma\boldsymbol{1}\right) \\ &+ \sum_{g\neq i}\bar{\boldsymbol{\mathcal{H}}}_{ng,i}(k)\times_3\left(\sqrt{\gamma(1-\gamma)}\bar{\boldsymbol{\Omega}}_{ng}(k)+\gamma\boldsymbol{I}\right) \\ &+ \sum_{g\neq i}\sum_{j\neq n}\bar{\boldsymbol{\mathcal{H}}}_{jg,i}(k)\times_3\left(\sqrt{\gamma(1-\gamma)}\bar{\boldsymbol{\Omega}}_{jg}(k)\right. \\ &\left.+\rho_1\gamma\boldsymbol{1}\right)+\boldsymbol{\mathcal{W}}_i'(k)\end{aligned} \tag{11}$$

where $\boldsymbol{I}$ is an identity matrix, $\boldsymbol{1}$ is the matrix with ones everywhere, and

$$\bar{\boldsymbol{\mathcal{H}}}_{ni,i}(k) \triangleq \boldsymbol{\mathcal{H}}_{ni,i}(k)\times_3\boldsymbol{P}^*(\boldsymbol{\omega}_{ni,i}); \tag{12}$$

Similarly, $\bar{\boldsymbol{\mathcal{H}}}_{ji,i}(k)$, $\bar{\boldsymbol{\mathcal{H}}}_{ng,i}(k)$ and $\bar{\boldsymbol{\mathcal{H}}}_{jg,i}(k)$ are defined in the same way; $\bar{\boldsymbol{\Omega}}_{jg}(k) \triangleq \bar{\boldsymbol{X}}_{jg}(k)\bar{\boldsymbol{S}}_{ni}^{H}(k) \in \mathbb{C}^{M\times M}$ is a random matrix whose element-wise definition is $\bar{\boldsymbol{x}}_{jg}(k)^H \bar{\boldsymbol{s}}_{ni}(k) = \varpi_{jg}\sqrt{\gamma(1-\gamma)}$, in which $\varpi_{jg}$ is a complex random variable which is asymptotically approximate to zero due to the independence between pilot and data symbols as $T$ increases; For $\bar{\boldsymbol{\Omega}}_{ni}(k)$, $\bar{\boldsymbol{\Omega}}_{ji}(k)$ and $\bar{\boldsymbol{\Omega}}_{ng}(k)$, they have similar definitions to $\bar{\boldsymbol{\Omega}}_{jg}(k)$; $\rho_1$ denotes the pilot correlation between any two users, i.e., $\bar{\boldsymbol{s}}_{ji}(k)^H \bar{\boldsymbol{s}}_{ni}(k) = \rho_1\gamma$. For ease of discussion, we further define the following notations

$$\boldsymbol{\mathcal{I}}_u(k) \triangleq \sum_{j\neq n} \bar{\boldsymbol{\mathcal{H}}}_{ji,i}(k)\times_3\big(\sqrt{\gamma(1-\gamma)}\bar{\boldsymbol{\Omega}}_{ji}(k) + \gamma\rho_1\mathbf{1}\big) \quad (13)$$

$$\boldsymbol{\mathcal{I}}_p(k) \triangleq \sum_{g\neq i} \bar{\boldsymbol{\mathcal{H}}}_{ng,i}(k)\times_3\big(\sqrt{\gamma(1-\gamma)}\bar{\boldsymbol{\Omega}}_{ng}(k) + \gamma\boldsymbol{I}\big) \quad (14)$$

$$\boldsymbol{\mathcal{I}}_e(k) \triangleq \sum_{g\neq i}\sum_{j\neq n} \bar{\boldsymbol{\mathcal{H}}}_{jg,i}(k)\times_3\big(\sqrt{\gamma(1-\gamma)}\bar{\boldsymbol{\Omega}}_{jg}(k) + \gamma\rho_1\mathbf{1}\big) \quad (15)$$

where $\boldsymbol{\mathcal{I}}_u(k)$, $\boldsymbol{\mathcal{I}}_p(k)$ and $\boldsymbol{\mathcal{I}}_e(k)$ respectively represent intra-cell interference, pilot contamination and inter-cell interference. Therefore, (11) can be rewritten as the following signal plus interference model

$$\begin{aligned}\boldsymbol{\mathcal{Y}}_{ni}(k) = \bar{\boldsymbol{\mathcal{H}}}_{ni,i}(k)\times_3\big(\sqrt{\gamma(1-\gamma)}\bar{\boldsymbol{\Omega}}_{ni}(k) + \gamma\boldsymbol{I}\big)\\ + \boldsymbol{\mathcal{I}}_p(k) + \boldsymbol{\mathcal{I}}_u(k) + \boldsymbol{\mathcal{I}}_e(k) + \boldsymbol{\mathcal{W}}_i'(k).\end{aligned} \quad (16)$$

When $\gamma = 1$, the above signal model represents the conventional orthogonal pilot-data scheme. When $\gamma = 0$, the observation of pilots in (16) is null. Therefore, the DOAs of the user $n$ cannot be identified. This is because the DOAs of the user $n$ associated with its data symbols are merged into the interference term. In general, when $\gamma$ increases, the received SNR in (16) increases as well, which results in more accurate CSI estimation.

## III. Uplink Angular Channel Estimation

From (2), we see the channel model is characterized by the propagation parameters such as $u$, $v$ and $\tau$, etc. Therefore, we can obtain the uplink CSI through these parameters instead of the whole channel matrix. In (16), for a fixed $k$, the "primary" channel parameters for channel estimation are the uplink DOAs $u$ and $v$. Here, the DOAs named as "primary parameters" is because other channel parameters such as path gain, $\alpha$, can be consequently estimated after the DOAs are obtained. Moreover, the delay parameters are equivalently estimated when the frequency fading on all the sub-carriers is obtained. The uplink DOAs of the user $n$ on the sub-carrier $k$ are embedded into an order-3 tensor as follows

$$\boldsymbol{\mathcal{R}}_{ni,i}(\boldsymbol{u}_{ni,i}, \boldsymbol{v}_{ni,i}) = \bar{\boldsymbol{\mathcal{H}}}_{ni,i}\times_3\big(\sqrt{\gamma(1-\gamma)}\bar{\boldsymbol{\Omega}}_{ni} + \gamma\boldsymbol{I}\big) \quad (17)$$

where the index $k$ is dropped off for ease of expression. Most importantly, the received signal can be expressed as a sum of rank-one tensors, namely CANDECOMP/PARAFAC (CP) decomposition. This inherent decomposed structure is utilized in our channel estimation algorithm design.

The likelihood estimation of $\boldsymbol{v}_{ni,i}$ and $\boldsymbol{u}_{ni,i}$ is given by

$$\max_{\boldsymbol{u}_{ni,i},\boldsymbol{v}_{ni,i}} \log P(\boldsymbol{\mathcal{Y}}_{ni}; \boldsymbol{u}_{ni,i}, \boldsymbol{v}_{ni,i})$$

where $\boldsymbol{\mathcal{Y}}_{ni}$ is the observation from (16), $\boldsymbol{v}_{ni,i}$ and $\boldsymbol{u}_{ni,i}$ are the parameters. By incorporating the interference as a hidden variable, $P(\boldsymbol{\mathcal{Y}}_{ni}; \boldsymbol{u}_{ni,i}, \boldsymbol{v}_{ni,i})$ is calculated by

$$\begin{aligned}P(\boldsymbol{\mathcal{Y}}_{ni}; \boldsymbol{u}_{ni,i}, \boldsymbol{v}_{ni,i}) &= \int P(\boldsymbol{\mathcal{Y}}_{ni}, \boldsymbol{\mathcal{I}}; \boldsymbol{u}_{ni,i}, \boldsymbol{v}_{ni,i}) d\boldsymbol{\mathcal{I}}\\ &= \int P(\boldsymbol{\mathcal{Y}}_{ni}|\boldsymbol{\mathcal{I}}; \boldsymbol{u}_{ni,i}, \boldsymbol{v}_{ni,i}) P(\boldsymbol{\mathcal{I}}) d\boldsymbol{\mathcal{I}}\end{aligned}$$

where $\boldsymbol{\mathcal{I}} \triangleq \boldsymbol{\mathcal{I}}_p + \boldsymbol{\mathcal{I}}_u + \boldsymbol{\mathcal{I}}_e$. Alternatively, we have

$$\begin{aligned}&\log P(\boldsymbol{\mathcal{Y}}_{ni}|\boldsymbol{\mathcal{I}}; \boldsymbol{u}_{ni,i}, \boldsymbol{v}_{ni,i})\\ &\quad\propto -\|\boldsymbol{\mathcal{Y}}_{ni} - \boldsymbol{\mathcal{R}}_{ni,i}(\boldsymbol{u}_{ni,i}, \boldsymbol{v}_{ni,i}) - \boldsymbol{\mathcal{I}}\|^2\end{aligned}$$

where $\propto$ means propotional to. The target parameters $\boldsymbol{u}$ and $\boldsymbol{v}$ are in the term $\boldsymbol{\mathcal{R}}_{ni,i}(\boldsymbol{u}_{ni,i}, \boldsymbol{v}_{ni,i})$ which has a CP decomposition structure. However, when we consider the integral calculation by incorporating $\boldsymbol{\mathcal{I}}$'s PDF, the tensor $\boldsymbol{\mathcal{R}}_{ni,i}(\boldsymbol{u}_{ni,i}, \boldsymbol{v}_{ni,i})$ needs to be vectorized first. Thus, the resulting likelihood function cannot preserve the tensor structure of $\boldsymbol{\mathcal{R}}_{ni,i}(\boldsymbol{u}_{ni,i}, \boldsymbol{v}_{ni,i})$ which leads the maximum likelihood estimation to be a far more complicated multi-dimensional optimization problem.

To deal with the aforementioned issue, we utilize the expectation-maximization (EM) algorithm [20]. In the EM algorithm, the hyper-parameters are iteratively estimated by maximizing a lower bound of the likelihood function, which is called M-step. On the other hand, the lower bound is gradually updated by using a new estimated hyper-parameters, which is called E-step. At the $t+1$th iteration of the EM algorithm, we denote the adaptive lower bound of the log-likelihood $\log P(\boldsymbol{\mathcal{Y}}_{ni}; \boldsymbol{u}_{ni,i}, \boldsymbol{v}_{ni,i})$ as

$$Q(\boldsymbol{u}_{ni,i}, \boldsymbol{v}_{ni,i}|\boldsymbol{u}_{ni,i}(t), \boldsymbol{v}_{ni,i}(t)) \quad (18)$$

where $\boldsymbol{u}_{ni,i}(t)$ and $\boldsymbol{v}_{ni,i}(t)$ are the estimated DOAs from the previous iteration. By choosing $\boldsymbol{u}_{ni,i}(t+1)$ and $\boldsymbol{v}_{ni,i}(t+1)$ through maximizing the $Q$ function, we can always obtain an improvement of the likelihood value, i.e.,

$$\begin{aligned}&\log P(\boldsymbol{\mathcal{Y}}_{ni}; \boldsymbol{u}_{ni,i}(t+1), \boldsymbol{v}_{ni,i}(t+1))\\ &\qquad\geq \log P(\boldsymbol{\mathcal{Y}}_{ni}; \boldsymbol{u}_{ni,i}(t), \boldsymbol{v}_{ni,i}(t))\end{aligned}$$

In our problem, the calculations of the E-steps and the M-steps are given by

*1) In E-step:*

$$\begin{aligned}&Q(\boldsymbol{u}_{ni,i}, \boldsymbol{v}_{ni,i}|\boldsymbol{u}_{ni,i}(t), \boldsymbol{v}_{ni,i}(t))\\ &\quad= \mathbb{E}_{\boldsymbol{\mathcal{I}}|\boldsymbol{\mathcal{Y}}_{ni};\boldsymbol{u}_{ni,i}(t),\boldsymbol{v}_{ni,i}(t)}(\log P(\boldsymbol{\mathcal{Y}}_{ni}, \boldsymbol{\mathcal{I}}; \boldsymbol{u}_{ni,i}, \boldsymbol{v}_{ni,i}))\end{aligned} \quad (19)$$

which implies

$$\begin{aligned}&Q(\boldsymbol{u}_{ni,i}, \boldsymbol{v}_{ni,i}|\boldsymbol{u}_{ni,i}(t), \boldsymbol{v}_{ni,i}(t))\\ &\quad\propto -\mathbb{E}_{\boldsymbol{\mathcal{I}}|\boldsymbol{\mathcal{Y}}_i;\boldsymbol{u}_{ni,i}(t),\boldsymbol{v}_{ni,i}(t)}\|\boldsymbol{\mathcal{Y}}_{ni} - \boldsymbol{\mathcal{R}}_{ni,i}(\boldsymbol{u}_{ni,i}, \boldsymbol{v}_{ni,i}) - \boldsymbol{\mathcal{I}}\|^2\end{aligned} \quad (20)$$

*2) In M-step:*

$$\max_{\boldsymbol{u}_{ni,i},\boldsymbol{v}_{ni,i}} Q(\boldsymbol{u}_{ni,i},\boldsymbol{v}_{ni,i}|\boldsymbol{u}_{ni,i}(t),\boldsymbol{v}_{ni,i}(t)) \tag{21}$$

which is equivalent to

$$\min_{\boldsymbol{u}_{ni,i},\boldsymbol{v}_{ni,i}} \mathbb{E}_{\boldsymbol{\mathcal{I}}|\boldsymbol{\mathcal{Y}}_{ni};\boldsymbol{u}_{ni,i}(t),\boldsymbol{v}_{ni,i}(t)} \times \|\boldsymbol{\mathcal{Y}}_{ni} - \boldsymbol{\mathcal{R}}_{ni,i}(\boldsymbol{u}_{ni,i},\boldsymbol{v}_{ni,i}) - \boldsymbol{\mathcal{I}}\|^2 \tag{22}$$

The derivation of the above E-step and M-step are presented in Appendix A. Accordingly, at each iteration of the M-step, the updates of $u$ and $v$ are given by

$$\hat{u}^{(p)}_{ni,i,l} = \arg\max_{u} \frac{|\boldsymbol{a}^H(u)\boldsymbol{F}^{(2)}(:;p,l)|}{\|\boldsymbol{a}(u)\|\|\boldsymbol{F}^{(2)}(:;p,l)\|},\ l=1,\ldots,L \tag{23}$$

$$\hat{v}^{(p)}_{ni,i,l} = \arg\max_{v} \frac{|\boldsymbol{a}^H(v)\boldsymbol{F}^{(1)}(:;p,l)|}{\|\boldsymbol{a}(v)\|\|\boldsymbol{F}^{(1)}(:;p,l)\|},\ l=1,\ldots,L \tag{24}$$

where

$$\boldsymbol{F}^{(1)} \triangleq \left[(\boldsymbol{Y}_{i(1)} - \mathbb{E}(\boldsymbol{I}_{(1)}))(\boldsymbol{F}^{(3)} \odot \boldsymbol{F}^{(2)})^* + \mu_l \boldsymbol{A}(\boldsymbol{u}_{ni,i})\right] \times \left[(\boldsymbol{F}^{(3)} \odot \boldsymbol{F}^{(2)})^T (\boldsymbol{F}^{(3)} \odot \boldsymbol{F}^{(2)})^* + \mu_l \boldsymbol{I}\right]^{-1} \tag{25}$$

$$\boldsymbol{F}^{(2)} \triangleq \left[(\boldsymbol{Y}_{i(2)} - \mathbb{E}(\boldsymbol{I}_{(2)}))(\boldsymbol{F}^{(3)} \odot \boldsymbol{F}^{(1)})^* + \mu_l \boldsymbol{A}(\boldsymbol{v}_{ni,i})\right] \times \left[(\boldsymbol{F}^{(3)} \odot \boldsymbol{F}^{(1)})^T (\boldsymbol{F}^{(3)} \odot \boldsymbol{F}^{(1)})^* + \mu_l \boldsymbol{I}\right]^{-1} \tag{26}$$

and

$$\boldsymbol{F}^{(3)} \triangleq \left(\boldsymbol{Y}_{i(3)} - \mathbb{E}(\boldsymbol{I}_{(3)})\right)\left(\left(\boldsymbol{F}^{(2)} \odot \boldsymbol{F}^{(1)}\right)^T\right)^+ \tag{27}$$

in which the expectation $\mathbb{E}$ stands for the posterior mean of $P(\boldsymbol{\mathcal{I}}|\boldsymbol{\mathcal{Y}}_{ni};\boldsymbol{u}_{ni,i}(t),\boldsymbol{v}_{ni,i}(t))$. These three matrices are the introduced auxiliary variables which essentially stand for the factor matrices of $\boldsymbol{\mathcal{R}}_{ni,i}(\boldsymbol{u}_{ni,i},\boldsymbol{v}_{ni,i})$, i.e.,

$$\boldsymbol{\mathcal{R}}_{ni,i}(\boldsymbol{u}_{ni,i},\boldsymbol{v}_{ni,i}) = \left[\!\left[\boldsymbol{F}^{(1)},\boldsymbol{F}^{(2)},\boldsymbol{F}^{(3)}\right]\!\right]$$

where the bracket $[\![\cdot]\!]$ stands for the CP decomposition [21]. For ease of later discussion, our introduced DOA estimation algorithm is summarized in Algorithm 1.

By looking at the resulting EM algorithm, we see that each iteration contains three interference cancellation stages which are offered by the calculations of $(\boldsymbol{Y}_{i(1)} - \mathbb{E}(\boldsymbol{I}_{(1)}))$, $(\boldsymbol{Y}_{i(2)} - \mathbb{E}(\boldsymbol{I}_{(2)}))$ and $(\boldsymbol{Y}_{i(3)} - \mathbb{E}(\boldsymbol{I}_{(3)}))$ defined in (25), (26) and (27) respectively, i.e., the LMMSE estimation of the interference is subtracted from the observation. The stopping criterion for tensor decomposition procedure inside the M-step is determined such that the objective cannot ascend, i.e.,

$$Q(\bar{\boldsymbol{u}}(k+1),\bar{\boldsymbol{v}}(k+1)|\boldsymbol{u}_{ni,i}(t),\boldsymbol{v}_{ni,i}(t)) < Q(\bar{\boldsymbol{u}}(k),\bar{\boldsymbol{v}}(k)|\boldsymbol{u}_{ni,i}(t),\boldsymbol{v}_{ni,i}(t))$$

where $\bar{\boldsymbol{u}}(k)$ and $\bar{\boldsymbol{v}}(k)$ respectively stands for the estimated DOA at the $k$th iterations of the decomposition. For the outer iterations among E-steps and M-steps, the termination condition is chosen as

$$Q(\boldsymbol{u}_{ni,i}(t+1),\boldsymbol{v}_{ni,i}(t+1)|\boldsymbol{u}_{ni,i}(t),\boldsymbol{v}_{ni,i}(t)) < Q(\boldsymbol{u}_{ni,i}(t),\boldsymbol{v}_{ni,i}(t)|\boldsymbol{u}_{ni,i}(t-1),\boldsymbol{v}_{ni,i}(t-1))$$

**Algorithm 1:** DOA Estimation by EM Algorithm.

**Input:** Interference Covariance Matrix
**Output:** $\hat{\boldsymbol{v}}_{ni,i}$, $\hat{\boldsymbol{u}}_{ni,i}$

```
(Initialization)
Initialize u_{ni,i}(0), v_{ni,i}(0), R^{(0)}_{ni,i}, t = 0, μ_0 and k = 0
while t < T do
  (E-Step)
  Update E(I) by LMMSE as introduced in (59)
  (M-Step)
  while k < K do
    Update F^(1), F^(2), F^(3) according to (53), (54)
    and (55).
    Update ū(k), v̄(k) using (23) and (24)
    Update J^(k) = Q(ū(k), v̄(k)|u_{ni,i}(t), v_{ni,i}(t))
    if J^(k) < J^(k−1) then
      break;
    end if
    Update μ_k = μ_{k−1} * β_2
    k = k × 1
  end while
  Update
  f^(t) = Q(ū(k−1), v̄(k−1)|u_{ni,i}(t), v_{ni,i}(t))
  if f^(t) < f^(t−1) then
    break
  else
    u_{ni,i}(t) = ū(k−1), v_{ni,i}(t) = v̄(k−1)
  end if
  t = t + 1
end while
```

It happens when the lower bound of the likelihood value no longer ascends. Moreover, we see that the parameters retrieval in (23) and (24) are conducted through one dimensional searching on each decomposed vector, rather than operating on multiple dimensional spaces. Due to the same reason as discussed in [23], this decomposed manner can significantly reduce the complexity since the decomposed structure of $\boldsymbol{\mathcal{Y}}$ is utilized. In detail, a complexity analysis is given in Section IV-C.

## IV. Theoretical Analyses

In this section, we consider theoretical analysis related to our introduced algorithm, including Cramér Rao lower bound of the DOA estimation, uniqueness condition of the CP decomposition in the M-step and algorithm complexity.

### A. Cramér Rao Lower Bound of DOA estimation

To evaluate the performance of our introduced algorithm, we utilize Cramér Rao lower bound as the performance benchmark. In order to obtain the CRLB, we calculate the Fisher information matrix via a vectorization of $\boldsymbol{\mathcal{Y}}_{ni}$. We write the mode-3 unfolding of $\boldsymbol{\mathcal{Y}}_{ni}$ as

$$\boldsymbol{y}_{ni} = \boldsymbol{A}(\boldsymbol{\varphi}_{ni,i})\boldsymbol{\alpha}_{ni,i} + \boldsymbol{i} + \boldsymbol{w} \tag{28}$$

where $\boldsymbol{A}(\boldsymbol{\varphi}_{ni,i}) \triangleq \boldsymbol{A}(\boldsymbol{v}_{ni,i}) \odot \boldsymbol{A}(\boldsymbol{u}_{ni,i}) \odot ((\sqrt{\gamma(1-\gamma)}\bar{\boldsymbol{\Omega}}_{ni}^T + \gamma \boldsymbol{I})\boldsymbol{P}^H(\boldsymbol{\omega}_{ni,i})\boldsymbol{A}(\boldsymbol{\omega}_{ni,i}))$ and $\boldsymbol{\varphi}_{ni,i} \triangleq [\boldsymbol{v}_{ni,i}, \boldsymbol{u}_{ni,i}]$. As shown in [27], the Fisher information can be calculated by

$$\boldsymbol{F}_{\varphi_{l_1},\varphi_{l_2}} = \frac{1}{2}\mathrm{Tr}\left\{\frac{\partial \boldsymbol{C}_{\boldsymbol{y}}}{\partial \varphi_{l_1}}\boldsymbol{C}_{\boldsymbol{y}}^{-1}\frac{\partial \boldsymbol{C}_{\boldsymbol{y}}}{\partial \varphi_{l_2}}\boldsymbol{C}_{\boldsymbol{y}}^{-1}\right\}$$

where

$$\boldsymbol{C}_y = \begin{bmatrix} \boldsymbol{R}_{\boldsymbol{y}} & \boldsymbol{M}_{\boldsymbol{y}} \\ \boldsymbol{M}_{\boldsymbol{y}}^* & \boldsymbol{R}_{\boldsymbol{y}}^* \end{bmatrix}$$

in which

$$\boldsymbol{R}_{\boldsymbol{y}} = \mathbb{E}\left[\boldsymbol{A}(\boldsymbol{\varphi})\boldsymbol{\alpha}\boldsymbol{\alpha}^H\boldsymbol{A}^H(\boldsymbol{\varphi})\right] + \mathbb{E}\left(\boldsymbol{i}\boldsymbol{i}^H\right) + \sigma^2\boldsymbol{I} \tag{29}$$

$$\boldsymbol{M}_{\boldsymbol{y}} = \mathbb{E}\left[\boldsymbol{A}(\boldsymbol{\varphi})\boldsymbol{\alpha}\boldsymbol{\alpha}^T\boldsymbol{A}^T(\boldsymbol{\varphi})\right] + \mathbb{E}\left(\boldsymbol{i}\boldsymbol{i}^T\right) \tag{30}$$

$\mathbb{E}(\boldsymbol{i}\boldsymbol{i}^H)$ and $\mathbb{E}(\boldsymbol{i}\boldsymbol{i}^T)$ respectively represent the covariance matrix and pseudo-covariance matrix of the interference. The resulting approximate expressions of $\mathbb{E}(\boldsymbol{i}\boldsymbol{i}^H)$ and $\mathbb{E}(\boldsymbol{i}\boldsymbol{i}^T)$ are given in Appendix D. Meanwhile, more details of the CRLB calculation are summarized in Appendix E. In Section VI, we will see the value of the CRLB decreases as $\gamma$ increases. This is because the effective SINR in the channel estimation stage is improved.

### B. Uniqueness of CP Decomposition in M-Step

In the M-step, the DOA is estimated from a CP decomposition manner. Generally, the uniqueness condition for arbitrary CP decomposition follows the Kruskal condition below

*Theorem 1:* Let $\boldsymbol{\mathcal{R}} = [\![\boldsymbol{F}^{(1)}, \boldsymbol{F}^{(2)}, \boldsymbol{F}^{(3)}]\!] \in \mathbb{C}^{I_1 \times I_2 \times I_3}$, where $\boldsymbol{F}^{(i)} \in \mathbb{C}^{I_i \times L}$. If

$$k\left(\boldsymbol{F}^{(1)}\right) + k\left(\boldsymbol{F}^{(1)}\right) + k\left(\boldsymbol{F}^{(1)}\right) \geq 2L + 2$$

where $k(\boldsymbol{F})$ stands for the k-rank of matrix $\boldsymbol{F}$, then the CP decomposition $[\![\boldsymbol{F}^{(1)}, \boldsymbol{F}^{(2)}, \boldsymbol{F}^{(3)}]\!]$ of $\boldsymbol{\mathcal{R}}$ is unique.

*Proof:* For the proof, please refer to [28], [29]. ■

In addition, the steering vector of FD-MIMO has a specific Vandermonde structure which can be further leveraged to derive a stronger uniqueness condition as follows,

*Theorem 2:* Let $\boldsymbol{\mathcal{R}} = [\![\boldsymbol{F}^{(1)}, \boldsymbol{F}^{(2)}, \boldsymbol{F}^{(3)}]\!] \in \mathbb{C}^{I_1 \times I_2 \times I_3}$, where $\boldsymbol{F}^{(1)} \in \mathbb{C}^{I_1 \times L}$ and $\boldsymbol{F}^{(2)} \in \mathbb{C}^{I_2 \times L}$ are Vandermonde matrix. We suppose any two columns from $\boldsymbol{F}^{(1)}$ and $\boldsymbol{F}^{(2)}$ are different. If $rank(\boldsymbol{F}^{(1)}) = L$ and $rank(\boldsymbol{F}^{(3)} \odot \boldsymbol{F}^{(2)}) = L$, the given CP decomposition $[\![\boldsymbol{F}^{(1)}, \boldsymbol{F}^{(2)}, \boldsymbol{F}^{(3)}]\!]$ of $\boldsymbol{\mathcal{R}}$ is unique.

*Proof:* The proof can be found at the Corollary III.4 in [30]. ■

According to the above theorem, if the matrices $\boldsymbol{F}^{(1)}$ and $\boldsymbol{F}^{(3)} \odot \boldsymbol{F}^{(2)}$ are full column-rank, the unique decomposition can be achieved. Therefore, when it is applied to our case, the resulting uniqueness of the M-step decomposition can be ensured when the number of distinguishable paths is smaller than the number of antennas at azimuth and elevation directions.

### C. Complexity

At each iteration of our introduced EM algorithm, the major computational task involves solving three regularized least squared problems, noted as (25), (26) and (27) . In (25), since $\boldsymbol{F}^{(3)} \odot \boldsymbol{F}^{(2)} \in \mathbb{C}^{N_{B_e}M \times L}$, the number of flops required to compute (25) is $\mathcal{O}(N_{Ba}N_{B_e}ML + N_{Ba}ML^2 + L^3)$, where the big $\mathcal{O}$ represents 'on the order of'. Similarly, the flops for (26) and (27) are equal to $\mathcal{O}(N_{Ba}N_{B_e}ML + N_{Be}ML^2 + L^3)$ and $\mathcal{O}(N_{Ba}N_{B_e}ML + N_{Be}N_{Ba}L^2 + L^3)$, respectively.

By following up a similar analysis in [23], when we conduct a joint estimation of azimuth and elevation angle directions, the major complexity is dominated by $\mathcal{O}(\bar{N}_{Ba}^p\bar{N}_{B_e}^p)$, where $\bar{N}_{Ba} \gg N_{Ba}$ and $\bar{N}_{B_e} \gg N_{B_e}$ represent the number of quantization grids on the azimuth and elevation angular spaces, respectively; $p$ is a positive integer related to selected algorithms. Therefore, the joint estimation method essentially has a higher complexity than the decomposition-based method.

## V. System Sum Rate

In the pilot-data superposition framework, the overall rate performance is characterized by the weighted summation of the uplink and downlink rate as introduced at the very beginning of this paper, defined in (1). Through the previous discussion on the CRLB, it reveals that the estimated channel is a function of the allocated pilot power. Therefore, the performance of the downlink rate is a function of the pilot power as well, since the downlink precoding is designed according to the estimated channel. Moreover, the uplink is more complicated than the downlink as: when we allocate more power to data, the pilot power is reduced accordingly, which leads to a deterioration of the CSI estimation performance. Thus, increasing the data power in uplink cannot sufficiently improve the uplink rate.

For simplicity, we define $\kappa_{DL} + \kappa_{UL} = 1$ and $\kappa = \kappa_{UL}$. Then, the overall sum-rate metric (1) can be rewritten as

$$R(\kappa, \gamma) = \kappa\left(R^{UL}(\gamma) - R^{DL}(\gamma)\right) + R^{DL}(\gamma) \tag{31}$$

Given a fixed power allocation $\gamma$, the difference between the uplink and downlink rate is the slope of the rate function $R(\kappa)$. Thus, different precoding and receiving techniques can result in different sum rate slopes. In order to numerically reveal this, we mainly introduce the sum-rate calculations of the uplink and downlink in this section.

### A. Uplink Sum-Rate by Linear Receiver

For the uplink transmission, it can be split into two steps. In the first step, the data symbols are contaminated by the superimposed pilot, in which the rate is denoted as $R^{UL_p}$. In the second step, the uplink resources are purely allocated for data transmission which we denote as $R^{UL_d}$.

Now, we consider the uplink rate calculation in the first step on a single carrier. To estimate the data symbols, the pilot term $\sum_j \boldsymbol{\mathcal{H}}_{ji,i} \times_3 \boldsymbol{S}_{ji,i}$ in (7) is first subtracted from the estimated one $\sum_j \hat{\boldsymbol{\mathcal{H}}}_{ji,i} \times_3 \boldsymbol{S}_{ji,i}$. Therefore, the observation of the data symbols $\bar{\boldsymbol{\mathcal{Z}}}_{ni,i}$ is expressed as

$$\bar{\boldsymbol{\mathcal{Z}}}_{ni,i} = \boldsymbol{\mathcal{H}}_{ni,i} \times_3 \boldsymbol{X}_{ni} + \sum_j (\boldsymbol{\mathcal{H}}_{ji,i} - \hat{\boldsymbol{\mathcal{H}}}_{ji,i}) \times_3 \boldsymbol{S}_{ji,i} + \boldsymbol{\mathcal{I}} \tag{32}$$

where $\boldsymbol{\mathcal{I}}$ includes the remained interference. By substituting the channel model in (32), the unfolding version of the tensor $\bar{\boldsymbol{\mathcal{Z}}}_{ni,i}$

on the third mode is given by

$$\bar{\boldsymbol{Z}}_{ni,i_{(3)}} = \bar{\boldsymbol{X}}_{ni}^{T}\boldsymbol{P}^{H}(\boldsymbol{\omega}_{ni,i})\boldsymbol{A}(\boldsymbol{\omega}_{ni,i})\boldsymbol{\Lambda}_0(\boldsymbol{\alpha}_{ni,i}) \times [\boldsymbol{A}(\boldsymbol{v}_{ni,i}) \odot \boldsymbol{A}(\boldsymbol{u}_{ni,i}))]^T + \boldsymbol{I}_{(3)}$$

Then, by conducting a linear receiver $\boldsymbol{E}_{ni,i}^{T} \in \mathbb{C}^{N_{BSa}N_{BSe}\times L}$ on the right hand of $\bar{\boldsymbol{Z}}_{ni,i_{(3)}}$, we have

$$\boldsymbol{Z}'_{ni,i} = \boldsymbol{E}_{ni,i}\boldsymbol{B}(\boldsymbol{v}_{ni,i},\boldsymbol{u}_{ni,i})\bar{\boldsymbol{X}}_{ni} + \boldsymbol{I}'$$

where $\boldsymbol{Z}'_{ni,i} \triangleq (\bar{\boldsymbol{Z}}_{ni,i_{(3)}}\boldsymbol{E}_{ni,i}^{T})^T \in \mathbb{C}^{L\times T}$, $\boldsymbol{I}' = \boldsymbol{E}_{ni,i}\boldsymbol{I}_{(3)}^{T}$,

$$\boldsymbol{B}(\boldsymbol{v}_{ni,i},\boldsymbol{u}_{ni,i}) \triangleq [\boldsymbol{A}(\boldsymbol{v}_{ni,i}) \odot \boldsymbol{A}(\boldsymbol{u}_{ni,i}))]\boldsymbol{\Lambda}_0(\boldsymbol{\alpha}_{ni,i})\boldsymbol{P}^{*}(\boldsymbol{\omega}_{ni,i})\boldsymbol{A}^{T}(\boldsymbol{\omega}_{ni,i})$$

Thus, for the $l$th stream data, we have

$$\boldsymbol{z}'_{ni,i,l} = \boldsymbol{e}_{ni,i,l}^{H}\boldsymbol{b}(\boldsymbol{v}_{ni,i,l},\boldsymbol{u}_{ni,i,l})\bar{\boldsymbol{x}}_{ni,l}^{H} + \boldsymbol{e}_{ni,i,l}^{H}\sum_{l'\neq l}\boldsymbol{b}(\boldsymbol{v}_{ni,i,l'},\boldsymbol{u}_{ni,i,l'})\bar{\boldsymbol{x}}_{ni,l'}^{H} + \boldsymbol{i}_{l}'^{H}$$

where $\boldsymbol{e}_{ni,i,l}^{H}$ represents the $l$th row of $\boldsymbol{E}_{ni,i}$, $\bar{\boldsymbol{x}}_{ni,l}^{H}$ is the $l$th row of $\bar{\boldsymbol{X}}_{ni}$ and $\boldsymbol{b}(\boldsymbol{v}_{ni,i,l},\boldsymbol{u}_{ni,i,l})$ is the $l$th column of $\boldsymbol{B}(\boldsymbol{v}_{ni,i},\boldsymbol{u}_{ni,i})$. Therefore, the SINR at the $l$th stream of the $n$th user in the uplink can be expressed as

$$\text{SINR}_{ni,i,l}^{UL_p} = \frac{(1-\gamma)|\tilde{b}(\boldsymbol{v}_{ni,i,l},\boldsymbol{u}_{ni,i,l})|^2}{(1-\gamma)I_{stream} + \gamma I_{res} + (1-\gamma)I_{intra} + I_{Inter} + T\sigma^2} \tag{33}$$

where $\tilde{b}(\boldsymbol{v}_{ni,i,l},\boldsymbol{u}_{ni,i,l}) \triangleq \boldsymbol{e}_{ni,i,l}^{H}\boldsymbol{b}(\boldsymbol{v}_{ni,i,l},\boldsymbol{u}_{ni,i,l})$,

$$I_{stream} \triangleq \sum_{l'\neq l}|\tilde{b}(\boldsymbol{v}_{ni,i,l'},\boldsymbol{u}_{ni,i,l'})|^2$$

$$I_{intra} \triangleq \sum_{j\neq n}\sum_{l'}|\tilde{b}(\boldsymbol{v}_{ji,i,l'},\boldsymbol{u}_{ji,i,l'})|^2$$

$$I_{res} \triangleq \sum_{j}\sum_{l'}|(\tilde{b}(\boldsymbol{v}_{ji,i,l'},\boldsymbol{u}_{ji,i,l'}) - \tilde{b}(\hat{\boldsymbol{v}}_{ji,i,l'},\hat{\boldsymbol{u}}_{ji,i,l'}))|^2$$

$$I_{inter} \triangleq \sum_{g\neq i}\sum_{j}\sum_{l'}|\tilde{b}(\boldsymbol{v}_{jg,i,l'},\boldsymbol{u}_{jg,i,l'})|^2$$

Similarly, when the pure data transmission is conducted through a length-$T$ block in the uplink, the SINR is given by

$$\text{SINR}_{ni,i,l}^{UL_d} = \frac{|\tilde{b}(\boldsymbol{v}_{ni,i,l},\boldsymbol{u}_{ni,i,l})|^2}{I_{stream} + I_{intra} + I_{Inter} + T\sigma^2} \tag{34}$$

where $T\sigma^2$ is the additive noise power on the length-$T$ symbols. Therefore, the uplink rate on the $i$th cell can be calculated by

$$R_i^{UL} = \sum_{j}\sum_{l}\log_2\left(1+\text{SINR}_{ji,i,l}^{UL_p}\right) + N_d\log_2\left(1+\text{SINR}_{ji,i,l}^{UL_d}\right) \tag{35}$$

where $N_d$ is the number of length-$T$ blocks allocated for pure data transmission in uplink. In multiple carriers case, we assume the first $T$ OFDM symbols are allocated with pilot-data superposition. Then, the uplink rate is directly obtained as $N_cR_i^{UL}$, where $N_c$ stands for the number of sub-carriers.

### *B. Downlink Sum-Rate by Linear Precoding*

By unfolding the third mode of the downlink channel tensor $\boldsymbol{\mathcal{H}}_{ni,g}$, the $n$th MS channel can be expressed as $\boldsymbol{H}_{ji,i_{(3)}}$. Letting $\boldsymbol{Q}_i \in \mathbb{C}^{N_{BSa}N_{BSe}\times T}$ represent the precoded symbols at the $i$th BS, we have

$$\boldsymbol{Q}_i = \sum_{j}\boldsymbol{P}_{ji,i}\boldsymbol{X}_{ji}$$

where $\boldsymbol{P}_{ji,i} \in \mathbb{C}^{N_{BSa}N_{BSe}\times L}$ represents the precoding matrix for the $j$th MS and $\boldsymbol{X}_{ji} \in \mathbb{C}^{L\times T}$ is the corresponding transmitted symbols. Therefore, the received signal at the $n$th MS is given by

$$\boldsymbol{Y}_{ni} = \boldsymbol{H}_{ni,i_{(3)}}\boldsymbol{P}_{ni,i}\boldsymbol{X}_{ni} + \boldsymbol{H}_{ni,i_{(3)}}\sum_{j\neq n}\boldsymbol{P}_{ji,i}\boldsymbol{X}_{ji} + \sum_{g\neq i}\sum_{j}\boldsymbol{H}_{ni,g_{(3)}}\boldsymbol{P}_{jg,g}\boldsymbol{X}_{jg,g} + \boldsymbol{W}_{ni,i} \tag{36}$$

For the $n$th MS in the $i$th cell, we assume it employs a linear receiver $\boldsymbol{E}_{ni,i}$. Thus the received signal at the $l$th stream can be expressed as

$$\boldsymbol{e}_{ni,i,l}^{H}\boldsymbol{Y}_{ni,i} = \boldsymbol{e}_{ni,i,l}^{H}\boldsymbol{H}_{ni,i_{(3)}}\boldsymbol{p}_{ni,i,l}\boldsymbol{x}_{ni,l}^{H} \tag{37}$$

$$+ \sum_{l'\neq l}\boldsymbol{e}_{ni,i,l}^{H}\boldsymbol{H}_{ni,i_{(3)}}\boldsymbol{p}_{ni,i,l}\boldsymbol{x}_{ni,l}^{H} + \boldsymbol{i}_{l}'^{H} \tag{38}$$

Therefore, the SINR at the $l$th stream of the $n$th user is given by

$$\text{SINR}_{ni,i,l}^{DL} = \frac{\left|\tilde{\boldsymbol{h}}_{ni,i,l}^{H}\boldsymbol{p}_{ni,i,l}\right|^2}{I_{stream} + I_{intra} + I_{inter} + T\sigma^2} \tag{39}$$

where $\tilde{\boldsymbol{h}}_{ni,i,l}^{H} \triangleq \boldsymbol{e}_{ni,i,l}^{H}\boldsymbol{H}_{ni,i(3)}$

$$I_{stream} = \sum_{l'\neq l}\left|\tilde{\boldsymbol{h}}_{ni,i,l}^{H}\boldsymbol{p}_{ni,i,l'}\right|^2$$

$$I_{intra} = \sum_{j\neq n}\sum_{l'}\left|\tilde{\boldsymbol{h}}_{ni,i,l}^{H}\boldsymbol{p}_{ji,i,l'}\right|^2$$

$$I_{inter} = \sum_{j}\sum_{g}\sum_{l'}\left|\tilde{\boldsymbol{h}}_{jg,g,l}^{H}\boldsymbol{p}_{jg,,l'}\right|^2$$

Hence, the downlink sum rate in the $i$th cell is given by

$$R_i^{DL} = \sum_{j}\sum_{l}\log_2(1+\text{SINR}_{ji,i,l}^{DL}) \tag{40}$$

Similarly, in multiple-carriers case, the downlink rate is scaled by $\bar{R}_i^{DL} = N_cR_i^{DL}$.

### *C. Linear Precoding for Downlink*

From the previous two sub-sections, we see the downlink/uplink rate are functions of the precoding/receiving processing methods. To further investigate the extent of the

downlink-uplink rate trade-off by precoding/receiving, we briefly discuss the downlink precoding design. A similar uplink receiving design can be attained through the uplink-downlink duality [31]. Now, we consider maximizing the sum rate (40)

$$\max_{\{\boldsymbol{p}_{ji,i,l}\}} \sum_{j=1}^{J}\sum_{l=1}^{L} R_{ji,i,l}^{DL}$$
$$\text{s.t.} \quad \sum_{l=1}^{L} \text{Tr}\big(\boldsymbol{p}_{ji,i,l}\boldsymbol{p}_{ji,i,l}^{H}\big) < P_j, \forall j \tag{41}$$

where the constraint represents the limitation of the allocated power power for each user. This optimization problem can be equivalently solved by a weighted minimum mean square error (WMMSE) minimization problem which is given by [32]

$$\min_{\{w_{ji,i,l},u_{ji,i,l},\boldsymbol{p}_{ji,i,l}\}} \sum_{j}\sum_{l} w_{ji,i,l}M_{ji,i,l} - \log(w_{ji,i,l})$$
$$\text{s.t.} \quad \sum_{l=1}^{L} \text{Tr}\big(\boldsymbol{p}_{ji,i,l}\boldsymbol{p}_{ji,i,l}^{H}\big) < P_j, \forall j \tag{42}$$

where $M_{ji,i,l}$ stands for the receiving mean squared error (MSE) on the $l$th stream of the $j$th user which is defined as

$$M_{ji,i,l} \triangleq \left|1 - u_{ji,i,l}\tilde{\boldsymbol{h}}_{ji,i,l}^{H}\boldsymbol{p}_{ji,i,l}\right|^2 + \sum_{(j',l')\neq(j,l)}\left(\left|u_{ji,i,l}\tilde{\boldsymbol{h}}_{ji,i,l}\boldsymbol{p}_{j'i,i,l'}\right|^2 + |u_{ji,i,l}|^2\sigma_{ji,i,l}^2\right)$$

where $w_{ji,i,l}$ and $u_{ji,i,l}$ are auxiliary variables provides the equivalence between (41) and (42) according to [32]. Furthermore, the WMMSE problem can be solved by the coordinate descent algorithm [33]. However, the aforementioned precoding method is too complicated to implement as the computational complexity is proportional to the number of antennas at BS. To alleviate this overhead, we can leverage the conjugate beamforming to construct an effective channel with reduced dimension and then conduct the WMMSE precoding to eliminate the remained interference. Therefore, the design of the precoding becomes

$$\max_{\{\boldsymbol{p}'_{ji,i,l}\}} \sum_{j=1}^{J}\sum_{l=1}^{L} R_{ji,i,l}^{DL}$$
$$\text{s.t.} \quad \sum_{l=1}^{L} \text{Tr}(\boldsymbol{p}_{ji,i,l}\boldsymbol{p}_{ji,i,l}^{H}) < P_j \forall j$$
$$\boldsymbol{p}_{ji,i,l} = \boldsymbol{P}_{ni,i}^{c}\boldsymbol{p}'_{ji,i,l} \tag{43}$$

where $\boldsymbol{p}_{ji,i,l} \in \mathbb{C}^{N_{BSa}N_{BSe}\times 1}$, $\boldsymbol{p}'_{ji,i,l} \in \mathbb{C}^{N_{MS}\times 1}$ and $\boldsymbol{P}_{ni,i}^{c} \triangleq \tilde{\boldsymbol{H}}_{ni,i}^{H}$ with column normalization. By comparing the optimization problem (41) to (43), we see that the problem (43) essentially restricts the form of the feasible solutions of (41) to $\boldsymbol{P}_{ni,i}^{c}\boldsymbol{p}'_{ji,i,l}$. Thus the dimension of the targeting precoding is reduced from $N_{BSa}N_{BSe}$ to $N_{MS}$, which also means designing the precoder on the effective channel $\tilde{\boldsymbol{H}}_{ni,i}\tilde{\boldsymbol{H}}_{ni,i}^{H}$.

## VI. Simulations

In this section, we evaluate the performance of the pilot-data superposition scheme as well as our introduced DOA estimation algorithm. In the deployment of the cellular network, we consider 7 hexagonal cells in one cluster and eight co-scheduled MSs within each cell. Thus, for the received signal of each MS, it contains intra-cell interference from the other 7 MSs, inter-cell interference from the neighboring six cells and pilot contamination. For any MS, the average relative path-loss to any other MSs in different cells is set as 0.1, i.e., $\sqrt{\Lambda_{ng,g}/\Lambda_{ji,g}} = 0.1$. The antenna spacing for both MS and BS is set to be half-wavelength. The number of antennas at MS is assumed to be 8. BS employs 64 antennas, where the number of antennas on azimuth and horizon is assumed to be $8 \times 8$. The channel is assumed to be dominated by 4 clusters, and each cluster is involved in 1 resolvable path. The pilot and data correlation matrix $\boldsymbol{\Omega}_{ni,i}$ is randomly generated according to the product of two circular Gaussian matrices with normalization. Moreover, for performance evaluation and comparison purpose, we assume the uplink precoding methods are matched to ideal DODs of MSs, i.e., $\boldsymbol{P}(\boldsymbol{\omega}_{jg,g}) = \boldsymbol{A}(\boldsymbol{\omega}_{jg,g})$. It is important to note that our method does not depend on the underlying uplink precoding methods used at the MSs. However, different uplink beamforming/precoding at MSs will lead to different distributions of the third-factor matrix resulting in different convergence behaviors of the tensor decomposition in the M-step. The consequent convergence analysis related to the distribution of the third factor matrix is out of the scope of this paper. Finally, the pilot correlation coefficient $\rho_1$ is set to be less than $1/T$, where $T$ is the length of uplink pilots.

### A. Uplink SINR For Channel Estimation

In order to explain the CSI estimation performance of the pilot-data superposition system, we first investigate the distribution of the received SINR in the channel estimation stage with different allocated pilot power $\gamma$. Here, the SINR is defined by the power ratio between the received signal to all the interference plus noise in (16),

$$\text{SINR} = \frac{\|\bar{\boldsymbol{\mathcal{H}}}_{ni,i}(k)\times_3(\sqrt{\gamma(1-\gamma)}\bar{\boldsymbol{\Omega}}_{ni}(k) + \gamma\boldsymbol{I})\|^2}{\|\boldsymbol{\mathcal{I}}_p + \boldsymbol{\mathcal{I}}_u + \boldsymbol{\mathcal{I}}_e + \boldsymbol{\mathcal{W}}'_i\|^2}. \tag{44}$$

For the measurement of the SINR distribution, we calculate the experimental CDF of the SINR by fixing a receiving channel realization and varying the interference pattern according to (44). From Fig. 1, we can observe the gap between two consecutive CDF curves shrinks as $\gamma$ increases. One is because of the magnified interference power from pilot contamination, though the interference caused by superimposed data becomes smaller at the same time. On the other hand, the power contributing to the CSI estimation decreases with decreasing $\gamma$.

### B. Convergence Behavior

We now investigate the convergence behavior of the EM-algorithm. For the EM algorithm, the initialization of the DOAs is drawn from uniform distributions in the angle space. Meanwhile, the interference estimation is initialized through the high

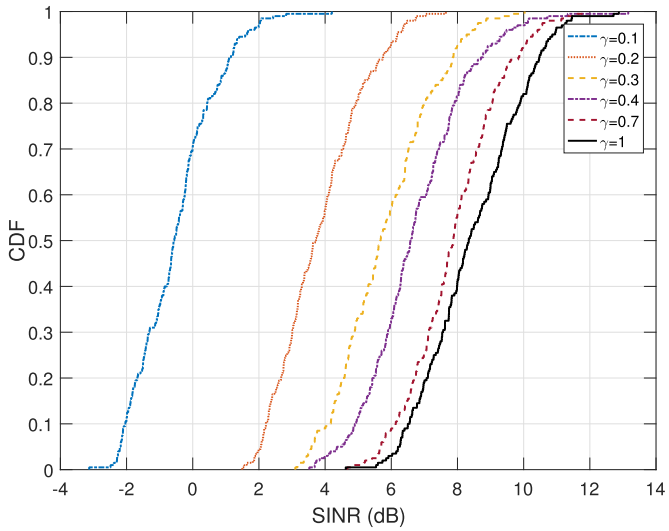



Fig. 1. The CDF of uplink pilot SINR under different allocation power $\gamma$ .

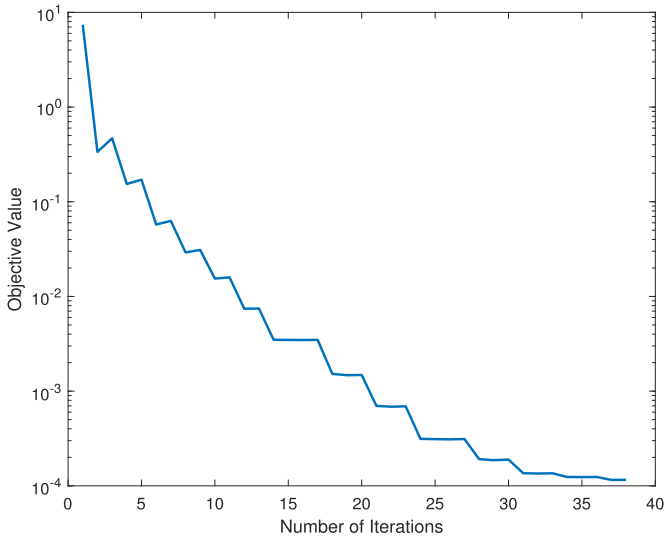



Fig. 2. The convergence of our introduced algorithm.

order SVD [21]. According to this initialization, the DOAs can be retrieved due to the essential uniqueness of the CP decomposition. The resulting convergence of the algorithm is shown in Fig. 2, where the objective value is calculated using $\|\boldsymbol{\mathcal{Y}}_{ni} - \boldsymbol{\mathcal{R}}_{ni,i}(\boldsymbol{u}, \boldsymbol{v}) - \boldsymbol{\mathcal{I}}\|^2$, where $\boldsymbol{u}, \boldsymbol{v}$ is the estimated DOAs according to every M-step. During the numerical simulation, we also observe that merely conducting the CP decomposition using alternative least squared (ALS) cannot lead to the convergence. The main reason is the direct ALS method does not take advantage of the Vandermonde structure of the tensor factor matrix. Meanwhile, the existing interference needs to be mitigated before processing. Moreover, in simulations, rank-deficiency of the underlying CP model can occur. To avoid the loss of ranks in the decomposition, the rank regularization term can be added in the objective as discussed in [34].

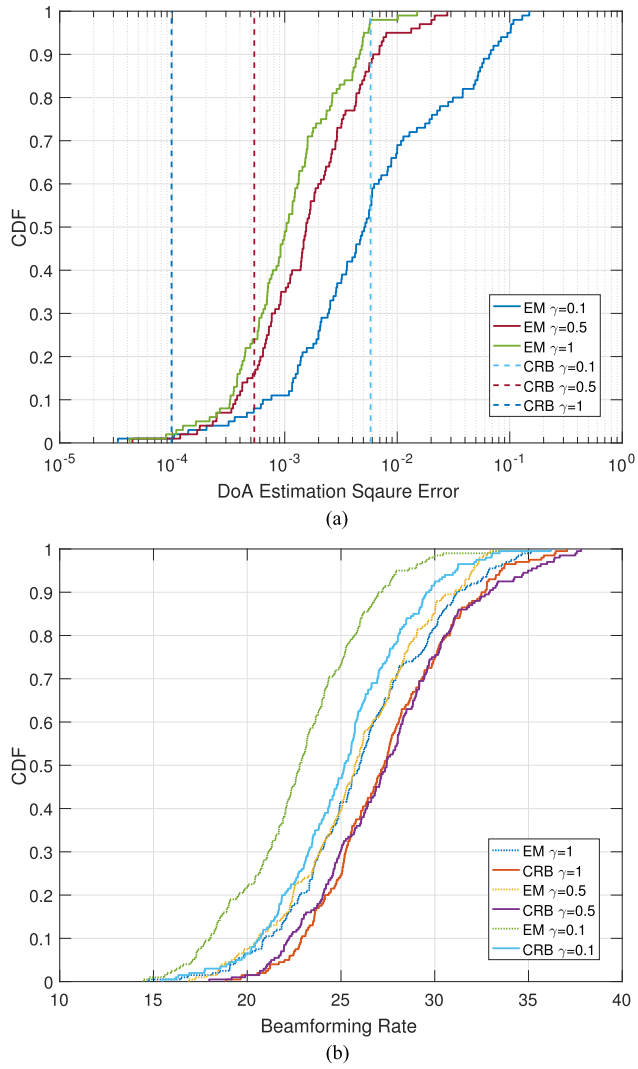



Fig. 3. Stochastic CRLB versus allocated pilot power $\gamma$ when $\rho_1 = 0$, $\sigma = 10^{-4}$: (a) the squared error of estimated DOAs (b) beamforming rate.

### C. Channel Estimation Performance

Fig. 3(a) shows our introduced channel estimation algorithm compared with the CRLB by squared error, where the curves for the CRLB and the EM algorithm are generated from the same channel realization. We can see the gap between two consecutive CDF curves shrinks as $\gamma$ increases. This shrinkage is also similar to the behavior observed in Fig. 1. In Fig 3(b), we compared the performance of conjugate beamforming by using the estimated DOAs and the ground truth DOAs combined with a Gaussian distributed error, where the variance of the Gaussian error is the CRLB. As we can see, the beamforming performance using these two DOAs are very close.

### D. Uplink and Downlink Sum Rate

We now consider the uplink rate distribution in Fig. 4. For the uplink, the receiving method is conducted by the matched filter. Fig. 4(a) depicts the rate performance without CSI estimation errors. It reveals that the pilot power is negatively correlated

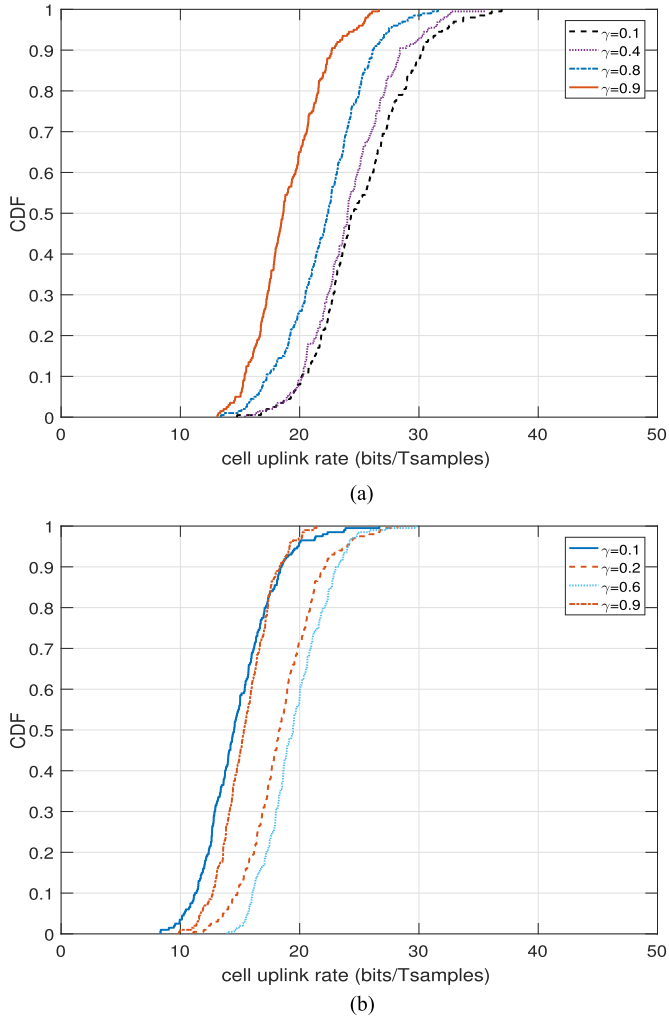



Fig. 4. The CDF of cell uplink rate under different pilot power $\gamma$: (a) perfect CSI and (b) with estimation error.

to the achieved rate. Moreover, when the CSI estimation error exists, lowing pilot power cannot lead to more accurate CSI, which results in the deterioration of the rate performance. However Fig. 4(b) reveals that the correlation between the pilot power and the uplink rate is no longer monotonic. Interestingly, we can observe that the best rate performance happens when allocating a moderate amount of pilot power. As a dual case to the uplink, Fig. 5 depicts the downlink sum-rate performance. Fig. 5(a) compares different precoding algorithms using the perfect CSI, where C-BF represents the conjugate beamforming, C-WMMSE represents the WMMSE combined with the conjugate beamforming as introduced in Section V-C and WMMSE is directly conducted by solving the problem (42). Moreover, Fig. 5(b) to Fig. 5(d) shows the precoding performance with estimation errors. From all these figures, we can observe that the WMMSE can obtain the best performance gain as the estimated channel becomes more accurate. On the other hand, the channel accuracy gain for conjugate beamforming is not significant compared with other methods. The WMMSE can alleviate the inter-user interference by leveraging other users' CSI through

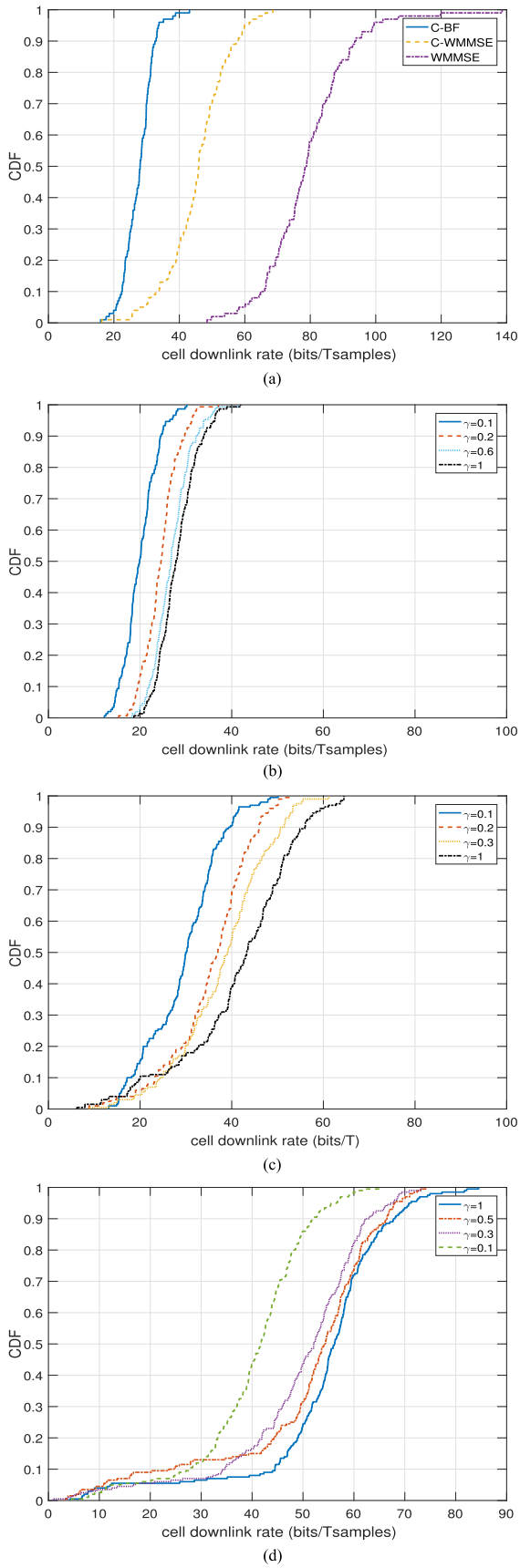



Fig. 5. The CDF of cell downlink transmission rate: (a) perfect CSI (b) C-BF with estimation error (c) C-WMMSE with estimation error (d) WMMMSE with estimation error.

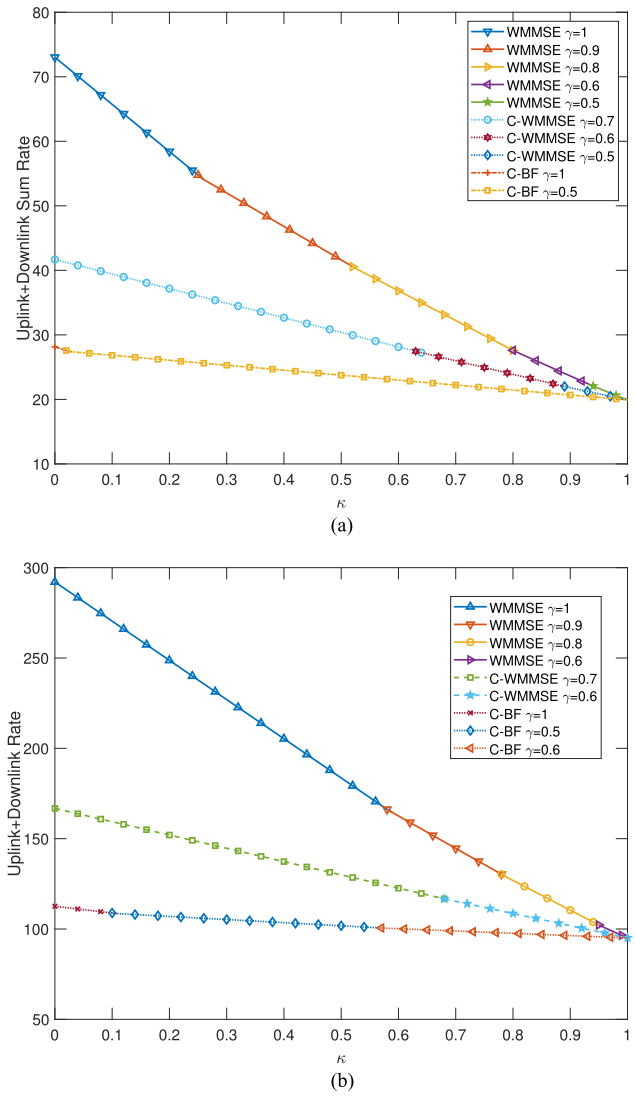



Fig. 6. The system sum rate versus the uplink importance factor $\kappa$: (a) $N_d = 0$ (b) $N_d = 3$.

a joint optimization procedure. However, the complexity of the WMMSE is too high to implement. Therefore, the C-WMMSE can be considered as a trade-off between complexity and performance. From Fig. 5(c), we see that when C-WMMSE uses the channel obtained at $\gamma = 0.1$, it still outperforms the conjugate beamforming at $\gamma = 1$.

### *E. Uplink and Downlink Tradeoff*

Finally, we consider the optimal trade-off curve of the uplink and downlink transmission, i.e., $\max_\gamma R(\kappa, \gamma)$, where $R(\kappa, \gamma)$ is defined in (31). $R^{UL}$ is calculated according to (35); $R^{DL}$ is calculated by (40) times $1 + N_d$. Therefore, the calculation of $R^{UL}$ and $R^{DL}$ are normalized to the same number of transmission resources. In system sum-rate, the ratio between $R^{UL}$ and $R^{DL}$ is controlled by the uplink importance factor $\gamma$. From Fig 6, we observe that when we employ the WMMSE, the downlink sum-rate can dominate the system rate. In this case, allocating more power on the pilot is always better. Otherwise, the pilot power and the data power should be carefully divided for the optimal transmission in the pilot-data superposition system. Note that in this paper, we only provide numerical evaluations for power allocation strategies, while the analytical characterization of the optimal resource allocation will be treated as a future extension to this work.

## VII. Conclusion

Conventional pilot-data structure in communications systems is conducted by splitting data and pilot through orthogonal manners. For the multi-user FD-MIMO system with a large number of active users, the conventional pilot-data structure requires greater overhead. To address this issue, the pilot-data superposition strategy is introduced to multi-user FD-MIMO systems. However, the pilot-data superposition scheme lowers the effective pilot power for channel estimation, hence reduces the accuracy of CSI estimation at the BS. A tensor decomposition based channel estimation algorithm is developed for uplink channel estimation through the EM framework to substantially increase the performance of the underlying CSI estimation. Simulation results demonstrate the superior performance of our algorithm in the limited pilot power regime. Furthermore, the sum-rate trade-off on uplink and downlink is considered as the most important aspect to determine the pilot power allocations with different downlink multi-user MIMO precoding strategies. The resulting optimal power allocation between pilot and data is numerically characterized depending on the underlying channel estimation algorithms, uplink receive processing, and downlink precoding.

## Appendix A
## Derivation of the EM Algorithm for DOAs Estimation

### *A. In E-step*

we have

$$Q(\boldsymbol{u}_{ni,i}, \boldsymbol{v}_{ni,i} | \boldsymbol{u}_{ni,i}(t), \boldsymbol{v}_{ni,i}(t)) = \mathbb{E}_{\boldsymbol{\mathcal{I}}|\boldsymbol{\mathcal{Y}}_i;\boldsymbol{u}_{ni,i}(t),\boldsymbol{v}_{ni,i}(t)}(\log P(\boldsymbol{\mathcal{Y}}_i, \boldsymbol{\mathcal{I}}; \boldsymbol{u}_{ni,i}, \boldsymbol{v}_{ni,i}))$$

where $t$ represents the $t$th iteration. From the Bayesian formula, we know that $P(\boldsymbol{\mathcal{Y}}|\boldsymbol{\mathcal{I}}; \boldsymbol{u}_{ni,i}(t), \boldsymbol{v}_{ni,i}(t))$

$$P(\boldsymbol{\mathcal{I}}|\boldsymbol{\mathcal{Y}}; \boldsymbol{u}_{ni,i}(t), \boldsymbol{v}_{ni,i}(t)) = P(\boldsymbol{\mathcal{Y}}|\boldsymbol{\mathcal{I}}; \boldsymbol{u}_{ni,i}(t), \boldsymbol{v}_{ni,i}(t)) P(\boldsymbol{\mathcal{I}}) / P(\boldsymbol{\mathcal{Y}}; \boldsymbol{u}_{ni,i}(t), \boldsymbol{v}_{ni,i}(t))$$

Since $\boldsymbol{\mathcal{W}}'_i$ is i.i.d Gaussian distribution, we have

$$\log P(\boldsymbol{\mathcal{Y}}_i|\boldsymbol{\mathcal{I}}; \boldsymbol{u}_{ni,i}, \boldsymbol{v}_{ni,i}) \propto -\|\boldsymbol{\mathcal{Y}}_i - \boldsymbol{\mathcal{R}}_{ni,i}(\boldsymbol{u}_{ni,i}, \boldsymbol{v}_{ni,i}) - \boldsymbol{\mathcal{I}}\|^2$$

where

$$\boldsymbol{\mathcal{R}}_{ni,i}(\boldsymbol{u}_{ni,i}, \boldsymbol{v}_{ni,i}) = \bar{\boldsymbol{\mathcal{H}}}_{ni,i} \times_3 (\sqrt{\gamma(1-\gamma)}\bar{\boldsymbol{\Omega}}_{ni}(k) + \gamma \boldsymbol{I}) \tag{45}$$

by (16). Therefore,

$$Q(\boldsymbol{u}_{ni,i}, \boldsymbol{v}_{ni,i} | \boldsymbol{u}_{ni,i}(t), \boldsymbol{v}_{ni,i}(t)) \propto \mathbb{E}_{\boldsymbol{\mathcal{I}}|\boldsymbol{\mathcal{Y}}_i;\boldsymbol{u}_{ni,i}(t),\boldsymbol{v}_{ni,i}(t)}(\log P(\boldsymbol{\mathcal{Y}}_i|\boldsymbol{\mathcal{I}}; \boldsymbol{u}_{ni,i}, \boldsymbol{v}_{ni,i})) \propto \mathbb{E}_{\boldsymbol{\mathcal{I}}|\boldsymbol{\mathcal{Y}}_i;\boldsymbol{u}_{ni,i}(t),\boldsymbol{v}_{ni,i}(t)} \|\boldsymbol{\mathcal{Y}}_i - \boldsymbol{\mathcal{R}}_{ni,i}(\boldsymbol{u}_{ni,i}, \boldsymbol{v}_{ni,i}) - \boldsymbol{\mathcal{I}}\|^2$$

### B. In M-Step

we have

$$\max_{\boldsymbol{u}_{ni,i},\boldsymbol{v}_{ni,i}} Q(\boldsymbol{u}_{ni,i},\boldsymbol{v}_{ni,i}|\boldsymbol{u}_{ni,i}(t),\boldsymbol{v}_{ni,i}(t))$$

which is equivalent to

$$\min_{\boldsymbol{u}_{ni,i},\boldsymbol{v}_{ni,i}} \mathbb{E}\|\boldsymbol{\mathcal{Y}}_i - \boldsymbol{\mathcal{R}}_{ni,i}(\boldsymbol{u}_{ni,i},\boldsymbol{v}_{ni,i}) - \boldsymbol{\mathcal{I}}\|^2 \tag{46}$$

where the expectation is taking into the conditional random variable $\boldsymbol{\mathcal{I}}|\boldsymbol{\mathcal{Y}}_i;\boldsymbol{u}_{ni,i}(t),\boldsymbol{v}_{ni,i}(t)$. To solve the problem (46), we can introduce an auxiliary tensor variable $\boldsymbol{\mathcal{F}}$. Then It can be equivalently written as

$$\begin{aligned}\min_{\boldsymbol{u}_{ni,i},\boldsymbol{v}_{ni,i},\boldsymbol{\mathcal{F}}} \ & \mathbb{E}\|\boldsymbol{\mathcal{Y}}_i - \boldsymbol{\mathcal{F}} - \boldsymbol{\mathcal{I}}\|^2 \\ \text{s.t.} \ & \boldsymbol{\mathcal{R}}_{ni,i}(\boldsymbol{u}_{ni,i},\boldsymbol{v}_{ni,i}) = \boldsymbol{\mathcal{F}}\end{aligned}$$

By $\boldsymbol{\mathcal{F}} = [\![\boldsymbol{F}^{(1)},\boldsymbol{F}^{(2)},\boldsymbol{F}^{(3)}]\!]$, the optimization problem then can be expressed as

$$\begin{aligned}\min_{\boldsymbol{u}_{ni,i},\boldsymbol{v}_{ni,i},\boldsymbol{F}^{(1)},\boldsymbol{F}^{(2)},\boldsymbol{F}^{(3)}} \ & \mathbb{E}\|\boldsymbol{\mathcal{Y}}_i - [\![\boldsymbol{F}^{(1)},\boldsymbol{F}^{(2)},\boldsymbol{F}^{(3)}]\!] - \boldsymbol{\mathcal{I}}\|^2 \\ \text{s.t.} \ & \boldsymbol{F}^{(1)} = \boldsymbol{A}(\boldsymbol{u}_{ni,i}) \\ & \boldsymbol{F}^{(2)} = \boldsymbol{A}(\boldsymbol{v}_{ni,i})\end{aligned}$$

Furthermore, the above problem can be solved by the following penalty method

$$\begin{aligned}\min_{\boldsymbol{u}_{ni,i},\boldsymbol{v}_{ni,i},\boldsymbol{F}^{(1)},\boldsymbol{F}^{(2)},\boldsymbol{F}^{(3)}} & \mathbb{E}\|\boldsymbol{\mathcal{Y}}_i - [\![\boldsymbol{F}^{(1)},\boldsymbol{F}^{(2)},\boldsymbol{F}^{(3)}]\!] - \boldsymbol{\mathcal{I}}\|^2 \\ & + \mu_l\left(\|\boldsymbol{F}^{(1)} - \boldsymbol{A}(\boldsymbol{u}_{ni,i})\|^2 + \|\boldsymbol{F}^{(2)} - \boldsymbol{A}(\boldsymbol{v}_{ni,i})\|^2\right)\end{aligned}$$

where $\mu_l$ will gradually increase during the iteration, i.e., $\mu_l \longrightarrow \infty$ as $l \longrightarrow \infty$. At the $l$th iteration, the solution of the previous iteration, i.e., $\boldsymbol{F}^{(1)}(\mu_{l-1})$ and $\boldsymbol{u}_{ni,i}(\mu_{l-1})$, can be used as the warm-starter. Furthermore, for any given $\mu_l$, the problem can be solved by the alternative linear minimum mean squared error (ALMMSE), where the factor matrices $\boldsymbol{F}^{(1)}$, $\boldsymbol{F}^{(2)}$, $\boldsymbol{F}^{(3)}$ and the parameters $\boldsymbol{u}_{ni,i}$, $\boldsymbol{v}_{ni,i}$ are iteratively updated by

$$\begin{aligned}\boldsymbol{F}^{(1)} = \arg\min_{\boldsymbol{F}^{(1)}} & \mathbb{E}\left\|\boldsymbol{Y}_{i(1)} - \boldsymbol{F}^{(1)}(\boldsymbol{F}^{(3)} \odot \boldsymbol{F}^{(2)})^T - \boldsymbol{I}_{(1)}\right\|^2 \\ & + \mu_l\|\boldsymbol{F}^{(1)} - \boldsymbol{A}(\boldsymbol{u}_{ni,i})\|^2\end{aligned} \tag{47}$$

$$\boldsymbol{F}^{(2)} = \arg\min_{\boldsymbol{F}^{(2)}} \mathbb{E}\left\|\boldsymbol{Y}_{i(2)} - \boldsymbol{F}^{(2)}(\boldsymbol{F}^{(3)} \odot \boldsymbol{F}^{(1)})^T - \boldsymbol{I}_{(2)}\right\|^2 \tag{48}$$

$$+ \mu_l\|\boldsymbol{F}^{(2)} - \boldsymbol{A}(\boldsymbol{v}_{ni,i})\|^2 \tag{49}$$

$$\boldsymbol{F}^{(3)} = \arg\min_{\boldsymbol{F}^{(3)}} \mathbb{E}\left\|\boldsymbol{Y}_{i(3)} - \boldsymbol{F}^{(3)}(\boldsymbol{F}^{(2)} \odot \boldsymbol{F}^{(1)})^T - \boldsymbol{I}_{(3)}\right\|^2 \tag{50}$$

$$\boldsymbol{u}_{ni,i} = \arg\min_{\boldsymbol{u}_{ni,i}} \|\boldsymbol{F}^{(1)} - \boldsymbol{A}(\boldsymbol{u}_{ni,i})\|^2 \tag{51}$$

$$\boldsymbol{v}_{ni,i} = \arg\min_{\boldsymbol{v}_{ni,i}} \|\boldsymbol{F}^{(2)} - \boldsymbol{A}(\boldsymbol{v}_{ni,i})\|^2 \tag{52}$$

where the above iterative decomposition is terminated as certain convergence condition meets, such as the difference between two consecutive updated factor matrices. Correspondingly, the close form solution of (47), (48) and (50) are respectively given by

$$\begin{aligned}\hat{\boldsymbol{F}}^{(1)} = & \left[(\boldsymbol{Y}_{i(1)} - \mathbb{E}(\boldsymbol{I}_{(1)}))(\boldsymbol{F}^{(3)} \odot \boldsymbol{F}^{(2)})^* + \mu_l\boldsymbol{A}(\boldsymbol{u}_{ni,i})\right] \\ & \times \left[(\boldsymbol{F}^{(3)} \odot \boldsymbol{F}^{(2)})^T(\boldsymbol{F}^{(3)} \odot \boldsymbol{F}^{(2)})^* + \mu_l\boldsymbol{I}\right]^{-1}\end{aligned} \tag{53}$$

$$\begin{aligned}\hat{\boldsymbol{F}}^{(2)} = & \left[(\boldsymbol{Y}_{i(2)} - \mathbb{E}(\boldsymbol{I}_{(2)}))(\boldsymbol{F}^{(3)} \odot \boldsymbol{F}^{(1)})^* + \mu_l\boldsymbol{A}(\boldsymbol{v}_{ni,i})\right] \\ & \times \left[(\boldsymbol{F}^{(3)} \odot \boldsymbol{F}^{(1)})^T(\boldsymbol{F}^{(3)} \odot \boldsymbol{F}^{(1)})^* + \mu_l\boldsymbol{I}\right]^{-1}\end{aligned} \tag{54}$$

$$\hat{\boldsymbol{F}}^{(3)} = (\boldsymbol{Y}_{i(3)} - \mathbb{E}(\boldsymbol{I}_{(3)}))\left((\boldsymbol{F}^{(2)} \odot \boldsymbol{F}^{(1)})^T\right)^+ \tag{55}$$

in which the proof can be found in Appendix C. Moreover, The solution of (51) and (52) can be easily verified as the following match filters

$$\hat{u}^{(p)}_{ni,i,l} = \arg\max_u \frac{|\boldsymbol{a}^H(u)\boldsymbol{F}^{(2)}(:,p,l)|}{\|\boldsymbol{a}(u)\|\|\boldsymbol{F}^{(2)}(:,p,l)\|}$$

$$\hat{v}^{(p)}_{ni,i,l} = \arg\max_v \frac{|\boldsymbol{a}^H(v)\boldsymbol{F}^{(1)}(:,p,l)|}{\|\boldsymbol{a}(v)\|\|\boldsymbol{F}^{(1)}(:,p,l)\|}$$

## APPENDIX B
## M-STEP: FACTOR MATRIX ESTIMATION

In M-step, all the factor matrix estimation problems (47), (48), (50) can be generally treated as the following LMMSE estimator

$$\min_{\boldsymbol{W}} \mathbb{E}_{\boldsymbol{X},\boldsymbol{Y}}\|\boldsymbol{X} - \boldsymbol{W}\boldsymbol{Y}\|^2$$

where $\boldsymbol{X}$ is the hidden variable, $\boldsymbol{Y}$ is the observation. The optimal condition is equivalent to

$$\frac{\partial\mathbb{E}\|\boldsymbol{X} - \boldsymbol{W}\boldsymbol{Y}\|^2}{\partial\boldsymbol{W}} = \mathbb{E}\{(\boldsymbol{X} - \boldsymbol{W}\boldsymbol{Y})\boldsymbol{Y}^{\mathrm{H}}\} = \boldsymbol{0}$$

which leads to

$$\boldsymbol{W} = \boldsymbol{C}_{\boldsymbol{X},\boldsymbol{Y}}\boldsymbol{C}^{-1}_{\boldsymbol{Y},\boldsymbol{Y}} \tag{56}$$

where $\boldsymbol{C}_{\boldsymbol{X},\boldsymbol{Y}} \triangleq \mathbb{E}[\boldsymbol{X}\boldsymbol{Y}^H]$ and $\boldsymbol{C}_{\boldsymbol{Y},\boldsymbol{Y}} \triangleq \mathbb{E}[\boldsymbol{Y}\boldsymbol{Y}^H]$.

For (47) and (48), it can be further expressed as (57) and (58) shown at the bottom of this page. We see that in

$$\boldsymbol{F}^{(1)} = \arg\min_{\boldsymbol{F}^{(1)}} \mathbb{E}\left\|[\boldsymbol{Y}_{i(1)},\boldsymbol{0}] - \boldsymbol{F}^{(1)}\left[(\boldsymbol{F}^{(3)} \odot \boldsymbol{F}^{(2)})^T, -\sqrt{\mu_l}\boldsymbol{I}\right] - \left[\boldsymbol{I}_{(1)}, \sqrt{\mu_l}\boldsymbol{A}(\boldsymbol{u}_{ni,i})\boldsymbol{\Lambda}^{(1)}\right]\right\|^2 \tag{57}$$

$$\boldsymbol{F}^{(2)} = \arg\min_{\boldsymbol{F}^{(2)}} \mathbb{E}\left\|[\boldsymbol{Y}_{i(2)},\boldsymbol{0}] - \boldsymbol{F}^{(2)}\left[(\boldsymbol{F}^{(3)} \odot \boldsymbol{F}^{(1)})^T, -\sqrt{\mu_l}\boldsymbol{I}\right] - \left[\boldsymbol{I}_{(2)}, \sqrt{\mu_l}\boldsymbol{A}(\boldsymbol{v}_{ni,i})\boldsymbol{\Lambda}^{(2)}\right]\right\|^2 \tag{58}$$

(57), when $\boldsymbol{F}^{(3)}$ and $\boldsymbol{F}^{(2)}$ are fixed, the estimation of $\boldsymbol{F}^{(1)}$ can be treated as an LMMSE estimator, where the observation is $[(\boldsymbol{F}^{(3)} \odot \boldsymbol{F}^{(2)})^T, -\sqrt{\mu_l}\boldsymbol{I}]$ and the hidden variable is $[\boldsymbol{Y}_{i(1)}, \boldsymbol{0}] - [\boldsymbol{I}_{(1)}, \sqrt{\mu_l}\boldsymbol{A}(\boldsymbol{u}_{ni,i})\boldsymbol{\Lambda}^{(1)}]$. Thus, by using (56), we have

$$\begin{aligned}\boldsymbol{C}_{\boldsymbol{X},\boldsymbol{Y}} &= \mathbb{E}\left\{(\boldsymbol{Y}_{i(1)} - \boldsymbol{I}_{(1)})(\boldsymbol{F}^{(3)} \odot \boldsymbol{F}^{(2)})^* + \mu_l \boldsymbol{A}(\boldsymbol{u}_{ni,i})\boldsymbol{\Lambda}^{(1)}\right\} \\ &= (\boldsymbol{Y}_{i(1)} - \mathbb{E}(\boldsymbol{I}_{(1)}))(\boldsymbol{F}^{(3)} \odot \boldsymbol{F}^{(2)})^* + \mu_l \boldsymbol{A}(\boldsymbol{u}_{ni,i})\boldsymbol{\Lambda}^{(1)} \\ \boldsymbol{C}_{\boldsymbol{Y},\boldsymbol{Y}} &= \left[(\boldsymbol{F}^{(3)} \odot \boldsymbol{F}^{(2)})^T, -\sqrt{\mu_l}\boldsymbol{I}\right]\left[(\boldsymbol{F}^{(3)} \odot \boldsymbol{F}^{(2)})^*, -\sqrt{\mu_l}\boldsymbol{I}\right]^T \\ &= (\boldsymbol{F}^{(3)} \odot \boldsymbol{F}^{(2)})^T(\boldsymbol{F}^{(3)} \odot \boldsymbol{F}^{(2)})^* + \mu_l \boldsymbol{I}\end{aligned}$$

Therefore,

$$\begin{aligned}\hat{\boldsymbol{F}}^{(1)} = &\left[(\boldsymbol{Y}_{i(1)} - \mathbb{E}(\boldsymbol{I}_{(1)}))(\boldsymbol{F}^{(3)} \odot \boldsymbol{F}^{(2)})^* + \mu_l \boldsymbol{A}(\boldsymbol{u}_{ni,i})\boldsymbol{\Lambda}^{(1)}\right] \\ &\times \left[(\boldsymbol{F}^{(3)} \odot \boldsymbol{F}^{(2)})^T(\boldsymbol{F}^{(3)} \odot \boldsymbol{F}^{(2)})^* + \mu_l \boldsymbol{I}\right]^{-1}\end{aligned}$$

Similarly, the solution of (48) can be derived as

$$\begin{aligned}\hat{\boldsymbol{F}}^{(2)} = &\left[(\boldsymbol{Y}_{i(2)} - \mathbb{E}(\boldsymbol{I}_{(2)}))(\boldsymbol{F}^{(3)} \odot \boldsymbol{F}^{(1)})^* + \mu_l \boldsymbol{A}(\boldsymbol{v}_{ni,i})\boldsymbol{\Lambda}^{(2)}\right] \\ &\times \left[(\boldsymbol{F}^{(3)} \odot \boldsymbol{F}^{(1)})^T(\boldsymbol{F}^{(3)} \odot \boldsymbol{F}^{(1)})^* + \mu_l \boldsymbol{I}\right]^{-1}\end{aligned}$$

## Appendix C
## E-Step Calculation

At the $t$th E-step, the major concern is the calculation of the posterior mean value $\mathbb{E}(\boldsymbol{I}_{(1)}), \mathbb{E}(\boldsymbol{I}_{(2)})$ and $\mathbb{E}(\boldsymbol{I}_{(3)})$. As we know,

$$P(\boldsymbol{\mathcal{I}}|\boldsymbol{\mathcal{Y}}; \boldsymbol{u}_{ni,i}(t), \boldsymbol{v}_{ni,i}(t)) \propto P(\boldsymbol{\mathcal{Y}}|\boldsymbol{\mathcal{I}}; \boldsymbol{u}_{ni,i}(t), \boldsymbol{v}_{ni,i}(t))P(\boldsymbol{\mathcal{I}})$$

where $P(\boldsymbol{\mathcal{I}})$ is the PDF of interference,

$$P(\boldsymbol{\mathcal{Y}}|\boldsymbol{\mathcal{I}}; \boldsymbol{u}_{ni,i}(t), \boldsymbol{v}_{ni,i}(t)) \sim \mathcal{CN}(\boldsymbol{\mathcal{Y}}_i - \boldsymbol{\mathcal{R}}_{ni,i}(\boldsymbol{\varphi}(t)), \sigma^2\boldsymbol{I}, \boldsymbol{0})$$

As the discussion in [35], we apply a Gaussian approximation to the interference. Here, PDF of the interference is defined in Appendix-D which is denoted as

$$P(\boldsymbol{\mathcal{I}}) \sim \mathcal{CN}(\boldsymbol{0}, \boldsymbol{R}_{\boldsymbol{\mathcal{I}}}, \boldsymbol{M}_{\boldsymbol{\mathcal{I}}})$$

where $\sim$ stands for satisfying the distribution, $\boldsymbol{R}_I$ and $\boldsymbol{M}_I$ are given by (61) and (60). Meanwhile, we have

$$\begin{aligned}P(\boldsymbol{\mathcal{Y}}|\boldsymbol{\mathcal{I}}; \boldsymbol{u}_{ni,i}(t), \boldsymbol{v}_{ni,i}(t)) \propto \exp\Bigg(&\left[(\boldsymbol{i}-\boldsymbol{\mu}_i)^H \quad (\boldsymbol{i}-\boldsymbol{\mu}_i)^T\right] \\ &\times \begin{bmatrix}\sigma^2\boldsymbol{I} & \boldsymbol{0} \\ \boldsymbol{0} & \sigma^2\boldsymbol{I}\end{bmatrix}^{-1}\begin{bmatrix}\boldsymbol{i}-\boldsymbol{\mu}_i \\ (\boldsymbol{i}-\boldsymbol{\mu}_i)^*\end{bmatrix}\Bigg)\end{aligned}$$

where $\boldsymbol{i} = vec(\boldsymbol{\mathcal{I}})$ and $\boldsymbol{\mu}_i = vec(\boldsymbol{\mathcal{Y}}_i - \boldsymbol{\mathcal{R}}_{ni,i}(\boldsymbol{\varphi}(t)))$. As the product of Gaussian PDFs are still Gaussian, we can obtain the posterior mean value as,

$$\begin{aligned}\begin{bmatrix}\boldsymbol{m_i} \\ \boldsymbol{m_i^*}\end{bmatrix} = &\left(\begin{bmatrix}\boldsymbol{R}_I & \boldsymbol{M}_I \\ \boldsymbol{M}_I^* & \boldsymbol{R}_I^*\end{bmatrix}^{-1} + \begin{bmatrix}\sigma^2\boldsymbol{I} & \boldsymbol{0} \\ \boldsymbol{0} & \sigma^2\boldsymbol{I}\end{bmatrix}^{-1}\right)^{-1} \\ &\times \begin{bmatrix}\sigma^2\boldsymbol{I} & \boldsymbol{0} \\ \boldsymbol{0} & \sigma^2\boldsymbol{I}\end{bmatrix}^{-1}\begin{bmatrix}\boldsymbol{\mu}_i \\ \boldsymbol{\mu}_i{}^*\end{bmatrix} \\ = &\left(\sigma^2\begin{bmatrix}\boldsymbol{R}_I & \boldsymbol{M}_I \\ \boldsymbol{M}_I^* & \boldsymbol{R}_I^*\end{bmatrix}^{-1} + \boldsymbol{I}\right)^{-1}\begin{bmatrix}\boldsymbol{\mu}_i \\ \boldsymbol{\mu}_i{}^*\end{bmatrix} \\ = &\begin{bmatrix}\boldsymbol{R}_I & \boldsymbol{M}_I \\ \boldsymbol{M}_I^* & \boldsymbol{R}_I^*\end{bmatrix}\left(\sigma^2\boldsymbol{I} + \begin{bmatrix}\boldsymbol{R}_I & \boldsymbol{M}_I \\ \boldsymbol{M}_I^* & \boldsymbol{R}_I^*\end{bmatrix}\right)^{-1}\begin{bmatrix}\boldsymbol{\mu}_i \\ \boldsymbol{\mu}_i{}^*\end{bmatrix}.\end{aligned} \tag{59}$$

By rearranging the entries of $\boldsymbol{m_i}$ according to the reverse order of vectorization, we can easily arrive at $\mathbb{E}(\boldsymbol{I}_{(1)})$, $\mathbb{E}(\boldsymbol{I}_{(2)})$ and $\mathbb{E}(\boldsymbol{I}_{(3)})$.

## Appendix D
## Prior Distribution of Interference

For analysis, we can derive the Gaussian approximation of the interference as

$$P(\boldsymbol{\mathcal{I}}) \sim \mathcal{CN}(\boldsymbol{0}, \boldsymbol{R}_{\boldsymbol{\mathcal{I}}}, \boldsymbol{M}_{\boldsymbol{\mathcal{I}}})$$

in which $\mathcal{CN}(\boldsymbol{\mu}, \boldsymbol{R}, \boldsymbol{C})$ represents complex Gaussian probability function with the mean value $\boldsymbol{\mu}$, covariance matrix $\boldsymbol{R}_{\boldsymbol{\mathcal{I}}}$ and pseudo covariance matrix $\boldsymbol{M}_{\boldsymbol{\mathcal{I}}}$. In details,

$$P(\boldsymbol{\mathcal{I}}) \propto \exp\left(\begin{bmatrix}\boldsymbol{i}^H & \boldsymbol{i}^T\end{bmatrix}\begin{bmatrix}\boldsymbol{R}_I & \boldsymbol{M}_I \\ \boldsymbol{M}_I^* & \boldsymbol{R}_I^*\end{bmatrix}^{-1}\begin{bmatrix}\boldsymbol{i} \\ \boldsymbol{i}^*\end{bmatrix}\right)$$

where $\boldsymbol{i} \triangleq vec(\boldsymbol{\mathcal{I}})$ and $\text{vec}(\boldsymbol{\mathcal{I}}) = \text{vec}(\boldsymbol{I}_{u(3)}) + \text{vec}(\boldsymbol{I}_{p(3)}) + \text{vec}(\boldsymbol{I}_{e(3)})$ in which

$$\begin{aligned}\boldsymbol{I}_{u(3)} = &\sum_{j\neq n}\sqrt{\Lambda_{ji,i}}\left(\sqrt{\gamma(1-\gamma)}\bar{\boldsymbol{\Omega}}_{ni}^T(k) + \gamma\rho_1\boldsymbol{1}\right) \\ &\times \boldsymbol{P}^H(\boldsymbol{\omega}_{ji,i})\boldsymbol{A}(\boldsymbol{\omega}_{ji,i})\boldsymbol{\Lambda}_0(\boldsymbol{\alpha}_{ji,i})[(\boldsymbol{A}(\boldsymbol{v}_{ji,i}) \odot \boldsymbol{A}(\boldsymbol{u}_{ji,i})]^T \\ \boldsymbol{I}_{p(3)} = &\sum_{g\neq i}\sqrt{\Lambda_{ng,i}}\left(\sqrt{\gamma(1-\gamma)}\bar{\boldsymbol{\Omega}}_{ng}^T(k) + \gamma\boldsymbol{I}\right) \\ &\times \boldsymbol{P}^H(\boldsymbol{\omega}_{ng,g})\boldsymbol{A}(\boldsymbol{\omega}_{ng,i})\boldsymbol{\Lambda}_0(\boldsymbol{\alpha}_{ng,i})[\boldsymbol{A}(\boldsymbol{v}_{ng,i}) \odot \boldsymbol{A}(\boldsymbol{u}_{ng,i})]^T \\ \boldsymbol{I}_{e(3)} = &\sum_{g\neq i}\sum_{j\neq n}\sqrt{\Lambda_{jg,i}}\left(\sqrt{\gamma(1-\gamma)}\boldsymbol{\Omega}_{jg}^T(k) + \rho_1\gamma\boldsymbol{1}\right) \\ &\times \boldsymbol{P}^H(\boldsymbol{\omega}_{jg,g})\boldsymbol{A}(\boldsymbol{\omega}_{jg,i})\boldsymbol{\Lambda}_0(\boldsymbol{\alpha}_{jg,i})[\boldsymbol{A}(\boldsymbol{v}_{jg,i}) \odot \boldsymbol{A}(\boldsymbol{u}_{jg,i})]^T\end{aligned}$$

Furthermore, we assume the path gains are independent to each others and $\alpha_{ji,i,l} \triangleq \mathcal{CN}(0, 1, \varsigma)$. Therefore, the pseudo-covariance matrix and covariance matrix of the interference

terms are respectively given by

$$\boldsymbol{M}_{\boldsymbol{\mathcal{I}}} \triangleq \mathbb{E}(\mathrm{vec}(\boldsymbol{\mathcal{I}})\mathrm{vec}(\boldsymbol{\mathcal{I}})^T) = \sum_{k=1}^{3} \boldsymbol{M}_k \tag{60}$$

$$\boldsymbol{R}_{\boldsymbol{\mathcal{I}}} \triangleq \mathbb{E}(\mathrm{vec}(\boldsymbol{\mathcal{I}})\mathrm{vec}(\boldsymbol{\mathcal{I}})^H) = \sum_{k=1}^{3} \boldsymbol{R}_k \tag{61}$$

where

$$\begin{aligned}
\boldsymbol{M}_1 &\triangleq \sum_{j\neq n} \Lambda_{ji,i} \mathbb{E}\left(\boldsymbol{p}_{ji,i}^{(1)} \boldsymbol{p}_{ji,i}^{(1)T}\right)\\
\boldsymbol{R}_1 &\triangleq \sum_{j\neq n} \Lambda_{ji,i} \mathbb{E}\left(\boldsymbol{p}_{ji,i}^{(1)} \boldsymbol{p}_{ji,i}^{(1)H}\right)\\
\boldsymbol{p}_{ji,i}^{(1)} &\triangleq ([\boldsymbol{A}(\boldsymbol{v}_{ji,i}) \odot \boldsymbol{A}(\boldsymbol{u}_{ji,i})] \odot [(\sqrt{\gamma(1-\gamma)}\boldsymbol{\Omega}_{ji}^T(k)\\
&\quad + \gamma\rho_1 \mathbf{1})\boldsymbol{P}^H(\boldsymbol{\omega}_{ji,i})\boldsymbol{A}(\boldsymbol{\omega}_{ji,i})])\boldsymbol{\alpha}_{ji,i}\\
\boldsymbol{M}_2 &\triangleq \sum_{g\neq i} \Lambda_{ji,i} \mathbb{E}\left(\boldsymbol{p}_{ng,i}^{(2)} \boldsymbol{p}_{ng,i}^{(2)T}\right)\\
\boldsymbol{R}_2 &\triangleq \sum_{g\neq i} \Lambda_{ji,i} \mathbb{E}\left(\boldsymbol{p}_{ng,i}^{(2)} \boldsymbol{p}_{ng,i}^{(2)H}\right)\\
\boldsymbol{p}_{ng,i}^{(2)} &\triangleq ([\boldsymbol{A}(\boldsymbol{v}_{ng,i}) \odot \boldsymbol{A}(\boldsymbol{u}_{ng,i})] \odot [(\sqrt{\gamma(1-\gamma)}\boldsymbol{\Omega}_{ng}^T(k)\\
&\quad + \gamma\boldsymbol{I})\boldsymbol{P}^H(\boldsymbol{\omega}_{ng,g})\boldsymbol{A}(\boldsymbol{\omega}_{ng,i})])\boldsymbol{\alpha}_{ng,i}\\
\boldsymbol{M}_3 &\triangleq \sum_{g\neq i}\sum_{j\neq n} \Lambda_{ng,i} \mathbb{E}\left(\boldsymbol{p}_{jg,i}^{(3)} \boldsymbol{p}_{jg,i}^{(3)T}\right)\\
\boldsymbol{R}_3 &\triangleq \sum_{g\neq i}\sum_{j\neq n} \Lambda_{jg,i} \mathbb{E}\left(\boldsymbol{p}_{jg,i}^{(3)} \boldsymbol{p}_{jg,i}^{(3)H}\right)\\
\boldsymbol{p}_{jg,i}^{(3)} &\triangleq ([\boldsymbol{A}(\boldsymbol{v}_{jg,i}) \odot \boldsymbol{A}(\boldsymbol{u}_{jg,i})] \odot [(\sqrt{\gamma(1-\gamma)}\boldsymbol{\Omega}_{jg}^T(k)\\
&\quad + \gamma\rho_1 \mathbf{1})\boldsymbol{P}^H(\boldsymbol{\omega}_{jg,g})\boldsymbol{A}(\boldsymbol{\omega}_{jg,i})])\boldsymbol{\alpha}_{jg,i}
\end{aligned}$$

Moreover, when we define

$$\begin{aligned}
\boldsymbol{M} &\triangleq \mathbb{E}_v(\boldsymbol{a}(v)\boldsymbol{a}^T(v))\\
\boldsymbol{M}' &\triangleq \mathbb{E}_{\boldsymbol{\omega}}(\boldsymbol{P}^H(\boldsymbol{\omega})\boldsymbol{A}(\boldsymbol{\omega})\boldsymbol{A}^T(\boldsymbol{\omega})\boldsymbol{P}^*(\boldsymbol{\omega}))\\
\boldsymbol{R} &= \mathbb{E}_v(\boldsymbol{a}(v)\boldsymbol{a}^H(v))\\
\boldsymbol{R}' &= \mathbb{E}_{\boldsymbol{\omega}}(\boldsymbol{P}^H(\boldsymbol{\omega})\boldsymbol{A}(\boldsymbol{\omega})\boldsymbol{A}^H(\boldsymbol{\omega})\boldsymbol{P}(\boldsymbol{\omega}))
\end{aligned}$$

we can obtain

$$\begin{aligned}
\mathbb{E}\left(\boldsymbol{p}_{ji,i}^{(1)} \boldsymbol{p}_{ji,i}^{(1)T}\right) &= \varsigma \boldsymbol{M} \otimes \boldsymbol{M} \otimes \mathbb{E}\left(\tilde{\boldsymbol{X}}_{ji}^T \boldsymbol{M}' \tilde{\boldsymbol{X}}_{ji}\right)\\
\mathbb{E}\left(\boldsymbol{p}_{ji,i}^{(1)} \boldsymbol{p}_{ji,i}^{(1)H}\right) &= \boldsymbol{R} \otimes \boldsymbol{R} \otimes \mathbb{E}\left(\tilde{\boldsymbol{X}}_{ji}^T \boldsymbol{R}' \tilde{\boldsymbol{X}}_{ji}^*\right)
\end{aligned}$$

where

$$\tilde{\boldsymbol{X}}_{ji} \triangleq \sqrt{\gamma(1-\gamma)}\boldsymbol{\Omega}_{ji} + \gamma\rho_1 \mathbf{1}$$

$$\begin{aligned}
\mathbb{E}\left(\boldsymbol{p}_{ng,i}^{(2)} \boldsymbol{p}_{ng,i}^{(2)T}\right) &= \varsigma \boldsymbol{M} \otimes \boldsymbol{M} \otimes \mathbb{E}\left(\tilde{\boldsymbol{X}}_{ng}^T \boldsymbol{M}' \tilde{\boldsymbol{X}}_{ng}\right)\\
\mathbb{E}\left(\boldsymbol{p}_{ng,i}^{(2)} \boldsymbol{p}_{ng,i}^{(2)H}\right) &= \boldsymbol{R}(v) \otimes \boldsymbol{R}(u) \otimes \mathbb{E}\left(\tilde{\boldsymbol{X}}_{ng}^T \boldsymbol{R}' \tilde{\boldsymbol{X}}_{ng}^*\right)
\end{aligned}$$

where

$$\tilde{\boldsymbol{X}}_{ng} = \sqrt{\gamma(1-\gamma)}\boldsymbol{\Omega}_{ng} + \gamma\rho_1 \boldsymbol{I}$$

in which

$$\begin{aligned}
\mathbb{E}\left(\boldsymbol{p}_{jg,i}^{(3)} \boldsymbol{p}_{jg,i}^{(3)T}\right) &= \varsigma \boldsymbol{M}(v) \otimes \boldsymbol{M}(u)\\
&\quad \otimes \mathbb{E}\left(\tilde{\boldsymbol{X}}_{jg}^T \boldsymbol{M}'(\omega_{jg,g}, \omega_{jg,i}) \tilde{\boldsymbol{X}}_{jg}\right)\\
\mathbb{E}\left(\boldsymbol{p}_{jg,i}^{(3)} \boldsymbol{p}_{jg,i}^{(3)H}\right) &= \boldsymbol{R}(v) \otimes \boldsymbol{R}(u)\\
&\quad \otimes \mathbb{E}\left(\tilde{\boldsymbol{X}}_{jg}^T \boldsymbol{R}'(\omega_{jg,g}, \omega_{jg,i}) \tilde{\boldsymbol{X}}_{jg}^*\right)
\end{aligned}$$

in which

$$\tilde{\boldsymbol{X}}_{jg} = \sqrt{\gamma(1-\gamma)}\boldsymbol{\Omega}_{jg} + \gamma\rho_1 \mathbf{1}$$

Note that closed-form expressions of the above covariance matrices do not exist. Therefore, we use sample covariance matrices in our simulations which is widely used when the closed-form cannot be directly obtained [27], [36].

## Appendix E
## Cramér Rao Lower bound

In the expression of $\boldsymbol{R}_{\boldsymbol{y}}$ and $\boldsymbol{M}_{\boldsymbol{y}}$, the user channel terms $\mathbb{E}[\boldsymbol{A}(\boldsymbol{\varphi})\boldsymbol{\alpha}\boldsymbol{\alpha}^H\boldsymbol{A}^H(\boldsymbol{\varphi})]$ and $\mathbb{E}[\boldsymbol{A}(\boldsymbol{\varphi})\boldsymbol{\alpha}\boldsymbol{\alpha}^T\boldsymbol{A}^T(\boldsymbol{\varphi})]$ can be further expanded as

$$\begin{aligned}
&\sum_{p,l}\left(\boldsymbol{a}\left(v_{ni,i,l}^{(p)}\right)\boldsymbol{a}^H\left(v_{ni,i,l}^{(p)}\right) \otimes \boldsymbol{a}\left(u_{ni,i,l}^{(p)}\right)\boldsymbol{a}^H\left(u_{ni,i,l}^{(p)}\right)\right.\\
&\otimes \mathbb{E}\left(\tilde{\boldsymbol{X}}_{ni}^T \boldsymbol{P}^H(\boldsymbol{\omega}_{ni,i})\boldsymbol{a}\left(\omega_{ni,i,l}^{(p)}\right)\boldsymbol{a}^H\left(\omega_{ni,i,l}^{(p)}\right)\boldsymbol{P}\left(\boldsymbol{\omega}_{ni,i}\right)\tilde{\boldsymbol{X}}_{ni}^*\right)\\
&\varsigma\sum_{p,l}\left(\boldsymbol{a}\left(v_{ni,i,l}^{(p)}\right)\boldsymbol{a}^T\left(v_{ni,i,l}^{(p)}\right) \otimes \boldsymbol{a}\left(u_{ni,i,l}^{(p)}\right)\boldsymbol{a}^T\left(u_{ni,i,l}^{(p)}\right)\right.\\
&\otimes \mathbb{E}\left(\tilde{\boldsymbol{X}}_{ni}^T \boldsymbol{P}^H(\boldsymbol{\omega}_{ni,i})\boldsymbol{a}\left(\omega_{ni,i,l}^{(p)}\right)\boldsymbol{a}^T\left(\omega_{ni,i,l}^{(p)}\right)\boldsymbol{P}^*(\boldsymbol{\omega}_{ni,i})\tilde{\boldsymbol{X}}_{ni}\right)
\end{aligned}$$

respectively, where

$$\tilde{\boldsymbol{X}}_{ni} \triangleq \left(\sqrt{\gamma(1-\gamma)}\bar{\boldsymbol{\Omega}}_{ni} + \gamma\boldsymbol{I}\right)$$

Next, we take the derivative of $\frac{\partial \boldsymbol{R}_{\boldsymbol{y}}}{\partial v_{ni,i,l}^{(p)}}$. As

$$\begin{aligned}
&\frac{\partial \boldsymbol{a}\left(v_{ni,i,l}^{(p)}\right)\boldsymbol{a}^H\left(v_{ni,i,l}^{(p)}\right)}{\partial v_{ni,i,l}^{(p)}}\\
&= \tilde{\boldsymbol{a}}\left(v_{ni,i,l}^{(p)}\right)\boldsymbol{a}^H\left(v_{ni,i,l}^{(p)}\right) + \boldsymbol{a}\left(v_{ni,i,l}^{(p)}\right)\tilde{\boldsymbol{a}}^H\left(v_{ni,i,l}^{(p)}\right)\\
&\frac{\partial \boldsymbol{a}\left(v_{ni,i,l}^{(p)}\right)\boldsymbol{a}^T\left(v_{ni,i,l}^{(p)}\right)}{\partial v_{ni,i,l}^{(p)}}\\
&= \tilde{\boldsymbol{a}}\left(v_{ni,i,l}^{(p)}\right)\boldsymbol{a}^T\left(v_{ni,i,l}^{(p)}\right) + \boldsymbol{a}\left(v_{ni,i,l}^{(p)}\right)\tilde{\boldsymbol{a}}^T\left(v_{ni,i,l}^{(p)}\right)
\end{aligned}$$

Therefore,

$$\frac{\partial \boldsymbol{R}_{\boldsymbol{y}}}{\partial v_{ni,i,l}^{(p)}} = \mathrm{Cj}\left(\tilde{\boldsymbol{a}}\left(v_{ni,i,l}^{(p)}\right)\boldsymbol{a}^H\left(v_{ni,i,l}^{(p)}\right)\right) \otimes \boldsymbol{a}\left(u_{ni,i,l}^{(p)}\right)\boldsymbol{a}^H\left(u_{ni,i,l}^{(p)}\right) \otimes \mathbb{E}\left(\tilde{\boldsymbol{X}}_{ni}^T \boldsymbol{R}'_{l,p} \tilde{\boldsymbol{X}}_{ni}^*\right)$$

$$\frac{\partial \boldsymbol{M}_{\boldsymbol{y}}}{\partial v_{ni,i,l}^{(p)}} = \varsigma\mathrm{Tj}\left(\tilde{\boldsymbol{a}}\left(v_{ni,i,l}^{(p)}\right)\boldsymbol{a}^T\left(v_{ni,i,l}^{(p)}\right)\right) \otimes \boldsymbol{a}\left(u_{ni,i,l}^{(p)}\right)\boldsymbol{a}^T\left(u_{ni,i,l}^{(p)}\right) \otimes \mathbb{E}\left(\tilde{\boldsymbol{X}}_{ni}^T \boldsymbol{M}'_{l,p} \tilde{\boldsymbol{X}}_{ni}^*\right)$$

where $Cj(\boldsymbol{A})$ and $Tj(\boldsymbol{A})$ respectively stands for $\boldsymbol{A} + \boldsymbol{A}^H$ and $\boldsymbol{A} + \boldsymbol{A}^T$.

$$\boldsymbol{R}'_{l,p} \triangleq \mathbb{E}_{\boldsymbol{\omega}_{ni,i}}\left[\boldsymbol{A}^H(\boldsymbol{\omega}_{ni,i})\,\boldsymbol{a}\left(\omega_{ni,i,l}^{(p)}\right)\boldsymbol{a}^H\left(\omega_{ni,i,l}^{(p)}\right)\boldsymbol{A}(\boldsymbol{\omega}_{ni,i})\right]$$

$$\boldsymbol{M}'_{l,p} \triangleq \mathbb{E}_{\boldsymbol{\omega}_{ni,i}}\left[\boldsymbol{A}^H(\boldsymbol{\omega}_{ni,i})\boldsymbol{a}\left(\omega_{ni,i,l}^{(p)}\right)\boldsymbol{a}^T\left(\omega_{ni,i,l}^{(p)}\right)\boldsymbol{A}^*(\boldsymbol{\omega}_{ni,i})\right]$$

$$\tilde{\boldsymbol{a}}\left(v_{ni,i,l}^{(p)}\right) \triangleq \mathrm{diag}(0, -j\pi, \ldots, -j(N_a-1)\pi)\boldsymbol{a}\left(v_{ni,i,l}^{(p)}\right)$$

## REFERENCES


[1] E. G. Larsson, O. Edfors, F. Tufvesson, and T. L. Marzetta, "Massive MIMO for next generation wireless systems," *IEEE Commun. Mag.*, vol. 52, no. 2, pp. 186–195, Feb. 2014.

[2] H. Q. Ngo, E. G. Larsson, and T. L. Marzetta, "Energy and spectral efficiency of very large multiuser MIMO systems," *IEEE Trans. Commun.*, vol. 61, no. 4, pp. 1436–1449, Apr. 2013.

[3] Y. Kim *et al.*, "Full dimension MIMO (FD-MIMO): The next evolution of MIMO in LTE systems," *IEEE Wireless Commun.*, vol. 21, no. 2, pp. 26–33, Apr. 2014.

[4] Y.-H. Nam *et al.*, "Full-dimension MIMO (FD-MIMO) for next generation cellular technology," *IEEE Commun. Mag.*, vol. 51, no. 6, pp. 172–179, Jun. 2013.

[5] K. Upadhya, S. A. Vorobyov, and M. Vehkapera, "Superimposed pilots are superior for mitigating pilot contamination in massive MIMO," *IEEE Trans. Signal Process.*, vol. 65, no. 11, pp. 2917–2932, Jun. 2017.

[6] K. Upadhya, S. A. Vorobyov, and M. Vehkaperä, "Downlink performance of superimposed pilots in massive MIMO systems," *IEEE Trans. Wireless Commun.*, vol. 17, no. 10, pp. 6630–6644, Oct. 2018.

[7] D. Verenzuela, E. Björnson, and L. Sanguinetti, "Spectral and energy efficiency of superimposed pilots in uplink massive MIMO," *IEEE Trans. Wireless Commun.*, vol. 17, no. 11, pp. 7099–7115, Nov. 2018.

[8] J. Ma, C. Liang, C. Xu, and L. Ping, "On orthogonal and superimposed pilot schemes in massive MIMO NOMA systems," *IEEE J. Sel. Areas Commun.*, vol. 35, no. 12, pp. 2696–2707, Dec. 2017.

[9] "Study on 3D channel model for LTE," 3GPP, Sophia Antipolis Cedex, France, TR 36.873, Dec. 2017.

[10] P. Stoica and A. Nehorai, "Performance study of conditional and unconditional direction-of-arrival estimation," *IEEE Trans. Acoust., Speech, Signal Process.*, vol. 38, no. 10, pp. 1783–1795, Oct. 1990.

[11] R. Shafin, L. Liu, J. Zhang, and Y.-C. Wu, "DoA estimation and capacity analysis for 3-D millimeter wave massive-MIMO/FD-MIMO OFDM systems," *IEEE Trans. Wireless Commun.*, vol. 15, no. 10, pp. 6963–6978, Oct. 2016.

[12] A. Paulraj, R. Roy, and T. Kailath, "Estimation of signal parameters via rotational invariance techniques-ESPRIT," in *Proc. IEEE 9th Asilomar Conf. Circuits, Syst., Comput.*, 1985, pp. 83–89.

[13] C. Qian, L. Huang, N. D. Sidiropoulos, and H. C. So, "Enhanced PUMA for direction-of-arrival estimation and its performance analysis," *IEEE Trans. Signal Process.*, vol. 64, no. 16, pp. 4127–4137, Aug. 2016.

[14] R. Shafin, L. Liu, Y. Li, A. Wang, and J. Zhang, "Angle and delay estimation for 3-D massive MIMO/FD-MIMO systems based on parametric channel modeling," *IEEE Trans. Wireless Commun.*, vol. 16, no. 8, pp. 5370–5383, Aug. 2017.

[15] L. Cheng, Y.-C. Wu, J. Zhang, and L. Liu, "Subspace identification for DOA estimation in massive/full-dimension MIMO systems: Bad data mitigation and automatic source enumeration," *IEEE Trans. Signal Process.*, vol. 63, no. 22, pp. 5897–5909, Nov. 2015.

[16] D. Zhu, J. Choi, and R. W. Heath, "Two-dimensional AoD and AoA acquisition for wideband millimeter-wave systems with dual-polarized MIMO," *IEEE Trans. Wireless Commun.*, vol. 16, no. 12, pp. 7890–7905, Dec. 2017.

[17] D. Fan *et al.*, "Angle domain channel estimation in hybrid millimeter wave massive MIMO systems," *IEEE Trans. Wireless Commun.*, vol. 17, no. 12, pp. 8165–8179, Dec. 2018.

[18] D. Fan, F. Gao, G. Wang, Z. Zhong, and A. Nallanathan, "Angle domain signal processing-aided channel estimation for indoor 60-GHz TDD/FDD massive MIMO systems," *IEEE J. Sel. Areas Commun.*, vol. 35, no. 9, pp. 1948–1961, Sep. 2017.

[19] C. Qian, X. Fu, N. Sidiropoulos, and Y. Yang, "Tensor-based channel estimation for dual-polarized massive MIMO systems," *IEEE Trans. Signal Process.*, vol. 66, no. 24, pp. 6390–6403, Dec. 2018.

[20] A. P. Dempster, N. M. Laird, and D. B. Rubin, "Maximum likelihood from incomplete data via the EM algorithm," *J. Roy. Statistical Soc., Ser. B (Methodological)*, vol. 39, pp. 1–38, 1977.

[21] T. G. Kolda and B. W. Bader, "Tensor decompositions and applications," *SIAM Rev.*, vol. 51, no. 3, pp. 455–500, 2009.

[22] R. Shafin and L. Liu, "Multi-cell multi-user massive FD-MIMO: Downlink precoding and throughput analysis," *IEEE Trans. Wireless Commun.*, vol. 18, no. 1, pp. 487–502, Jan. 2019.

[23] Z. Zhou, J. Fang, L. Yang, H. Li, Z. Chen, and R. S. Blum, "Low-rank tensor decomposition-aided channel estimation for millimeter wave MIMO-OFDM systems," *IEEE J. Sel. Areas Commun.*, vol. 35, no. 7, pp. 1524–1538, Jul. 2017.

[24] Z. Zhou, J. Fang, L. Yang, H. Li, Z. Chen, and S. Li, "Channel estimation for millimeter-wave multiuser MIMO systems via PARAFAC decomposition," *IEEE Trans. Wireless Commun.*, vol. 15, no. 11, pp. 7501–7516, Nov. 2016.

[25] N. D. Sidiropoulos, L. De Lathauwer, X. Fu, K. Huang, E. E. Papalexakis, and C. Faloutsos, "Tensor decomposition for signal processing and machine learning," *IEEE Trans. Signal Process.*, vol. 65, no. 13, pp. 3551–3582, Jul. 2017.

[26] A. M. Sayeed, "Deconstructing multiantenna fading channels," *IEEE Trans. Signal Process.*, vol. 50, no. 10, pp. 2563–2579, Oct. 2002.

[27] J.-P. Delmas and H. Abeida, "Stochastic Cramer-Rao bound for noncircular signals with application to DOA estimation," *IEEE Trans. Signal Process.*, vol. 52, no. 11, pp. 3192–3199, Nov. 2004.

[28] J. B. Kruskal, "Three-way arrays: Rank and uniqueness of trilinear decompositions, with application to arithmetic complexity and statistics," *Linear Algebra Its Appl.*, vol. 18, no. 2, pp. 95–138, 1977.

[29] A. Stegeman and N. D. Sidiropoulos, "On Kruskal's uniqueness condition for the Candecomp/Parafac decomposition," *Linear Algebra Its Appl.*, vol. 420, no. 2-3, pp. 540–552, 2007.

[30] M. Sørensen and L. De Lathauwer, "Blind signal separation via tensor decomposition with Vandermonde factor: Canonical polyadic decomposition," *IEEE Trans. Signal Process.*, vol. 61, no. 22, pp. 5507–5519, Nov. 2013.

[31] S. Vishwanath, N. Jindal, and A. Goldsmith, "Duality, achievable rates, and sum-rate capacity of Gaussian MIMO broadcast channels," *IEEE Trans. Inf. Theory*, vol. 49, no. 10, pp. 2658–2668, Oct. 2003.

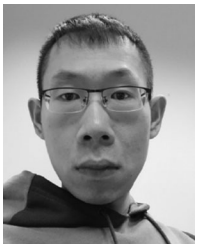
[32] Q. Shi, M. Razaviyayn, Z.-Q. Luo, and C. He, "An iteratively weighted MMSE approach to distributed sum-utility maximization for a MIMO interfering broadcast channel," *IEEE Trans. Signal Process.*, vol. 59, no. 9, pp. 4331–4340, Sep. 2011.

[33] Z.-Q. Luo and P. Tseng, "On the convergence of the coordinate descent method for convex differentiable minimization," *J. Optim. Theory Appl.*, vol. 72, no. 1, pp. 7–35, 1992.

[34] K. Huang, N. D. Sidiropoulos, and A. P. Liavas, "A flexible and efficient algorithmic framework for constrained matrix and tensor factorization," *IEEE Trans. Signal Process.*, vol. 64, no. 19, pp. 5052–5065, Oct. 2016.

[35] H. Inaltekin, "Gaussian approximation for the wireless multi-access interference distribution," *IEEE Trans. Signal Process.*, vol. 60, no. 11, pp. 6114–6120, Nov. 2012.

[36] Z. Koldovsky, P. Tichavsky, and E. Oja, "Efficient variant of algorithm fastICA for independent component analysis attaining the Cramer-Rao lower bound," *IEEE Trans. Neural Netw.*, vol. 17, no. 5, pp. 1265–1277, Sep. 2006.



**Zhou Zhou** received the B.S. degree in communications engineering from the University of Electronic Science and Technology of China, Chengdu, China, in 2011. Since 2018, he has been a Research Assistant with the ECE Department of Virginia Tech., Blacksburg, VA, USA. His current research interests are in the broad area of neural networks, machine intelligence, and wireless communications.

**Lingjia Liu** (SM'15) received the B.S. degree in electronic engineering from Shanghai Jiao Tong University, Shanghai, China and the Ph.D. degree in electrical and computer engineering from Texas A&M University, College Station, TX, USA. He is currently an Associate Professor with the Bradley Department of Electrical and Computer Engineering, Virginia Tech., Blacksburg, VA, USA. He is also the Associate Director for Affiliate Relations with Wireless@Virginia Tech. Prior to that, he was an Associate Professor with the EECS Department, University of Kansas (KU). From 2008 and 2011, he was with the Standards & Mobility Innovation Lab., Samsung Research America, where he received the Global Samsung Best Paper Award in 2008 and 2010. He was leading Samsung's efforts on multiuser MIMO, CoMP, and HetNets in 3GPP LTE/LTE-advanced standards. His general research interests mainly lie in emerging technologies for 5G cellular networks including machine learning for wireless networks, massive MIMO, massive machine-type communications, and mm-wave communications. He received the Air Force Summer Faculty Fellowship from 2013 to 2017, the Miller Scholarship at KU in 2014, the Miller Professional Development Award for Distinguished Research at KU in 2015, the 2016 IEEE GLOBECOM Best Paper Award, the 2018 IEEE ISQED Best Paper Award, the 2018 IEEE TCGCC Best Conference Paper Award, and the 2018 IEEE TAOS Best Paper Award.

**Jianzhong Zhang** received the Ph.D. degree from the University of Wisconsin Madison, Madison, WI, USA. He is a VP and the Head of the Standards and Mobility Innovation Lab of Samsung Research America, where he leads research, prototyping, and standards for 5G and future multimedia networks. From 2009 to 2013, he was the Vice Chairman of the 3GPP RAN1 working group and led development of LTE and LTE-Advanced technologies such as 3-D channel modeling, UL-MIMO, CoMP, Carrier Aggregation for TD-LTE.