Decentralized MIMO Systems With Imperfect CSI Using LMMSE Receivers

Zeyan Zhuang, *Graduate Student Member, IEEE*, Xin Zhang, *Member, IEEE*, Dongfang Xu, *Member, IEEE*, Shenghui Song, *Senior Member, IEEE*, and Yonina C. Eldar, *Fellow, IEEE*

***Abstract*—Centralized baseband processing (CBP) is required to achieve the full potential of massive multiple-input multiple-output (MIMO) systems. However, due to the large number of antennas, CBP suffers from two major issues: 1) Extensive data interconnection between radio frequency (RF) circuitry and the central processing unit; and 2) high-dimensional computation. To this end, decentralized baseband processing (DBP) has been proposed, where the antennas at the base station are partitioned into clusters connected to separate RF circuits and equipped with separate computing units. However, the optimal fusion scheme that maximizes signal-to-interference-and-noise ratio (SINR) and the related performance analysis for DBP with general spatial correlation and imperfect channel state information (CSI) have not been studied. In this paper, we consider a decentralized MIMO system where all clusters adopt linear minimum mean-square error (LMMSE) receivers. We first establish an optimal linear fusion scheme that has high computational and data input/output costs. To reduce the cost, we then propose two suboptimal fusion schemes with reduced complexity. For all three schemes, we study the SINR performance by leveraging random matrix theory and demonstrate conditions under which the suboptimal schemes are optimal. Furthermore, we determine the optimal regularization parameter for the LMMSE receiver, identify the best antenna partitioning strategy, and prove that the SINR will decrease as the number of clusters increases. Numerical simulations validate the accuracy of the theoretical results.**

***Index Terms*—Decentralized baseband processing (DBP), linear minimum mean-square error (LMMSE), massive multiple-input-multiple-output (MIMO), random matrix theory (RMT).**

I. Introduction

MASSIVE multi-input-multi-output (MIMO) technology has been shown very effective in achieving high spectral efficiency, energy efficiency, and link reliability [1]. However, conventional implementations require centralized baseband processing (CBP) where all baseband data are collected and processed by a centralized baseband processing unit (BBU) [2], [3]. Such centralized processing causes significant challenges when the size of the antenna arrays becomes extremely large. On the one hand, transmitting baseband data from a large number of base station (BS) antennas to the BBU results in very high throughput demands. On the other hand, traditional estimation or beamforming algorithms require the inversion of high-dimensional matrices, which causes heightened computational complexity. These challenges make CBP very difficult to implement in practice [4], [5].

To address the bandwidth and computation bottlenecks of CBP, a more efficient architecture called decentralized baseband processing (DBP) was recently proposed [2], [4], [6], [7], [8], [9], [10]. With DBP, BS antennas are divided into several clusters, each equipped with its own processing unit that performs signal processing tasks in a parallel and decentralized manner. Li et al. [4], [6] developed iterative decentralized algorithms to minimize the mean-squared error (MSE) of the received symbols. A gradient descent based decentralized algorithm was proposed in [7] to obtain zero-forcing (ZF) equalization for a daisy chain architecture. In [2], [11], the feedforward DBP architecture was proposed to reduce latency by avoiding the information exchange between clusters required by previous works [12].

With the feedforward structure, the estimates from all clusters and some middle-stage information are transferred to and fused at the central unit (CU). As a result, the fusion algorithm at the CU is crucial for successful estimation. In [2], Jeon et al. provided an optimal linear fusion that maximizes the signal-to-interference-and-noise ratio (SINR) in uncorrelated Rayleigh channels with perfect channel state information (CSI). Closed-form deterministic approximations of the SINR for maximum ratio combining (MRC), ZF, and linear minimum mean-square error (LMMSE) receivers were also derived. The design of the fusion coefficients for decentralized MIMO systems resembles that of the collaboration among access points (APs) in cell-free massive MIMO (CF-mMIMO) systems, where distributed APs are deployed and linked to a central processing unit (CPU) [13]. The difference between DBP and CF-mMIMO lies in the spatial correlation among different clusters, which makes the DBP architecture more general. In [14], four levels of cooperation with minimum mean-square error (MMSE) receiver were

Received 15 August 2024; revised 6 December 2024; accepted 20 January 2025. Date of publication 10 February 2025; date of current version 5 May 2025. This work was supported in part by the NSFC/RGC Joint Research Scheme through the Research Grants Council of the Hong Kong Special Administrative Region, China and National Natural Science Foundation of China under Project N_HKUST656/22 and in part by General Research Fund through Hong Kong Research Grants Council under Project 16209524. The guest editor coordinating the review of this article and approving it for publication was Prof. Xiao Li. *(Corresponding author: Shenghui Song.)*

Zeyan Zhuang, Xin Zhang, and Shenghui Song are with the Department of Electronic and Computer Engineering, The Hong Kong University of Science and Technology, Hong Kong, Hong Kong (e-mail: zzhuangac@connect.ust.hk; eezhangxin@ust.hk; eeshsong@ust.hk).

Dongfang Xu is with the Division of Integrative Systems and Design, The Hong Kong University of Science and Technology, Hong Kong, Hong Kong (e-mail: eedxu@ust.hk).

Yonina C. Eldar is with the Department of Mathematics and Computer Science, The Weizmann Institute of Science, Rehovot 7610001, Israel (e-mail: yonina.eldar@weizmann.ac.il).

Digital Object Identifier 10.1109/JSTSP.2025.3539098

proposed for CF-mMIMO systems, considering imperfect CSI and spatially correlated fading. However, the fusion scheme and performance analysis for DBP with imperfect CSI and general spatial correlation have not been studied.

In this paper, we consider feedforward decentralized massive MIMO systems with LMMSE receivers due to their near-optimal performance and ease of implementation [12]. We also consider the case with imperfect CSI, where the channel estimation (CE) errors are assumed to follow Gaussian distributions [14], [15], due to the difficulty in obtaining perfect CSI in massive MIMO systems [16], [17]. Specifically, we first determine the optimal linear fusion that maximizes the SINR, which requires global CSI and has high complexity. To reduce the cost, we propose two linear fusion schemes with reduced computation and communication workloads. In particular, the first approach utilizes a moderate amount of intermediate results computed by the clusters, while the second method can be implemented by a fully decentralized (FD) architecture [2], where the CU only needs to calculate the weighted sum of the local estimates, without requiring any other intermediate results.

Unfortunately, the presence of CE errors and the decentralized structure make it very challenging to evaluate the receive SINR. To this end, we adopt large random matrix theory (RMT) to tackle the random vector channels, which has been proven effective in analyzing the fundamental limits of centralized LMMSE receivers [18], [19]. The SINR analysis for the DBP architecture has been considered in [2] utilizing RMT. However, they assumed perfect CSI and uncorrelated Rayleigh channel, while here we consider the more general case of imperfect CSI and general spatial correlation. Based on the analysis results, we further investigate the impacts of several key system parameters, including the regularization parameter, the antenna partitioning strategy, and the number of clusters.

The main contributions can be summarized as follows:

1) For decentralized massive MIMO systems, we first derive the optimal linear fusion scheme to maximize the SINR with imperfect CSI and demonstrate that the optimal coefficients for the fusion scheme also minimize the MSE for estimating the transmit signal. We then propose two suboptimal fusion schemes assuming local CSI, which have reduced levels of complexity.
2) We derive the deterministic approximations of the SINR for all three fusion schemes using RMT. The analysis results reveal that, without spatial correlation between clusters or CE errors, the first suboptimal fusion scheme achieves the optimal SINR. The second suboptimal fusion scheme, which has lower complexity, can achieve the optimal SINR when there is no spatial correlation between clusters.
3) Based on the theoretical analysis, we investigate the scenario with heterogeneous CSI inaccuracy among clusters. The results show that when the CSI error is high for certain clusters, the system's performance is asymptotically equal to that without considering those clusters.
4) To obtain more physical insights, we derive closed-form expressions for the SINR over independent and identically distributed (i.i.d.) channels. Based on the results, we further determine the optimal regularization parameters and optimal antenna partition strategies that maximize SINR, and evaluate the impact of the number of clusters.
5) From the RMT perspective, we derive the deterministic approximations for linear functions of the resolvent where the covariance matrices have generally correlated columns, and demonstrate the convergence rates of the means are $\mathcal{O}(N^{-\frac{1}{2}})$. These findings can be applied to the analysis of regularized zero-forcing (RZF) precoding in downlink MIMO systems [20].

The rest of the paper is organized as follows. In Section II, we introduce the system model and the DBP architecture. In Section III, we derive the optimal linear fusion scheme that maximizes the SINR and propose two linear fusion schemes for decentralized implementation. In Section IV, we derive deterministic approximations for the SINR of proposed fusion schemes and investigate the results with i.i.d. channels. Numerical results are given in Sections V and VI concludes the paper.

Notations: Throughout the paper, lowercase and uppercase boldface letters represent vectors and matrices, respectively. We use $\mathbb{C}^N$ and $\mathbb{C}^{N\times M}$ to denote the space of N-dimensional complex vectors and the space of N-by-M complex matrices, respectively. The conjugate transpose and transpose operator are denoted by $(\cdot)^H$ and $(\cdot)^T$, respectively, $[\mathbf{A}]_{i,j}$ and $[\mathbf{a}]_i$ represent the (i,j)-th entry of matrix $\mathbf{A}$ and i-th element of $\mathbf{a}$, respectively. For integer vectors $\boldsymbol{\tau}=[\tau_1,\ldots,\tau_{N_1}]^T$ and $\boldsymbol{\beta}=[\beta_1,\ldots,\beta_{N_2}]^T$, we use notations $\mathbf{A}[\boldsymbol{\tau},\boldsymbol{\beta}]$ and $\mathbf{A}[\boldsymbol{\tau},:]$ to denote the submatrices of $\mathbf{A}$ with $[\mathbf{A}[\boldsymbol{\tau},\boldsymbol{\beta}]]_{i,j}=[\mathbf{A}]_{\tau_i,\beta_j}$ and $[\mathbf{A}[\boldsymbol{\tau},:]]_{i,j}=[\mathbf{A}]_{\tau_i,j}$, respectively. We write $\|\cdot\|$ for the spectral norm of a matrix or the Euclidean norm of a vector, Tr$\mathbf{A}$ refers to the trace of $\mathbf{A}$, $\mathbf{I}_N$ represents the identity matrix of size N, and $\mathbf{1}_N$ denotes the vector with all 1 entries of size N. The probability measure and the expectation operator are denoted by $\mathbb{P}[\cdot]$ and $\mathbb{E}[\cdot]$, respectively, $\underline{x}=x-\mathbb{E}x$ denotes the centered form of random variable x, and $\xrightarrow{a.s.}$ indicates almost sure convergence. We use $\frac{\partial(\cdot)}{\partial x}$ to represent the partial derivative. The notations $[N]$ and $[N]_0$ represent the set $\{1,2,\ldots,N\}$ and $\{0\}\cup[N]$, respectively. The indicator function is represented by $\mathbb{I}_{\{\cdot\}}$, and $\mathcal{O}(\cdot)$ is the standard Big-O notation. Specifically, we have $f(N)=\mathcal{O}(g(N))$ if and only if there exists a positive real number C and positive integer N_0 such that $|f(N)|\le Cg(N)$ for all $N\ge N_0$.

II. System Model

A. Signal Model

Consider an uplink MIMO system where $M+1$ single-antenna users transmit signals to one BS with N antennas. We denote the transmit signal of all users as $\mathbf{x}=[x_0,\ldots,x_M]^T\sim\mathcal{CN}(0,\mathbf{I}_{M+1})$, where x_j represents the signal of user j. The channel between user j and the BS is denoted by $\mathbf{h}_j\in\mathbb{C}^N$. Thus, the received signal at the BS is given by

$$\mathbf{y}=\boldsymbol{\Sigma}\mathbf{x}+\mathbf{n}, \tag{1}$$

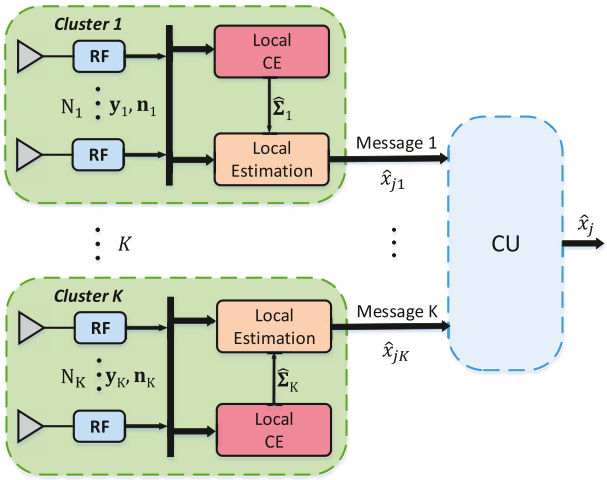


Fig. 1. The feedforward DBP architecture for massive MIMO uplink.

where $\mathbf{y} \in \mathbb{C}^N$, $\mathbf{\Sigma} = [\mathbf{h}_0, \ldots, \mathbf{h}_M] \in \mathbb{C}^{N\times(M+1)}$, and $\mathbf{n} \sim \mathcal{CN}(0, \sigma^2\mathbf{I}_N)$ denotes the additive Gaussian white noise (AWGN) at the BS.

Decentralized Architecture: In this paper, we consider the decentralized architecture [2], [4] as shown in Fig. 1, where the N antennas of the BS are divided into K clusters, connected to one CU. The k-th cluster consists of N_k antennas, and each cluster contains its own RF components and computing unit. For simplicity, we partition the received signal $\mathbf{y} = [\mathbf{y}_1^T, \ldots, \mathbf{y}_K^T]^T$, the channel matrix $\mathbf{\Sigma} = [\mathbf{\Sigma}_1^T, \ldots, \mathbf{\Sigma}_K^T]^T$, the channel vector $\mathbf{h}_j = [\mathbf{h}_{j1}^T, \ldots, \mathbf{h}_{jK}^T]^T, j \in [M]_0$, and the noise vector $\mathbf{n} = [\mathbf{n}_1^T, \ldots, \mathbf{n}_K^T]^T$, such that the received signal at the k-th cluster is given by

$$\mathbf{y}_k = \mathbf{\Sigma}_k\mathbf{x} + \mathbf{n}_k,\ k \in [K], \tag{2}$$

with $\mathbf{y}_k \in \mathbb{C}^{N_k}$, $\mathbf{\Sigma}_k \in \mathbb{C}^{N_k\times(M+1)}$, and $\mathbf{n}_k \in \mathbb{C}^{N_k}$. To estimate the signal from user j, the k-th cluster independently processes the baseband signal to compute the estimate $\widehat{x}_{jk}$. The local estimate $\widehat{x}_{jk}$ and a moderate amount of intermediate results are then transferred to the CU. These intermediate results are designed based on the estimation scheme and typically include the Gram matrices of the channels [21], [22] or variances of the estimation errors [2]. The CU will obtain the final estimate $\widehat{x}_j$ through linear fusion as

$$\widehat{x}_j = \sum_{k=1}^{K} \alpha_k \widehat{x}_{jk}, \tag{3}$$

where α_k represents the fusion coefficient that can be chosen as constant or determined based on the information from clusters. To estimate x_j, LMMSE estimation is widely used in conventional CBP due to its simplicity in implementation and near-optimal performance [12]. Therefore, we consider the case where all clusters adopt LMMSE estimator and the detailed design for the fusion coefficients $\alpha_k, k \in [K]$ will be discussed in Section III.

B. Channel Model

Due to limited angular spread and insufficient antenna spacing, spatial correlation between BS antennas is inevitable. In this paper, we consider the Rayleigh channel model [15], [23]

$$\mathbf{h}_j = \mathbf{R}_j^{\frac{1}{2}}\mathbf{z}_j,\ j \in [M]_0, \tag{4}$$

where $\mathbf{R}_j \in \mathbb{C}^{N\times N}$ is a nonnegative definite matrix that models the spatial correlation of the BS antennas with respect to (w.r.t.) user j and $\mathbf{z}_j \sim \mathcal{CN}(0, \mathbf{I}_N)$. Note that this model takes the variance profile model [18], [24] as a special case. With DBP, each cluster can only know its local spatial correlation matrix, which is a principal sub matrix of $\mathbf{R}_j$. To this end, we denote the index vector

$$\boldsymbol{\varsigma}_k = \left[\sum_{i=1}^{k-1} N_i + 1, \sum_{i=1}^{k-1} N_i + 2, \ldots, \sum_{i=1}^{k} N_i\right]^T,\ k \in [K], \tag{5}$$

and for a N-by-N matrix $\mathbf{R}$, we define

$$[\mathbf{R}]_{[k,l]} = \mathbf{R}[\boldsymbol{\varsigma}_k, \boldsymbol{\varsigma}_l]. \tag{6}$$

As a result, the spatial correlation matrix $\mathbf{R}_j$ can be expressed in the following block matrix form

$$\mathbf{R}_j = \begin{bmatrix} [\mathbf{R}_j]_{[1,1]} & [\mathbf{R}_j]_{[1,2]} & \cdots & [\mathbf{R}_j]_{[1,K]} \\ [\mathbf{R}_j]_{[2,1]} & [\mathbf{R}_j]_{[2,2]} & \cdots & [\mathbf{R}_j]_{[2,K]} \\ \vdots & \vdots & \ddots & \vdots \\ [\mathbf{R}_j]_{[K,1]} & [\mathbf{R}_j]_{[K,2]} & \cdots & [\mathbf{R}_j]_{[K,K]} \end{bmatrix}, \tag{7}$$

where the off-diagonal block $[\mathbf{R}_j]_{[k,l]}$ denotes the spatial correlation between the k-th and the l-th clusters for $k \neq l$, and the main-diagonal block $[\mathbf{R}_j]_{[k,k]}$ represents the spatial correlation within the k-th cluster.

Decentralized Channel Estimation: For simplicity, we assume $M+1$ mutually orthogonal pilot signals $\boldsymbol{\phi}_j \in \mathbb{C}^\tau, j \in [M]_0$ with $\|\boldsymbol{\phi}_j\|^2 = \tau$ ($\tau \geq M+1$). During the training phase, the received signal $\widetilde{\mathbf{Y}}_k \in \mathbb{C}^{N_k\times\tau}$ at the k-th cluster is given by

$$\widetilde{\mathbf{Y}}_k = \sum_{i=0}^{M} \mathbf{h}_{ik}\boldsymbol{\phi}_i^T + \widetilde{\mathbf{N}}_k, \tag{8}$$

where $\widetilde{\mathbf{N}}_k \in \mathbb{C}^{N_k\times\tau}$ represents the noise matrix with i.i.d. entries, each following $\mathcal{CN}(0, \sigma_{\mathrm{tr},k}^2)$. To estimate $\mathbf{h}_{jk}$, the k-th cluster will first correlate the training signal with the scaled pilot $\widetilde{\mathbf{y}}_{jk} = \widetilde{\mathbf{Y}}_k\frac{\boldsymbol{\phi}_j^*}{\tau}$, which yields

$$\widetilde{\mathbf{y}}_{jk} = \left(\sum_{i=0}^{M} \mathbf{h}_{ik}\boldsymbol{\phi}_i^T + \widetilde{\mathbf{N}}_k\right)\frac{\boldsymbol{\phi}_j^*}{\tau} = \mathbf{h}_{jk} + \widetilde{\mathbf{n}}_{jk}, \tag{9}$$

where $\widetilde{\mathbf{n}}_{jk} \sim \mathcal{CN}(0, \frac{\sigma_{\mathrm{tr},k}^2}{\tau})$. Denote the power of the resulting noise as $\widetilde{\sigma}_k^2 = \frac{\sigma_{\mathrm{tr},k}^2}{\tau}$. Thus, the training signal-to-noise ratio (SNR) is $\frac{1}{\widetilde{\sigma}_k^2}$, which is assumed to be known at each cluster. Under such circumstances, the MMSE estimate $\widehat{\mathbf{h}}_{jk}$ of $\mathbf{h}_{jk}$ is given by [25, Theorem 3.1]

$$\widehat{\mathbf{h}}_{jk} = [\mathbf{R}_j]_{[k,k]}\left([\mathbf{R}_j]_{[k,k]} + \widetilde{\sigma}_k^2\mathbf{I}_{N_k}\right)^{-1}\widetilde{\mathbf{y}}_{jk}. \tag{10}$$

Stacking the channel estimate of all clusters as $\widehat{\mathbf{h}}_j = [\widehat{\mathbf{h}}_{j1}^T, \ldots, \widehat{\mathbf{h}}_{jK}^T]^T$, the distribution of the channel vector can be expressed as

$$\begin{aligned} \widehat{\mathbf{h}}_j &\sim \mathcal{CN}(\mathbf{0}, \mathbf{\Phi}_j), \\ \mathbf{h}_j|\widehat{\mathbf{h}}_j &\sim \mathcal{CN}(\widetilde{\mathbf{h}}_j, \mathbf{W}_j), \end{aligned} \tag{11}$$

where the distribution of $\widehat{\mathbf{h}}_j$ is derived from the zero mean channel (4) and zero mean training noise (8), using the properties of the linear transformations of Gaussian distributions [26, Section III]. Here, $\mathbf{W}_j$ represents the covariance matrix of the CE error and the posterior mean is given by $\widetilde{\mathbf{h}}_j = \mathbf{V}_j\widehat{\mathbf{h}}_j$, by the properties of conditional Gaussian distributions [26, Section IV]. Matrices $\mathbf{\Phi}_j$, $\mathbf{V}_j$, and $\mathbf{W}_j$ are defined as

$$\begin{aligned} \mathbf{\Phi}_j &= \mathbf{D}_{T,j}(\mathbf{D}_{\widetilde{\sigma}} + \mathbf{R}_j)\mathbf{D}_{T,j}, \\ \mathbf{V}_j &= \mathbf{T}_j\mathbf{D}_{T,j}^{-1}, \ \mathbf{W}_j = \mathbf{T}_j\mathbf{D}_{\widetilde{\sigma}}, \end{aligned} \tag{12}$$

where $\mathbf{D}_{\widetilde{\sigma}} = \text{diag}(\widetilde{\sigma}_k^2\mathbf{I}_{N_k}; k \in [K])$, $\mathbf{T}_j = \mathbf{R}_j(\mathbf{D}_{\widetilde{\sigma}} + \mathbf{R}_j)^{-1}$, $\mathbf{D}_{R,j} = \text{diag}([\mathbf{R}_j]_{[k,k]}; k \in [K])$, and $\mathbf{D}_{T,j} = \mathbf{D}_{R,j}(\mathbf{D}_{\widetilde{\sigma}} + \mathbf{D}_{R,j})^{-1}$. When the spatial correlation between antennas of different clusters is negligible, i.e., the off-diagonal block $[\mathbf{R}_j]_{[k,l]} = \mathbf{0}_{N_k \times N_l}$ for $k \neq l$, it can be verified that $\mathbf{R}_j = \mathbf{D}_{R,j}$, $\mathbf{V}_j = \mathbf{I}_N$, and $\widehat{\mathbf{h}}_j$ is the same as that estimated by the centralized MMSE estimator. Therefore, $\mathbf{V}_j$ indicates the effect of decentralized CE. For ease of notation, we define the channel matrices $\widehat{\mathbf{\Sigma}} = [\widehat{\mathbf{h}}_0, \ldots, \widehat{\mathbf{h}}_M]$, $\widehat{\mathbf{\Sigma}}_k = [\widehat{\mathbf{h}}_{0k}, \ldots, \widehat{\mathbf{h}}_{Mk}]$, $\widetilde{\mathbf{\Sigma}} = [\widetilde{\mathbf{h}}_0, \ldots, \widetilde{\mathbf{h}}_M]$, and the sum of the covariance $\mathbf{W} = \sum_{j\in[M]_0} \mathbf{W}_j$.

C. SINR

Without loss of generality, we consider the estiomation of x_0 using linear reveiver at each cluster. Specifically, the k-th cluster will estimate $\widehat{x}_{0k} = \mathbf{r}_k^H\mathbf{y}_k$, where $\mathbf{r}_k = \mathbf{r}_k(\widehat{\mathbf{\Sigma}}_k, \{[\mathbf{R}_j]_{[k,k]}\}_{j=0}^M, \widetilde{\sigma}_k^2)$ depends on the local CSI of the k-th cluster. With the linear fusion in (3), the achievable rate of user 0 can be bounded by a standard lower bound $R_0 = \log(1+\gamma_0)$ based on the worst-case uncorrelated additive noise [27]. Here, γ_0 is the associated SINR given by

$$\gamma_0 = \frac{\left|\mathbb{E}_{\mathbf{h}_0|\widehat{\mathbf{h}}_0}\left[\mathbf{r}_\alpha^H\mathbf{h}_0\right]\right|^2}{\mathbb{E}_{\mathbf{\Sigma}|\widehat{\mathbf{\Sigma}},\mathbf{x},\mathbf{n}}\left|\mathbf{r}_\alpha^H\left[\mathbf{\Sigma}\mathbf{x} - \widetilde{\mathbf{h}}_0 x_0 + \mathbf{n}\right]\right|^2}, \tag{13}$$

with $\mathbf{r}_\alpha = [\alpha_1^* \cdot (\mathbf{r}_1)^T, \ldots, \alpha_K^* \cdot (\mathbf{r}_K)^T]^T \in \mathbb{C}^N$. In this paper, we will determine the optimal fusion coefficients $\alpha_1, \ldots, \alpha_K$ that can maximize the SINR γ_0 in (13) [2] and investigate the asymptotic characterization of γ_0.

III. Decentralized LMMSE Receiver and Linear Fusion

In this section, we first introduce the decentralized LMMSE receiver, and derive the optimal linear fusion scheme which requires global CSI and has high complexity. Then, we propose two linear fusion schemes, which only require local CSI and have reduced complexity.

A. Decentralized LMMSE Receiver

With perfect CSI, the local MSE of the estimate $\widehat{x}_{0k} = \mathbf{r}_k^H\mathbf{y}_k$ is given by

$$\text{MSE}_k = \mathbb{E}_{\mathbf{x},\mathbf{n}}|\widehat{x}_{0k} - x_0|^2, \tag{14}$$

where the expectation is w.r.t. the transmit signal and the AWGN. However, to obtain the linear receiver that minimizes the local MSE with imperfect CSI, the expectation should be performed over $\mathbf{\Sigma}_k|\widehat{\mathbf{\Sigma}}_k$ [1] [28, Eq. (6)], which yields

$$\text{MSE}_k|_{\widehat{\mathbf{\Sigma}}_k} = \mathbb{E}_{\mathbf{\Sigma}_k|\widehat{\mathbf{\Sigma}}_k,\mathbf{x},\mathbf{n}}|\widehat{x}_{0k} - x_0|^2. \tag{15}$$

Hence, the optimal $\mathbf{r}_k$ that minimizes $\text{MSE}_k|_{\widehat{\mathbf{\Sigma}}_k}$ is given by [28, Proposition 1]

$$\mathbf{r}_k^{\text{mmse}} = \left(\widehat{\mathbf{\Sigma}}_k\widehat{\mathbf{\Sigma}}_k^H + N_k\mathbf{Z}_k + N_k\rho_k\mathbf{I}_{N_k}\right)^{-1}\widehat{\mathbf{h}}_{0k}, \tag{16}$$

where $\rho_k = \frac{\sigma^2}{N_k}$ and

$$\mathbf{Z}_k = \frac{\widetilde{\sigma}_k^2}{N_k}\sum_{j=0}^M [\mathbf{R}_j]_{[k,k]}\left([\mathbf{R}_j]_{[k,k]} + \widetilde{\sigma}_k^2\mathbf{I}_{N_k}\right)^{-1}. \tag{17}$$

Note that the above ρ_k and $\mathbf{Z}_k$ minimize the local MSE, but may not be optimal after fusion. As a result, we treat ρ_k and $\mathbf{Z}_k$ as design parameters that can be optimized and refer to ρ_k as the regularization parameter in the subsequent.

B. Linear Fusion With Optimal Coefficients

The following proposition provides the optimal linear fusion scheme that maximizes the SINR and demonstrates that it also minimizes the MSE for estimating the transmit signal.

Proposition 1: The optimal coefficients $\boldsymbol{\alpha} = [\alpha_1, \ldots, \alpha_K]$ for maximizing (13) are given by

$$\boldsymbol{\alpha}^{\text{opt}} = c\widetilde{\mathbf{h}}_0^H\mathbf{D}_r\left(\mathbf{D}_r^H\left(\widetilde{\mathbf{\Sigma}}\widetilde{\mathbf{\Sigma}}^H + \mathbf{W} + \sigma^2\mathbf{I}_N\right)\mathbf{D}_r\right)^{-1}, \tag{18}$$

where $c \in \mathbb{C}\backslash\{0\}$ and $\mathbf{D}_r = \text{diag}(\mathbf{r}_k^{\text{mmse}}; k \in [K])$. When $c = 1$, $\boldsymbol{\alpha}^{\text{opt}}$ is also the optimal solution for the following MMSE problem

$$\underset{\boldsymbol{\alpha}}{\text{minimize}}\ \text{MSE}|_{\widehat{\mathbf{\Sigma}}} = \mathbb{E}_{\mathbf{\Sigma}|\widehat{\mathbf{\Sigma}},\mathbf{x},\mathbf{n}}\left|\sum_{k=1}^K \alpha_k\widehat{x}_{0k} - x_0\right|^2. \tag{19}$$

Proof: The proof of Proposition 1 is given in Appendix B. □

Remark 1: The optimality of $\boldsymbol{\alpha}^{\text{opt}}$ in Proposition 1 holds for any linear receivers, e.g., MRC, and LMMSE is a special case.

Remark 2: It can be verified that when $\boldsymbol{\alpha} = \boldsymbol{\alpha}^{\text{opt}}$ and $c = 1$, we have $\text{MSE}|_{\widehat{\mathbf{\Sigma}}} = (1+\gamma_0)^{-1} = \frac{\partial\log(1+\gamma_0)}{\partial\gamma_0}$, which reveals the connection between mutual information and MSE. This relation can be generalized to the case with non-Gaussian signals [29].

Remark 3: Proposition 1 shows that the fusion coefficients that maximize the SINR are actually a scaled version of the parameters that minimize MSE.

[1] According to the properties of conditional Gaussian distributions [26, Section IV], the distribution of the channel vector is given by $\mathbf{h}_{jk}|\widehat{\mathbf{h}}_{jk} \sim \mathcal{CN}(\widehat{\mathbf{h}}_{jk}, \widetilde{\sigma}_k^2[\mathbf{R}_j]_{[k,k]}([\mathbf{R}_j]_{[k,k]} + \widetilde{\sigma}_k^2\mathbf{I}_{N_k})^{-1})$.

TABLE I
COMPARISON FOR THE COMPLEXITY OF DIFFERENT ESTIMATION SCHEMES

Scheme	I/O Throughput of Cluster k	Computational Complexity of Cluster k	Computational Complexity of the CU
Centralized LMMSE	-	-	$C_{\mathrm{L}}(N,M)$
Local LMMSE & LFOC	N_kM+2N_k+1	$C_{\mathrm{L}}(N_k,M)$	$N^2M+2N^2+NM+2N+K^2M+C_{\mathrm{F}}(K)$
Local LMMSE & LFSC	$M+4$	$C_{\mathrm{L}}(N_k,M)+N_kM+2N_k^2+2N_k$	$K^2M+C_{\mathrm{F}}(K)$
Local LMMSE & LFCC	1	$C_{\mathrm{L}}(N_k,M)$	K

We will refer to the fusion scheme in Proposition 1 as linear fusion with optimal coefficients (LFOC) in the following. Note that, due to the presence of spatial correlation between different clusters, one cluster cannot determine the off-diagonal elements of $\mathbf{R}_j-\mathbf{D}_{R,j}$ and the channel $\widetilde{\boldsymbol{\Sigma}}$ in the feedforward architecture. As a result, to obtain $\boldsymbol{\alpha}^{\text{opt}}$, the CU needs to know the global CSI, which leads to high input/output (I/O) and computational cost. To reduce the complexity, we proposed two suboptimal linear fusion schemes in the following.

C. Linear Fusion With Suboptimal Coefficients

Based on (18), we can design the suboptimal fusion parameter as

$$\boldsymbol{\alpha}^{\text{s-opt}}=\widehat{\mathbf{h}}_0^H\mathbf{D}_r\left(\mathbf{D}_r^H(\widehat{\boldsymbol{\Sigma}}\widehat{\boldsymbol{\Sigma}}^H+\mathbf{D}_W+\sigma^2\mathbf{I}_N)\mathbf{D}_r\right)^{-1},\quad(20)$$

where $\mathbf{D}_W=\sum_{j\in[M]_0}\mathbf{D}_{T,j}\mathbf{D}_{\widetilde{\sigma}}$. In particular, the suboptimal solution first replaces $\widetilde{\boldsymbol{\Sigma}}$ in (18) with the estimated channel $\widehat{\boldsymbol{\Sigma}}$, and then adopts a block diagonal version of $\mathbf{W}$. These terms are all computed based on the local CSI and spatial correlations. In fact, we can write

$$[\widehat{\mathbf{M}}]_{k,l}=(\mathbf{r}_k^{\text{mmse}})^H\widehat{\boldsymbol{\Sigma}}_k\widehat{\boldsymbol{\Sigma}}_l^H\mathbf{r}_l^{\text{mmse}}+\mathbb{I}_{\{k=l\}}(\mathbf{r}_k^{\text{mmse}})^H[\mathbf{D}_W+\sigma^2\mathbf{I}_N]_{[k,k]}\mathbf{r}_k^{\text{mmse}},\quad(21)$$

$$[\widehat{\mathbf{m}}]_k=(\mathbf{r}_k^{\text{mmse}})^H\widehat{\mathbf{h}}_{0k},\quad(22)$$

where $\widehat{\mathbf{M}}=\mathbf{D}_r^H(\widehat{\boldsymbol{\Sigma}}\widehat{\boldsymbol{\Sigma}}^H+\mathbf{D}_W+\sigma^2\mathbf{I}_N)\mathbf{D}_r$ and $\widehat{\mathbf{m}}=\mathbf{D}_r^H\widehat{\mathbf{h}}_0$. Therefore, to obtain $\boldsymbol{\alpha}^{\text{s-opt}}$, the k-th cluster needs to compute and transfer the intermediate parameter set

$$\mathcal{P}_k=\left\{(\mathbf{r}_k^{\text{mmse}})^H\widehat{\mathbf{h}}_{0k},\ (\mathbf{r}_k^{\text{mmse}})^H\widehat{\boldsymbol{\Sigma}}_k,\ (\mathbf{r}_k^{\text{mmse}})^H[\mathbf{D}_W+\sigma^2\mathbf{I}_N]_{[k,k]}\mathbf{r}_k^{\text{mmse}}\right\}\quad(23)$$

to the CU. Then the CU will use $\mathcal{P}_1,\dots\mathcal{P}_K$ to determine $\widehat{\mathbf{M}}$ and $\widehat{\mathbf{m}}$ by (21) and (22), respectively, and compute $\boldsymbol{\alpha}^{\text{s-opt}}=\widehat{\mathbf{m}}^H\widehat{\mathbf{M}}^{-1}$. Finally, based on (3), $\widehat{x}_0$ could be obtained. We will refer to this scheme as linear fusion with suboptimal coefficients (LFSC) in the following. Note that the computation and I/O complexity of LFSC is manageable since each parameter in $\mathcal{P}_k$ is a vector or a scalar, and the detailed computational complexity is given in Table I. It can be observed that the overall complexity of LFSC is significantly reduced compared to CBP and LFOC.

D. Linear Fusion With Constant Coefficients

Although LFSC has reduced complexity, it still requires to compute and transmit additional information to the CU. When the number of clusters is very large, the computational complexity, memory requirements, and I/O cost may be very high. To further reduce the complexity, we provide a simple linear fusion scheme with constant coefficients α_k. For example, we can take

$$\alpha_k=\frac{1}{K},\ k\in[K],\quad(24)$$

or more heuristically

$$\alpha_k=\frac{N_k}{N},\ k\in[K].\quad(25)$$

We will refer to this scheme as linear fusion with constant coefficients (LFCC) in the following. The complexity of LFCC is also given in Table I which is much lower than that of LFSC. Note that LFCC can be implemented by the FD feedforward architecture [2], where each cluster only needs to take the weighted sum of local estimates and does not require any additional intermediate results or processing.

E. Complexity Analysis

The complexities for the proposed schemes and centralized LMMSE, including I/O and computations are summarized in Table I. We assume that under LFOC, the CU has access to global statistical CSI.

1) Computational Complexity: The computational complexity primarily arises from multiplication operations. For decentralized estimation, the k-th cluster calculates the channel covariance $\widehat{\boldsymbol{\Sigma}}_k\widehat{\boldsymbol{\Sigma}}_k^H$, which requires $N_k^2(M+1)$ multiplications. The matrix inversion and the equalization vector $\mathbf{r}_k^{\text{mmse}}$ require N_k^3 and N_k^2 multiplications, respectively. Thus, the total computation complexity for the LMMSE estimate $\widehat{x}_{0k}$ is

$$C_{\mathrm{L}}(N_k,M)=N_k^2M+N_k^3+2N_k^2+N_k.\quad(26)$$

For LFOC and LFCC, the k-th cluster only needs to compute the local estimate which requires $C_{\mathrm{L}}(N_k,M)$ multiplications. The computation of the parameter set $\mathcal{P}_k$ introduces extra $N_kM+2N_k^2+2N_k$ multiplications for LFSC. With LFOC, the CU computes the channel's posterior mean $\widetilde{\boldsymbol{\Sigma}}$ with $N^2(M+1)$ multiplications and the fusion coefficients in (18) with $N^2+NM+2\,N+K^2\,M+K^3+2\,K^2$ multiplications. The fusion introduces K multiplications, so the final computational complexity of the CU is given by $N^2\,M+2\,N^2+NM+2\,N+K^2\,M+C_{\mathrm{F}}(K)$ where $C_{\mathrm{F}}(K)=K^3+2\,K^2+K$. For LFSC, the CU needs $K^2\,M+K^3+2\,K^2$ multiplications to obtain $\boldsymbol{\alpha}^{\text{s-opt}}$ based on $\{\mathcal{P}_k\}_{k=1}^K$. Therefore, the overall complexity of obtaining the final estimate for the CU is $K^2\,M+C_{\mathrm{F}}(K)$. For LFCC, K multiplications are used to fuse the local estimates.

2) I/O Throughput: For LFOC, the k-th cluster needs to transmit the N_k-by-$(M+1)$ matrix $\widehat{\boldsymbol{\Sigma}}_k$, the length-$N_k$ vector

$\mathbf{r}_k^{\text{mmse}}$, and the local estimate $\widehat{x}_{0k}$. LFSC requires $M+3$ complex scalars in $\mathcal{P}_k$, while LFCC only transmits the local estimate.

We note that in practice, K is much smaller than N and M, such that the complexity $C_{\text{F}}(K)$ can be neglected. It can be observed that LFOC has high complexity at the CU. Furthermore, the CU requires global statistical CSI $\{\mathbf{R}_j\}_{j=0}^M$ for LFOC, resulting in a significant memory overhead. In contrast, LFSC reduces computational complexity at the CU. Overall, the complexity of LFCC is much lower than that of both LFSC and LFOC.

IV. Asymptotic SINR Analysis

The expression for the SINR γ_0 in (59) contains the correlated terms between the channels $\widehat{\boldsymbol{\Sigma}}$ and $\widetilde{\boldsymbol{\Sigma}}$, making it very difficult to analyze. In this section, we will investigate the asymptotic behaviors of γ_0. Specifically, we will show that as the number of antennas and users tend to infinity at the same pace, γ_0 will converge almost surely. Then, we will derive the deterministic approximations for the SINR of the fusion schemes proposed in Section III. To achieve this, we require the following large-scale system assumptions.

Assumption A.1: For a given number of clusters K, we assume $0 < \liminf_N \frac{N_k}{M} \leq \limsup_N \frac{N_k}{M} < +\infty$ for each $k \in [K]$.

Assumption A.2: Assume

$$\limsup_N \sup_{j\in[M]_0} \lambda_{\text{M}}(\mathbf{R}_j) < +\infty, \tag{27a}$$

$$\liminf_N \inf_{j\in[M]_0} \lambda_{\text{m}}(\mathbf{R}_j) > 0, \tag{27b}$$

where $\lambda_{\text{M}}(\mathbf{R}_j)$ and $\lambda_{\text{m}}(\mathbf{R}_j)$ denote the largest and smallest eigenvalue of $\mathbf{R}_j$, respectively.

Assumption A.3: Assume $\rho_k > 0$ and $\mathbf{Z}_k$ is Hermitian nonnegative such that $\limsup_N \sup_{k\in[K]} \|\mathbf{Z}_k\| < +\infty$ for each $k \in [K]$.

A.1 implies that the number of antennas of the k-th cluster N_k and the number of users M approach infinity at the same pace [2]. For simplicity, we define $c_k = \frac{N_k}{M}$ and use $N \xrightarrow{c_1,\ldots,c_K} +\infty$ to indicate this asymptotic regime. According to interlacing inequalities [30], **A.2** ensures the eigenvalues of the local spatial correlation matrices $[\mathbf{R}_j]_{[k,k]}$ are bounded away from 0, which makes $\mathbf{V}_j$ legal. **A.3** implies that the spectral norm of $\mathbf{Z}_k$ does not diverge to infinity. If **A.1** and **A.2** hold, (17) satisfies this assumption.

Let γ^{lfoc}, γ^{lfsc}, and γ^{lfcc} denote the SINR with LFOC, LFSC, and LFCC, respectively. The following theorem provides the deterministic approximations for γ^{lfoc}, γ^{lfsc}, and γ^{lfcc}, respectively.

Theorem 1: When Assumptions **A.1**-**A.3** hold, the deterministic approximations for γ^{lfoc}, γ^{lfsc}, and γ^{lfcc} are given by

$$\gamma^{\text{lfoc}} \xrightarrow[N\xrightarrow{c_1,\ldots,c_K}+\infty]{a.s.} \overline{\gamma}^{\text{lfoc}} = \mathbf{v}^H \boldsymbol{\Delta}^{-1}\mathbf{v}, \tag{28a}$$

$$\gamma^{\text{lfsc}} \xrightarrow[N\xrightarrow{c_1,\ldots,c_K}+\infty]{a.s.} \overline{\gamma}^{\text{lfsc}} = \frac{(\mathbf{v}^H \boldsymbol{\Delta}_I{}^{-1}\mathbf{v})^2}{\mathbf{v}^H \boldsymbol{\Delta}_I{}^{-1}\boldsymbol{\Delta}\boldsymbol{\Delta}_I{}^{-1}\mathbf{v}}, \tag{28b}$$

$$\gamma^{\text{lfcc}} \xrightarrow[N\xrightarrow{c_1,\ldots,c_K}+\infty]{a.s.} \overline{\gamma}^{\text{lfcc}} = \frac{|\boldsymbol{\alpha}\mathbf{J}\mathbf{v}|^2}{\boldsymbol{\alpha}\mathbf{J}\boldsymbol{\Delta}\mathbf{J}\boldsymbol{\alpha}^H}, \tag{28c}$$

with

$$[\mathbf{v}]_k = \frac{\overline{F}_k([\boldsymbol{\Phi}_0]_{[k,k]})}{N_k}, \tag{29a}$$

$$\mathbf{J} = \left[\mathbf{I}_K + \text{diag}(\mathbf{v})\right]^{-1}, \tag{29b}$$

$$[\boldsymbol{\Delta}]_{k,l} = \frac{\overline{\Upsilon}_{kl}([\boldsymbol{\Phi}_0]_{[l,k]}, [\mathbf{W}+\sigma^2\mathbf{I}_N]_{[k,l]})}{N_k N_l} + \frac{\overline{\Pi}_{B,kl}([\boldsymbol{\Phi}_0]_{[l,k]})}{\sqrt{N_k N_l}}, \tag{29c}$$

$$[\boldsymbol{\Delta}_I]_{k,l} = \frac{\overline{\Upsilon}_{kl}([\boldsymbol{\Phi}_0]_{[l,k]}, [\mathbf{D}_W+\sigma^2\mathbf{I}_N]_{[k,l]})}{N_k N_l} + \frac{\overline{\Pi}_{kl}([\boldsymbol{\Phi}_0]_{[l,k]})}{\sqrt{N_k N_l}}, \tag{29d}$$

where functions $\overline{F}_k$, $\overline{\Upsilon}_{kl}$, $\overline{\Pi}_{B,kl}$, and $\overline{\Pi}_{kl}$ can be obtained from Lemma 3 in Appendix A by setting $z_i = -\rho_i$, $\mathbf{S}_i = \mathbf{Z}_i$, $\mathbf{A}_j = \boldsymbol{\Phi}_j^{\frac{1}{2}}$, and $\mathbf{B}_j = \mathbf{V}_j\boldsymbol{\Phi}_j^{\frac{1}{2}}$ for each $i \in [K]$ and $j \in [M]$. Moreover, the following approximation for the mean holds

$$\mathbb{E}[\gamma^{\text{fs}}] = \overline{\gamma}^{\text{fs}} + \mathcal{O}\left(N^{-1}\right), \tag{30}$$

where $\text{fs} \in \{\text{lfoc}, \text{lfsc}, \text{lfcc}\}$.

Proof: The proof of Theorem 1 is given in Appendix C. □

Remark 4: When $K = 1$, which means there is only one cluster, $\overline{\gamma}^{\text{lfoc}}$, $\overline{\gamma}^{\text{lfsc}}$, and $\overline{\gamma}^{\text{lfcc}}$ degenerate to the results with centralized LMMSE [18], [19].

Remark 5: Under certain conditions, the performance of LFSC approaches that of LFOC. In particular, when $\boldsymbol{\Delta}_I = \boldsymbol{\Delta}$, we have $\overline{\gamma}^{\text{lfoc}} = \overline{\gamma}^{\text{lfsc}}$. The above condition can be achieved when $\mathbf{V}_j = \mathbf{I}_N$, for each $j \in [M]_0$, which is equivalent to

$$\mathbf{D}_{\widetilde{\sigma}}\mathbf{D}_{R,j} = \mathbf{D}_{\widetilde{\sigma}}\mathbf{R}_j. \tag{31}$$

Denote the index set of the clusters with perfect CSI as $\mathcal{K}_1 \subset [K]$ such that $\widetilde{\sigma}_{k_1}^2 = 0$, for $k_1 \in \mathcal{K}_1$. The above condition (31) can be achieved if $[\mathbf{R}_j]_{[k,l]} = \mathbf{0}_{N_k\times N_l}$ when the indexes k and l $(k \neq l)$ do not belong to the set $\mathcal{K}_1$ simultaneously. Furthermore, when the spatial correlations among all clusters are negligible, i.e., $\mathbf{R}_j = \mathbf{D}_{R,j}$ for each $j \in [M]_0$ or the CSI is perfect for all of the clusters, i.e., $\widetilde{\sigma}_k^2 = 0$ for each $k \in [K]$, (31) is satisfied automatically.

Remark 6: When there is no spatial correlation between clusters, the matrix $\boldsymbol{\Delta}$ becomes diagonal. Under such circumstances, it can be verified that when $\boldsymbol{\alpha} = C\left[\frac{(1+[\mathbf{v}]_1)[\mathbf{v}]_1}{[\boldsymbol{\Delta}]_{1,1}}, \ldots, \frac{(1+[\mathbf{v}]_K)[\mathbf{v}]_K}{[\boldsymbol{\Delta}]_{K,K}}\right]$ with $C \neq 0$ being an arbitrary constant in (28c), we have

$$\overline{\gamma}^{\text{lfoc}} = \overline{\gamma}^{\text{lfsc}} = \overline{\gamma}^{\text{lfcc}} = \sum_{k=1}^K \frac{[\mathbf{v}]_k^2}{[\boldsymbol{\Delta}]_{k,k}}, \tag{32}$$

which is equal to the sum of the SINRs from all clusters. This is consistent with the conclusion of [2, Theorem 6], which assumes uncorrelated Rayleigh channels and perfect CSI. In addition, $[\mathbf{v}]_k$ and $[\boldsymbol{\Delta}]_{k,k}$ are only related to the local spatial correlation, which varies slowly in time. As a result, the fusion

coefficients for LFCC can be set as $\alpha_k = \frac{(1+[\mathbf{v}]_k)[\mathbf{v}]_k}{[\mathbf{\Delta}]_{k,k}}$ to achieve the asymptotically optimal performance.

Remark 7: By applying Taylor expansion, it can be verified that the achievable rate can be calculated as $\mathbb{E}[\log(1+\gamma^{\text{fs}})] = \log(1+\overline{\gamma}^{\text{fs}}) + \mathcal{O}(N^{-1})$, where fs $\in \{\text{lfoc}, \text{lfsc}, \text{lfcc}\}$. These results can be utilized to evaluate the ergodic rate.

Remark 8: The result in (28) can be significantly simplified and given in a closed form with i.i.d. channels, providing useful insights for system design. More details are available in Section IV-A. It will be shown that there is a trade-off between complexity and performance. It is recommended to adopt LMMSE estimator locally, minimize the number of clusters, and avoid uniform partitioning.

Next, we investigate the case with heterogeneous CSI inaccuracy levels among the clusters. Specifically, the following corollary illustrates the behavior of the SINR when certain clusters have severe CSI errors.

Corollary 1: Let $\mathcal{K}_e$ be the index set of clusters with severe CSI errors, i.e., $\widetilde{\sigma}_k^2 \to +\infty$ for $k \in \mathcal{K}_e$, while $\widetilde{\sigma}_l^2$ is fixed for $l \in \mathcal{K}_g = [K] - \mathcal{K}_e = \{k_1, k_2, \ldots, k_G\}$, where $G = K - |\mathcal{K}_e|$. Then the deterministic approximation of the SINR satisfies

$$\overline{\gamma}^{\text{fs}} \to \overline{\gamma}^{\text{fs}}_{\mathcal{K}_g}, \tag{33}$$

where fs $\in \{\text{lfoc}, \text{lfsc}, \text{lfcc}\}$ and $\overline{\gamma}^{\text{fs}}_{\mathcal{K}_g}$ is obtained by replacing the channel $\mathbf{h}_j$ by $\mathbf{h}_j^{\mathcal{K}_g} = [\mathbf{h}_{jk_1}^T, \ldots, \mathbf{h}_{jk_G}^T]^T$ for each $j \in [M]_0$ and $\boldsymbol{\alpha}$ by $\boldsymbol{\alpha}^{\mathcal{K}_g} = [\alpha_{k_1}, \ldots, \alpha_{k_G}]$ in Theorem 1.

Proof: The proof of Corollary 1 is given in Appendix D. □

Remark 9: Corollary 1 indicates that for decentralized estimation, when several clusters have severe CSI errors, the performance will be equal to that of a system where these clusters are ignored.

A. Special Case Study: I.I.D. Channels

In this section, we will consider the simplified channel model with $\mathbf{R}_j = \mathbf{I}_N$ to get some physical insights. The parameter $\mathbf{Z}_k$ is set as (17), i.e., $\mathbf{Z}_k = \frac{(M+1)\widetilde{\sigma}_k^2}{N_k(\widetilde{\sigma}_k^2+1)}\mathbf{I}_{N_k}$.

1) SINR Approximation: With i.i.d. channels, the SINR in Theorem 1 can be given in a closed-form.

Corollary 2: Given assumptions **A.1**-**A.3** and the i.i.d. channels, $\overline{\gamma}^{\text{lfoc}}$, $\overline{\gamma}^{\text{lfsc}}$ and $\overline{\gamma}^{\text{lfcc}}$ can be given by

$$\overline{\gamma}^{\text{lfoc}} = \overline{\gamma}^{\text{lfsc}} = \sum_{k=1}^{K} \frac{\delta_k}{\varpi_k}, \tag{34a}$$

$$\overline{\gamma}^{\text{lfcc}} = \frac{|\sum_{k=1}^{K} \alpha_k \frac{\delta_k}{1+\delta_k}|^2}{\sum_{k=1}^{K} \frac{|\alpha_k|^2 \delta_k}{(1+\delta_k)^2}\varpi_k}, \tag{34b}$$

where

$$\varpi_k = 1 + \left(\frac{\sigma^2}{Mc_k} - \rho_k\right) \frac{(\widetilde{\sigma}_k^2+1)\delta_k}{1 - \frac{\delta_k^2}{c_k(1+\delta_k)^2}} \tag{35}$$

and δ_k is given by

$$\delta_k = \frac{1 - \frac{1}{c_k} - A_k + \sqrt{\left(A_k + \frac{1}{c_k} - 1\right)^2 + 4A_k}}{2A_k} \tag{36}$$

with $A_k = \rho_k(\widetilde{\sigma}_k^2+1) + \frac{M+1}{Mc_k}\widetilde{\sigma}_k^2$.

Proof: The proof can be obtained by setting $\mathbf{R}_j = \mathbf{I}_N$ and $\mathbf{Z}_k = \frac{(M+1)\widetilde{\sigma}_k^2}{N_k(\widetilde{\sigma}_k^2+1)}$ in Theorem 1 and is omitted here. □

Remark 10: According to the definition, it can be verified $\delta_k = [\mathbf{v}]_k$ and $\varpi_k = \frac{[\mathbf{\Delta}]_{k,k}}{[\mathbf{v}]_k}$ in Theorem 1.

Denote $\boldsymbol{\rho}_K = [\rho_1, \ldots, \rho_K]^T$ and $\mathbf{c}_K = [c_1, \ldots, c_K]^T$. We next investigate the impact of parameters $\boldsymbol{\rho}_K$, $\mathbf{c}_K$, and the number of clusters K on the SINR $\overline{\gamma}^{\text{lfoc}} = \overline{\gamma}^{\text{lfoc}}_K(\boldsymbol{\rho}_K, \mathbf{c}_K)$ where we keep N and M fixed. We also assume the CSI inaccuracy level is identical among clusters, i.e., $\widetilde{\sigma}_k^2 = \widetilde{\sigma}^2$ for each $k \in [K]$ in the following.

2) Regularization Parameter: The following corollary gives the optimal parameter $\boldsymbol{\rho}_K$.

Corollary 3: Assume the same conditions as Corollary 2. For any $\boldsymbol{\rho}_K > 0$, $\mathbf{c}_K \geq 0$, we have

$$\overline{\gamma}^{\text{lfoc}}_K(\boldsymbol{\rho}_K, \mathbf{c}_K) \leq \overline{\gamma}^{\text{lfoc}}_K\left(\boldsymbol{\rho}_K^{\text{opt}}, \mathbf{c}_K\right), \tag{37}$$

where $\boldsymbol{\rho}_K^{\text{opt}} = [\frac{\sigma^2}{Mc_1}, \ldots, \frac{\sigma^2}{Mc_K}]^T$.

Proof: The proof of Corollary 3 is omitted due to limited space, and can be found in the extended version [31]. □

Remark 11: We can observe that the optimal parameter $\boldsymbol{\rho}_K^{\text{opt}}$ that maximizes the SINR is identical to that achieving the minimal local MSE as derived in Section III-A. This can be intuitively obtained from (34a) where the SINR $\overline{\gamma}^{\text{lfoc}}$ is the sum of the SINRs of all clusters.

3) Antenna Partitioning Strategy: The following corollary shows the impact of the parameter $\mathbf{c}_K$.

Corollary 4: Assume the same conditions as Corollary 2 and denote $\boldsymbol{\rho}_K = [\frac{a}{c_1}, \ldots, \frac{a}{c_K}]^T$. For any $\mathbf{c}_K \geq 0$ and $\frac{\sigma^2}{M} \leq a$, we have

$$\overline{\gamma}^{\text{lfoc}}_K(\boldsymbol{\rho}_K, \mathbf{c}_K) \leq \overline{\gamma}^{\text{lfoc}}_K(\boldsymbol{\rho}_K, \mathbf{c}_{K,\text{M}}), \tag{38}$$

and for any $\mathbf{c}_K \geq 0$ and $\frac{\sigma^2}{M} \leq a \leq \frac{2\sigma^2}{M} + \frac{(M+1)\widetilde{\sigma}^2}{M(\widetilde{\sigma}^2+1)}$, we have

$$\overline{\gamma}^{\text{lfoc}}_K\left(\boldsymbol{\rho}_K, \mathbf{c}_{K,\text{m}}\right) \leq \overline{\gamma}^{\text{lfoc}}_K\left(\boldsymbol{\rho}_K, \mathbf{c}_K\right), \tag{39}$$

where $\mathbf{c}_{K,\text{m}} = [\overline{c}, \ldots, \overline{c}]^T$, $\mathbf{c}_{K,\text{M}} = [K\overline{c}, 0, \ldots, 0]^T$, and $\overline{c} = \frac{1}{K}\sum_{k=1}^{K} c_k$.

Proof: The proof of Corollary 4 is omitted due to limited space, and can be found in the extended version [31]. □

Remark 12: Inequalities (38) and (39) imply that when the regularization parameter $\boldsymbol{\rho}_K$ is within a certain range, i.e., $\frac{\sigma^2}{M} \leq a$, the system achieves the maximum SINR when the receive antennas are concentrated in one cluster, i.e., the centralized scheme. On the other side, when the antennas are equally partitioned, the system achieves the lowest SINR. This finding aligns with the conclusions in [2, Lemma 7, Lemma 8]. Note that [2] considers a given parameter $\boldsymbol{\rho}_K$ with perfect CSI, but Corollary 4 provides the range for $\boldsymbol{\rho}_K$ with imperfect CSI.

Remark 13: It can be observed that the optimal regularization parameter $\boldsymbol{\rho}_K^{\text{opt}}$ in Corollary 3 is within the range such that (38) and (39) hold.

4) Number of Clusters: Corollary 4 compares different antenna partitioning strategies with a fixed number of clusters K. Another interesting question to investigate is how the number

of clusters K affects the SINR. The following corollary demonstrates this impact.

Corollary 5: Assume the same conditions as Corollary 2 and denote $\boldsymbol{\rho}_K = [\frac{a}{c_1}, \ldots, \frac{a}{c_K}]^T$. For any $\frac{\sigma^2}{M} \leq a$, we have

$$\overline{\gamma}_K^{\text{lfoc}}(\boldsymbol{\rho}_K, \mathbf{c}_{K,\text{m}}) \geq \overline{\gamma}_{K+1}^{\text{lfoc}}(\boldsymbol{\rho}_{K+1}, \mathbf{c}_{K+1,\text{m}}). \tag{40}$$

Proof: The proof of Corollary 5 is omitted due to limited space, and can be found in the extended version [31]. □

Remark 14: Since the total number of receive antennas N is fixed, Corollary 5 indicates that as the number of clusters increases, the SINR will monotonically decrease.

Remark 15: It can be verified from the proof of Corollary 5 that

$$\overline{\gamma}_K^{\text{lfoc}}(\boldsymbol{\rho}_K, \mathbf{c}_{K,\text{m}}) \geq \frac{N}{(\sigma^2 + M)(\widetilde{\sigma}^2 + 1) + \widetilde{\sigma}^2}, \tag{41}$$

and the equality holds when $K \to +\infty$, which is not achievable since $K \leq N$. From Corollary 4, we know that the equal antenna partitioning strategy has the worst performance. Therefore, for decentralized LMMSE receivers, the SINR is strictly lower bounded by the right hand side (RHS) of (41). When the numbers of antennas and users are large, this lower bound is approximately $\frac{N}{M}$.

V. Numerical Results

In this section, we will verify the accuracy of the theoretical results in Section IV. We will also demonstrate the impact of several key system parameters over i.i.d. channels, including the regularization parameter, antenna partitioning strategy, and the number of clusters. For simplicity, we use the notation $\mathcal{N}_K = (N_1, \ldots, N_K)$ to denote the antenna partitioning strategy. The Monte Carlo (MC) simulation results were obtained from 5000 realizations and are illustrated by markers in all figures.

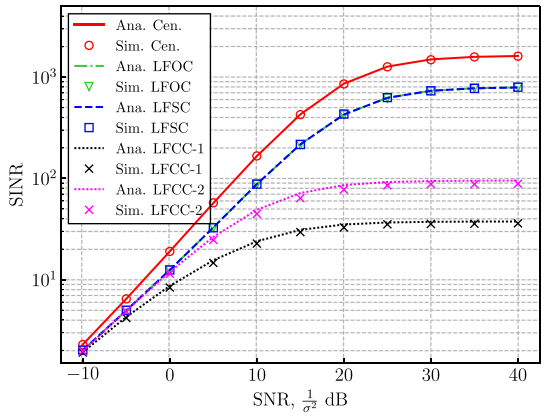
A. Correlated Channels

We consider a uniform linear array of antennas at the BS. The (m, n)-th entry of the spatial correlation matrix at BS is modeled as [32], [33]

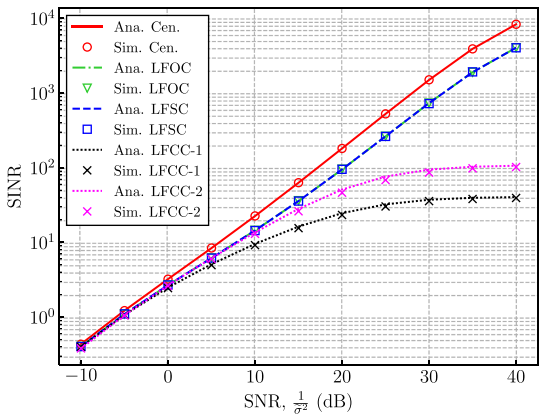
$$[\mathbf{C}(\eta, \delta, d_s)]_{m,n} = \int_{-180^\circ}^{180^\circ} \frac{1}{\sqrt{2\pi\delta^2}} e^{2\pi \jmath d_s (m-n)\sin(\frac{\pi\phi}{180}) - \frac{(\phi-\eta)^2}{2\delta^2}} \mathrm{d}\phi. \tag{42}$$

Here, d_s and ϕ denote the receive antenna spacing and angular spread, respectively, η represents the mean angle, and δ denotes the root-mean-square angle spread. In the simulation, we set $\mathbf{R}_j = \mathbf{C}(\eta_j, \delta_j, 1)$ with $\eta_j = (\frac{j}{180M})^\circ$ and $\delta_j = 10^\circ + (\frac{j}{M10})^\circ$. The parameter ρ_k is set as $\rho_k = \frac{\sigma^2}{N_k}$ and $\mathbf{Z}_k$ is set according to (17) for each $k \in [K]$.

Approximation of SINR: Fig. 2 illustrates the SINR with different linear fusion schemes. The system dimensions are set as $N = 32$, $M = 12$, $K = 2$, and $\mathcal{N}_2 = (10, 22)$. The red line and the two dotted lines represent the results of centralized LMMSE and LFCC, respectively. The fusion parameter is set as (24) for LFCC-1 and (25) for LFCC-2, respectively. It is shown that centralized LMMSE performs better than the three decentralized



(a) SINR versus signal SNR $\frac{1}{\sigma^2}$ with training SNR $\frac{1}{\widetilde{\sigma}_1^2} = \frac{1}{\widetilde{\sigma}_2^2} = 30$ dB.



(b) SINR versus training SNR $\frac{1}{\widetilde{\sigma}_1^2} = \frac{1}{\widetilde{\sigma}_2^2} = \frac{1}{\widetilde{\sigma}^2}$ with signal SNR $\frac{1}{\sigma^2} = 30$ dB.

Fig. 2. SINR with different linear fusion schemes.

schemes. By comparing the results with MC simulations, we can observe that the approximation results in Theorem 1 are accurate. By comparing the results of LFSC and LFOC, it can be observed that LFSC is highly efficient. Furthermore, Fig. 2 shows that when the antenna partition is unbalanced, using LFCC will lead to performance degradation.

BER Performance: Fig. 3 shows the bit error rate (BER) with 16-QAM modulation for different fusion schemes. The system parameters are set as $N = 108$, $M = 52$, $K = 2$, and $\mathcal{N}_2 = (20, 88)$. The fusion parameter for LFCC is set as $\boldsymbol{\alpha} = [0.5, 0.5]$ and the signal SNR and training SNR of two clusters are set to be equal, with $\frac{1}{\sigma^2} = \frac{1}{\widetilde{\sigma}_1^2} = \frac{1}{\widetilde{\sigma}_2^2}$, represented by the horizontal axis in Fig. 3. It can be observed that the BER for centralized architecture is much lower than that with the decentralized scheme.

Heterogeneous CSI Inaccuracy: Fig. 4 illustrates the performance of SINR with different levels of CSI inaccuracy among clusters. The system parameters are set as $N = 40$, $M = 16$, $K = 2$, $\mathcal{N}_2 = (22, 18)$, and $\frac{1}{\sigma^2} = \frac{1}{\widetilde{\sigma}_2^2} = 30$ dB. The fusion parameter for LFCC is set as $\boldsymbol{\alpha} = [0.5, 0.5]$. The red line represents the performance of centralized estimation considering only the

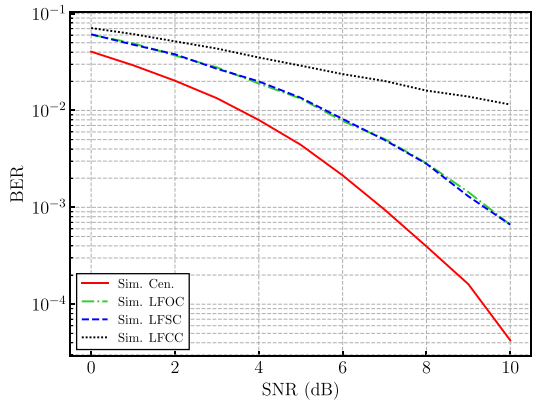


Fig. 3. BER versus SNR.

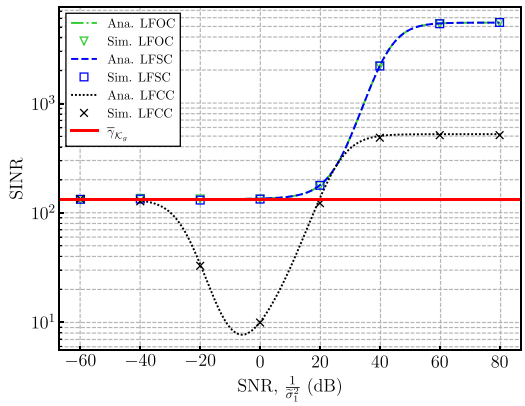


Fig. 4. SINR versus training SNR $\frac{1}{\sigma_1^2}$.

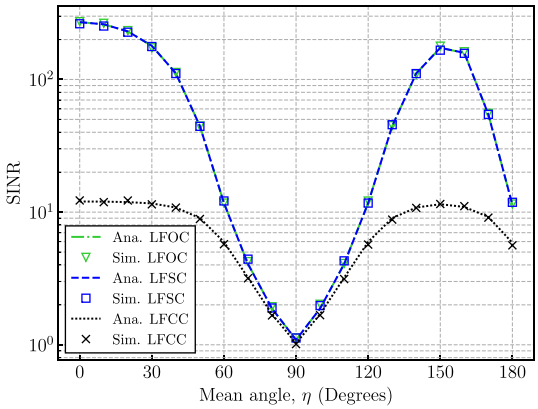


Fig. 5. SINR versus mean angle η.

second cluster. It can be observed that when the training SNR of the first cluster is very low, the performance approaches that without considering the first cluster, which validates Corollary 1.

Impact of Spatial Correlations: Fig. 5 illustrates the impact of the mean angle of the spatial correlation matrices on SINR for different fusion schemes. The spatial correlations are set to be $\mathbf{R}_j = \mathbf{C}(\eta, 20, 1)$. The parameters are set as $N = 40$, $M = 18$, $\mathcal{N}_2 = (28, 12)$, and $\frac{1}{\sigma^2} = \frac{1}{\widetilde{\sigma}_1^2} = \frac{1}{\widetilde{\sigma}_2^2} = 30$ dB. For LFCC, $\boldsymbol{\alpha}$ is set as $[0.5, 0.5]$. It can be observed that when $\eta = 90°$, the SINR is the lowest. This is because when $\eta = 90°$, the spatial correlation is strong, leading to many eigenvalues of $\mathbf{R}_j$ being close to zero.

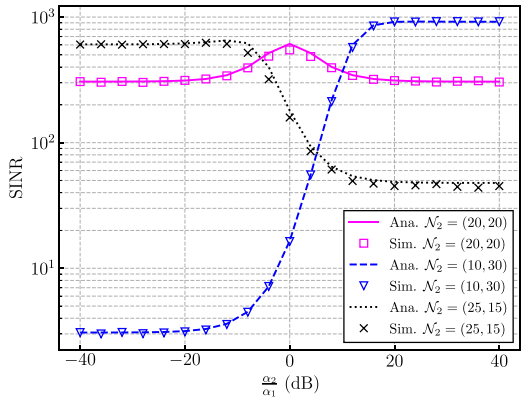


Fig. 6. SINR versus the ratio of fusion coefficients $\frac{\alpha_2}{\alpha_1}$.

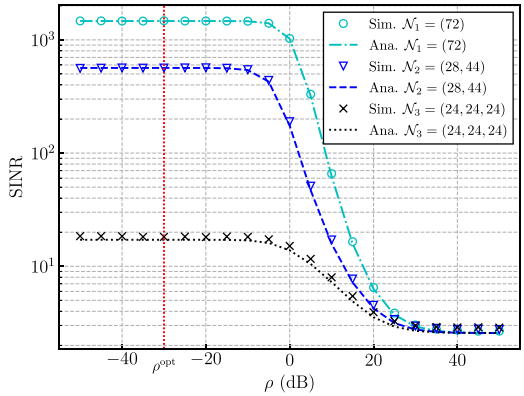


Fig. 7. SINR versus the regularization parameter ρ.

Impact of Fusion Coefficients: Fig. 6 shows the impact of the fusion coefficients with different antenna partitioning strategies. The system dimensions are set as $N = 40$, $M = 15$, and $K = 2$. The signal and training SNR are both set as 30 dB. It can be observed that when antennas are uniformly allocated to different clusters, i.e., $\mathcal{N}_2 = (20, 20)$, SINR is not sensitive to the value of $\boldsymbol{\alpha}$, but the maximum SINR is lower than that with unbalanced antenna partitioning. This observation for correlated channels is consistent with the conclusion for i.i.d. channels in Corollary 4. Moreover, as the number of antennas in a cluster increases, the corresponding linear fusion weights should also increase.

B. I.I.D. Channels

In the experiment, the training and signal SNR are both 30 dB, and the parameter $\mathbf{Z}_k$ is set according to (17) for each $k \in [K]$.

Optimal Regularization Parameter: Fig. 7 depicts the impact of $\boldsymbol{\rho}_K$ on the SINR with i.i.d. channels. The system dimensions are set as $N = 72$ and $M = 40$. The parameter $\boldsymbol{\rho}_K$ is of the form $\boldsymbol{\rho}_K = [\frac{\rho}{N_1}, \ldots, \frac{\rho}{N_K}]^T$. It can be observed that SINR is not sensitive to ρ. Significant performance loss is only observed when ρ is very large. Furthermore, according to Corollary 3, the theoretical optimal parameter $\boldsymbol{\rho}_K^{\text{opt}}$ should be $\boldsymbol{\rho}_K^{\text{opt}} = (-30 \text{ dB}) \cdot [\frac{1}{N_1}, \ldots, \frac{1}{N_1}]^T$. The optimal value $\rho^{\text{opt}} = -30$ dB shown in the

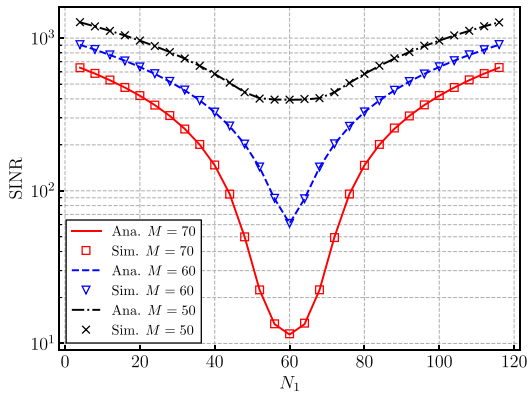


Fig. 8. SINR versus the number of antennas allocated to the first cluster N_1.

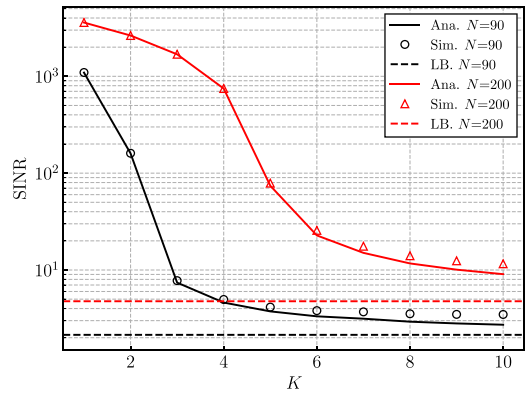


Fig. 9. SINR versus K.

figure obtained by the exhausted search validates the accuracy of Corollary 3.

Impact of Antenna Partitioning Strategy: In Fig. 8, we plot the SINR for different antenna partitioning strategies with different numbers of users. The parameters are set as $N = 120$, $K = 2$, and $\rho_k = \frac{\sigma^2}{N_k}$. It is observed that when clusters are of equal size, the SINR is the lowest, which agrees with Corollary 4. Furthermore, as the number of users increases, the performance degradation with uniform antenna partitioning becomes more obvious.

Impact of Number of Clusters: In Fig. 9, we plot the SINR for different numbers of clusters K where antennas are almost equally allocated with $\mathcal{N}_K = (\lfloor \frac{N}{K} \rfloor, \lfloor \frac{N}{K} \rfloor, \ldots, N-(K-1)\lfloor \frac{N}{K} \rfloor)$. The parameters are set as $M = 40$ and $\rho_k = \frac{\sigma^2}{N_k}$. It can be observed that as the number of clusters increases, the SINR decreases monotonically and is lower bounded by (41), which is shown by the dashed lines. This validates Corollary 5.

VI. Conclusion

In this paper, we considered the decentralized MIMO systems and imperfect CSI using LMMSE receivers. It was shown that with DBP, the linear fusion coefficients that maximize the receive SINR also minimize the MSE for estimating the transmit signal. To reduce the complexity, we further proposed two linear fusion schemes and derived the deterministic approximations of the SINR with imperfect CSI and general spatial correlations. Based on the results, we investigated the optimal LMMSE regularization parameters, optimal antenna partitioning strategies, and the impact of the number of clusters. It was shown that with i.i.d. channels, the SINR is minimized when the clusters are of equal size. Conversely, the optimal performance is achieved when antennas are mostly unbalanced distributed, i.e., all antennas are allocated to a single cluster. We also proved that as the number of clusters increases, the SINR monotonically decreases.

From the perspective of RMT, we investigated the asymptotic convergence properties of the resolvent for the covariance matrix with generally correlated columns and its related functions. These results can be utilized for the second-order SINR analysis, e.g., the outage probability, and be applied to downlink scenarios, e.g., the analysis of RZF.

Appendix A
Useful Results

In this section, we will give the main mathematical results, which will be frequently used throughout the proofs in this paper.

A. Matrix Inversion Lemma

Lemma 1: [23, Lemma 1]: Let $\mathbf{M} \in \mathbb{C}^{N\times N}$ be an invertible matrix, $\mathbf{m} \in \mathbb{C}^N$, and $a \in \mathbb{C}$. If $1 + a\mathbf{m}^H\mathbf{M}^{-1}\mathbf{m} \neq 0$, we have

$$\mathbf{m}^H\left(\mathbf{M}+a\mathbf{m}\mathbf{m}^H\right)^{-1} = \frac{\mathbf{m}^H\mathbf{M}^{-1}}{1+a\mathbf{m}^H\mathbf{M}^{-1}\mathbf{m}}. \tag{43}$$

B. Convergence of Random Vectors

Lemma 2: [23, Lemma 4]: Let $\mathbf{A}_N \in \mathbb{C}^{N\times N}$ be a random matrix and $\mathbf{x}_N \in \mathbb{C}^N$ be a random vector with i.i.d. entries with normalized Gaussian distribution $\mathcal{CN}(0,1)$. Assume that for each j, $[\mathbf{x}]_j$ is independent with sequence $(\mathbf{A}_N)_{N\geq 1}$ and $\mathbb{P}(\sup_{N\geq 1}\|\mathbf{A}_N\| < +\infty) = 1$. Then, we have

$$\frac{\mathrm{Tr}(\mathbf{x}_N^H\mathbf{A}_N\mathbf{x}_N)}{N} - \frac{\mathrm{Tr}\mathbf{A}_N}{N} \xrightarrow[N\to+\infty]{a.s.} 0. \tag{44}$$

Proof: Here we provide a proof different from [23] which relies on Tonelli's theorem. To describe independence, [23] assumes the underlying probability space is the product probability space of $(\mathbf{x}_N)_{N\geq 1}$ and $(\mathbf{A}_N)_{N\geq 1}$, whereas here we discuss within the same probability space.

Denoting the underline probability space $(\Omega, \mathcal{F}, \mathbb{P})$. Define the set $A_{n,K} = \{\omega \in \Omega : \sup_{j\geq n}\|\mathbf{A}_j(\omega)\| \leq K\}$ and $B_{N,\epsilon} = \{\omega \in \Omega : d_N(\omega) \leq \epsilon\}$, where

$$d_N(\omega) = \left|\frac{\mathrm{Tr}[\mathbf{x}_N^H(\omega)\mathbf{A}_N(\omega)\mathbf{x}_N(\omega)]}{N} - \frac{\mathrm{Tr}\mathbf{A}_N(\omega)}{N}\right|. \tag{45}$$

To prove $d_N \xrightarrow[N\to+\infty]{a.s.} 0$, we only need to show that for any positive integer n and $\epsilon > 0$, $\mathbb{P}(\bigcup_{n=1}^{+\infty}\bigcap_{N=n}^{+\infty} B_{N,\epsilon}) = 1$ holds. In the following, we will prove this by a stronger result $\mathbb{P}(\bigcup_{n=1}^{+\infty}\bigcap_{N=n}^{+\infty} B_{N,\epsilon}) \geq 1$. To this end, we write

$$\mathbb{P}\left(\bigcup_{n=1}^{+\infty}\bigcap_{N=n}^{+\infty} B_{N,\epsilon}\right) \geq \mathbb{P}\left(\bigcup_{n=1}^{+\infty}\bigcap_{N=n}^{+\infty} B_{N,\epsilon}\bigcap\bigcup_{K=0}^{+\infty} A_{n,K}\right)$$

$$= \mathbb{P}\left(\bigcup_{K=0}^{+\infty}\bigcup_{n=1}^{+\infty}\bigcap_{N=n}^{+\infty} B_{N,\epsilon}\bigcap A_{n,K}\right)$$

$$\overset{(a)}{=} \lim_{K\to+\infty}\lim_{n\to+\infty}\mathbb{P}\left(\bigcap_{N=n}^{+\infty} B_{N,\epsilon}\bigcap A_{n,K}\right), \quad (46)$$

where step (a) follows from $A_{n,K}\subset A_{n,K+1}$, $A_{n,K}\subset A_{n+1,K}$, and continuity of probability measure. On the other side, we have

$$\mathbb{P}\left(\bigcap_{N=n}^{+\infty} B_{N,\epsilon}\bigcap A_{n,K}\right) = 1-\mathbb{P}\left(\bigcup_{N=n}^{+\infty} B^c_{N,\epsilon}\bigcup A^c_{n,K}\right)$$

$$= 1-\mathbb{P}(A^c_{n,K})-\mathbb{P}\left(\bigcup_{N=n}^{+\infty} B^c_{N,\epsilon}\bigcap A_{n,K}\right)$$

$$\geq \mathbb{P}(A_{n,K})-\sum_{N\geq n}\mathbb{P}(B^c_{N,\epsilon}\bigcap A_{n,K}). \quad (47)$$

By using Markov's inequality, we can get

$$\mathbb{P}(B^c_{N,\epsilon}\bigcap A_{n,K}) \leq \mathbb{E}\mathbb{I}_{\{d_N\geq\epsilon\}}\mathbb{I}_{\{\|\mathbf{A}_N\|\leq K\}}$$

$$\leq \mathbb{E}\left[\mathbb{I}_{\{\|\mathbf{A}_N\|\leq K\}}\mathbb{E}(\epsilon^{-p}d_N^p|\mathbf{A}_N)\right]$$

$$\overset{(a)}{\leq} C_p N^{-\frac{p}{2}}\mathbb{E}[\mathbb{I}_{\{\|\mathbf{A}_N\|\leq K\}}\|\mathbf{A}_N\|^p] = \mathcal{O}(N^{-\frac{p}{2}}K^p), \quad (48)$$

where $p>2$, C_p is a constant, and step (a) follows from the trace lemma [23, Lemma 3]. Hence, $\sum_{N=n}^{+\infty}\mathbb{P}(B^c_{N,\epsilon}\bigcap A_{n,K})\to 0$ as $n\to+\infty$ and it holds true that

$$\lim_{n\to+\infty}\mathbb{P}\left(\bigcap_{N=n}^{+\infty} B_{N,\epsilon}\bigcap A_{n,K}\right) \geq \lim_{n\to+\infty}\mathbb{P}(A_{n,K}). \quad (49)$$

By combining (49) and (46), we can obtain

$$\mathbb{P}\left(\bigcup_{n=1}^{+\infty}\bigcap_{N=n}^{+\infty} B_{N,\epsilon}\right) \geq \lim_{K\to+\infty}\lim_{n\to+\infty}\mathbb{P}(A_{n,K})$$

$$\overset{(a)}{=} \mathbb{P}\left(\bigcup_{k=0}^{+\infty}\bigcup_{n=1}^{+\infty} A_{n,K}\right) = \mathbb{P}(\bigcup_{n=1}^{+\infty}\bigcup_{K=0}^{+\infty} A_{n,K}) \overset{(b)}{=} 1, \quad (50)$$

where step (a) follows from continuity of probability measure and step (b) follows from the identity $\mathbb{P}(\bigcup_{K=0}^{+\infty}A_{n,k})=1$. This completes the proof of Lemma 2. □

C. Convergence of Random Matrices

The SINR involves terms associated with random matrices $\widehat{\mathbf{\Sigma}}$ and $\widetilde{\mathbf{\Sigma}}$ whose columns are generally correlated. This makes the evaluation of the SINR very challenging. In the following, we introduce a more general setting and provide the relevant results, which can be utilized to prove Theorem 1.

Let $\widetilde{\mathbf{X}}=[\mathbf{A}_1\mathbf{x}_1,\ldots,\mathbf{A}_M\mathbf{x}_M]\in\mathbb{C}^{N\times M}$ and $\widetilde{\mathbf{Y}}=[\mathbf{B}_1\mathbf{x}_1,\ldots,\mathbf{B}_M\mathbf{x}_M]\in\mathbb{C}^{N\times M}$, where $\mathbf{A}_j,\mathbf{B}_j\in\mathbb{C}^{N\times N_i^x}$, and $\mathbf{x}_j\sim\mathcal{CN}(0,\mathbf{I}_{N_j^x})$ for each $j\in[M]$. Define two correlation matrices sets $A=\{\mathbf{A}_j: j\in[M]\}$ and $B=\{\mathbf{B}_j: j\in[M]\}$. Denote $\mathbf{\Omega}_j=\mathbf{A}_j\mathbf{A}_j^H$, $\mathbf{A}_{jk}=\mathbf{A}_j[\varsigma_k,:]$, and $\mathbf{B}_{jk}=\mathbf{B}_j[\varsigma_k,:]$. Let $\mathbf{X}_k=\frac{1}{\sqrt{N_k}}\widetilde{\mathbf{X}}[\varsigma_k,:]$ and $\mathbf{Y}_k=\frac{1}{\sqrt{N_k}}\widetilde{\mathbf{Y}}[\varsigma_k,:]$. The resolvent matrix of $\mathbf{X}_k\mathbf{X}_k^H+\mathbf{S}_k$ is given by

$$\mathbf{Q}_k(z)=\left(\mathbf{X}_k\mathbf{X}_k^H+\mathbf{S}_k-z\mathbf{I}_{N_k}\right)^{-1}, \quad (51)$$

with $\mathbf{S}_k$ being deterministic and Hermitian nonnegative. We further define some functions of the resolvents in Table II, where we set $\mathbf{Q}_k=\mathbf{Q}_k(z_k)$, for each $k\in[K]$ and omit the variable z_k for the sake of notation simplicity. Furthermore, $\boldsymbol{\mathcal{A}}$ and $\boldsymbol{\mathcal{B}}$ are appropriately sized matrices, and $\mathbf{b}$ is a properly dimensioned vector. We next introduce the stochastic equivalent notation.

TABLE II
TABLE OF NOTATIONS

Notation	Expression
$\overline{\Gamma}_k(\boldsymbol{\mathcal{A}})$	$\mathrm{Tr}\,\boldsymbol{\mathcal{A}}\mathbf{Q}_k$
$\overline{\Phi}_{B,kl}(\boldsymbol{\mathcal{A}},\mathbf{b})$	$\mathrm{Tr}\,\boldsymbol{\mathcal{A}}\mathbf{Q}_k\mathbf{X}_k\mathrm{diag}(\mathbf{b})\mathbf{Y}_l^H$
$\overline{\Phi}_{kl}(\boldsymbol{\mathcal{A}},\mathbf{b})$	$\mathrm{Tr}\,\boldsymbol{\mathcal{A}}\mathbf{Q}_k\mathbf{X}_k\mathrm{diag}(\mathbf{b})\mathbf{X}_l^H$
$\overline{\Upsilon}_{kl}(\boldsymbol{\mathcal{A}},\boldsymbol{\mathcal{B}})$	$\mathrm{Tr}\,\boldsymbol{\mathcal{A}}\mathbf{Q}_k\boldsymbol{\mathcal{B}}\mathbf{Q}_l$
$\overline{\Pi}_{B,kl}(\boldsymbol{\mathcal{A}})$	$\mathrm{Tr}\,\boldsymbol{\mathcal{A}}\mathbf{Q}_k\mathbf{Y}_k\mathbf{Y}_l^H\mathbf{Q}_l$
$\overline{\Pi}_{kl}(\boldsymbol{\mathcal{A}})$	$\mathrm{Tr}\,\boldsymbol{\mathcal{A}}\mathbf{Q}_k\mathbf{X}_k\mathbf{X}_l^H\mathbf{Q}_l$

Definition 1: (Stochastic Equivalent) For random sequence $(X_N)_{N\geq1}$ and deterministic sequence $(\overline{X}_N)_{N\geq1}$, we use $X_N\asymp_l\overline{X}_N$ to denote

$$|\mathbb{E}X_N-\overline{X}_N|=\mathcal{O}(N^{-l}),\ \frac{|X_N-\overline{X}_N|}{N}\xrightarrow[N\to+\infty]{a.s.}0. \quad (52)$$

The following lemma gives the stochastic equivalents for the functions in Table II.

Lemma 3: Assume Assumption **A.1** holds, the maximal spectral norm of the correlation matrices is bounded, i.e., $\limsup_N\max_{j\in[M]}\{\|\mathbf{A}_j\|,\|\mathbf{B}_j\|\}<+\infty$, and the spectral norms of matrices $\boldsymbol{\mathcal{A}},\boldsymbol{\mathcal{B}}$ and $\mathrm{diag}(\mathbf{b})$ are uniformly bounded. For any $z_1,\ldots,z_K<0$, we have $\Gamma_k\asymp_{\frac{1}{2}}\overline{\Gamma}_k$, $\Phi_{B,kl}\asymp_{\frac{1}{2}}\overline{\Phi}_{B,kl}$, $\Phi_{kl}\asymp_{\frac{1}{2}}\overline{\Phi}_{kl}$, $\Upsilon_{kl}\asymp_{\frac{1}{2}}\overline{\Upsilon}_{kl}$, $\Pi_{B,kl}\asymp_{\frac{1}{2}}\overline{\Pi}_{B,kl}$, and $\Pi_{kl}\asymp_{\frac{1}{2}}\overline{\Pi}_{kl}$, where

$$\overline{\Gamma}_k(\boldsymbol{\mathcal{A}})=\mathrm{Tr}\boldsymbol{\mathcal{A}}\mathbf{\Theta}_k, \quad (53)$$

$$\overline{\Phi}_{B,kl}(\boldsymbol{\mathcal{A}},\mathbf{b})=\frac{\sum_{j=1}^M[\widetilde{\mathbf{F}}_k\mathbf{b}]_j\mathrm{Tr}\boldsymbol{\mathcal{A}}\mathbf{\Theta}_k\mathbf{A}_{jk}\mathbf{B}_{jl}^H}{\sqrt{N_kN_l}}, \quad (54)$$

$$\overline{\Upsilon}_{kl}(\boldsymbol{\mathcal{A}},\boldsymbol{\mathcal{B}})=\mathrm{Tr}\boldsymbol{\mathcal{A}}\mathbf{\Theta}_k\boldsymbol{\mathcal{B}}\mathbf{\Theta}_l + \widetilde{\boldsymbol{\lambda}}_{kl}(\boldsymbol{\mathcal{A}})\widetilde{\mathbf{F}}_k\widetilde{\mathbf{F}}_l\mathbf{\Xi}_{kl}\boldsymbol{\lambda}_{kl}(\boldsymbol{\mathcal{B}}), \quad (55)$$

and $\overline{\Pi}_{kl,B}$ is given in (56) shown at the bottom of the next page. Functions $\overline{\Phi}_{kl}$ and $\overline{\Pi}_{kl}$ can be obtained by setting $\mathbf{B}_j=\mathbf{A}_j$ for each $j\in[M]$ in $\overline{\Phi}_{kl,B}$ and $\overline{\Pi}_{kl,B}$, respectively and we omit them here. The deterministic variables are defined in Table III where P and O can be any symbol from the set $\{A,B\}$. The matrix $\mathbf{\Theta}_k$ is defined by the positive solutions of the following system of equations w.r.t. δ_{jk}

$$\delta_{jk}=\frac{1}{N_k}\mathrm{Tr}[\mathbf{\Omega}_j]_{[k,k]}\mathbf{\Theta}_k, j\in[M], \quad (57)$$

TABLE III
TABLE OF DETERMINISTIC QUANTITIES

Notation	Expression	Notation	Expression
$\widetilde{\mathbf{F}}_k$	$\operatorname{diag}\left(\left(1+\delta_{jk}\right)^{-1}; j\in[M]\right)$	$[\mathbf{\Gamma}_{kl}]_{i,j}$	$\frac{1}{N_kN_l}\operatorname{Tr}[\mathbf{\Omega}_i]_{[l,k]}\mathbf{\Theta}_k[\mathbf{\Omega}_j]_{[k,l]}\mathbf{\Theta}_l$
$\mathbf{\Xi}_{kl}$	$\left(\mathbf{I}_M-\mathbf{\Gamma}_{kl}\widetilde{\mathbf{F}}_k\widetilde{\mathbf{F}}_l\right)^{-1}$	$\widetilde{\boldsymbol{\lambda}}_{kl}(\boldsymbol{\mathcal{A}})$	$\frac{\mathbf{1}_M^T}{\sqrt{N_kN_l}}\operatorname{diag}\left(\operatorname{Tr}\boldsymbol{\mathcal{A}}\mathbf{\Theta}_k[\mathbf{\Omega}_j]_{[k,l]}\mathbf{\Theta}_l; j\in[M]\right)$
$\widetilde{\boldsymbol{\lambda}}_{P,O,kl}(\boldsymbol{\mathcal{A}})$	$\frac{\mathbf{1}_M^T}{\sqrt{N_kN_l}}\operatorname{diag}\left(\operatorname{Tr}\boldsymbol{\mathcal{A}}\mathbf{\Theta}_k\mathbf{P}_{jk}\mathbf{O}_{jl}^H\mathbf{\Theta}_l; j\in[M]\right)$	$\boldsymbol{\lambda}_{kl}(\boldsymbol{\mathcal{B}})$	$\frac{1}{\sqrt{N_kN_l}}\operatorname{diag}\left(\operatorname{Tr}[\mathbf{\Omega}_j]_{[l,k]}\mathbf{\Theta}_k\boldsymbol{\mathcal{B}}\mathbf{\Theta}_l; j\in[M]\right)\mathbf{1}_M$
$[\mathbf{\Lambda}_{P,O,kl}]_{i,j}$	$\frac{1}{N_kN_l}\operatorname{Tr}[\mathbf{\Omega}_i]_{[l,k]}\mathbf{\Theta}_k\mathbf{P}_{jk}\mathbf{O}_{jl}^H\mathbf{\Theta}_l$	$\mathbf{D}_{P,O,k}$	$\frac{1}{N_k}\operatorname{diag}\left(\operatorname{Tr}\mathbf{P}_{jk}\mathbf{O}_{jk}^H\mathbf{\Theta}_k; j\in[M]\right)$

$$\mathbf{\Theta}_k=\left(-z_k\mathbf{I}_{N_k}+\mathbf{S}_k+\frac{1}{N_k}\sum_{j=1}^{M}\frac{[\mathbf{\Omega}_j]_{[k,k]}}{1+\delta_{jk}}\right)^{-1}. \tag{58}$$

Proof: The proof of Lemma 3 is given in Appendix E. □

APPENDIX B
PROOF OF PROPOSITION 1

Denote $\widetilde{\mathbf{H}}=[\widetilde{\mathbf{h}}_1,\ldots,\widetilde{\mathbf{h}}_M]$ and $\widehat{\mathbf{H}}_k=[\widehat{\mathbf{h}}_{1k},\ldots,\widehat{\mathbf{h}}_{Mk}]$. By taking $\mathbf{r}_k=\mathbf{r}_k^{\text{mmse}}$ in (13), straight-forward computations yield

$$\gamma_0=\frac{\boldsymbol{\alpha}\mathbf{m}\mathbf{m}^H\boldsymbol{\alpha}^H}{\boldsymbol{\alpha}\mathbf{M}\boldsymbol{\alpha}^H}, \tag{59}$$

where $\mathbf{M}=\mathbf{D}_r^H(\widetilde{\mathbf{H}}\widetilde{\mathbf{H}}^H+\mathbf{W}+\sigma^2\mathbf{I}_N)\mathbf{D}_r$ and $\mathbf{m}=\mathbf{D}_r^H\widetilde{\mathbf{h}}_0$. One can validate that $\mathbf{M}$ is almost surely invertible. By letting $\widetilde{\boldsymbol{\alpha}}=\boldsymbol{\alpha}\mathbf{M}^{\frac{1}{2}}$ and $\widetilde{\mathbf{m}}=\mathbf{M}^{-\frac{1}{2}}\mathbf{m}$, we have

$$\gamma_0=\frac{\widetilde{\boldsymbol{\alpha}}\widetilde{\mathbf{m}}\widetilde{\mathbf{m}}^H\widetilde{\boldsymbol{\alpha}}^H}{\widetilde{\boldsymbol{\alpha}}\widetilde{\boldsymbol{\alpha}}^H}, \tag{60}$$

where γ_0 is maximized when $\widetilde{\boldsymbol{\alpha}}=\widetilde{c}\widetilde{\mathbf{m}}^H$ for $\widetilde{c}\neq 0$ according to Cauchy-Schwartz inequality. The optimal $\boldsymbol{\alpha}$ is given by

$$\begin{aligned}\boldsymbol{\alpha}^{\text{opt}}&=\widetilde{c}\mathbf{m}^H\mathbf{M}^{-1}=\frac{c\mathbf{m}^H\mathbf{M}^{-1}}{1+\mathbf{m}^H\mathbf{M}^{-1}\mathbf{m}}\\&\overset{(a)}{=}c\mathbf{m}^H\left(\mathbf{M}+\mathbf{m}\mathbf{m}^H\right)^{-1},\end{aligned} \tag{61}$$

where $c=\widetilde{c}(1+\mathbf{m}^H\mathbf{M}^{-1}\mathbf{m})$ and step (a) follows from Lemma 1. Thus we proved (18). Since (19) is convex, by setting the derivative of the optimization objective w.r.t. $\boldsymbol{\alpha}$ to be to 0, we can obtain that $\boldsymbol{\alpha}^{\text{opt}}$ is the optimal solution when $c=1$. Therefore, we complete the proof of Proposition 1. □

APPENDIX C
PROOF OF THEOREM 1

A. Proof of the Almost Sure Convergence

According to Lemma 1, we can write

$$\mathbf{D}_r=\mathbf{Q}\mathbf{D}_{h,0}\mathbf{L}\mathbf{F}, \tag{62}$$

where

$$\begin{aligned}&\mathbf{Q}=\operatorname{diag}(\mathbf{Q}_k;k\in[K]),\ \mathbf{Q}_k=\left(\frac{\widehat{\mathbf{H}}_k\widehat{\mathbf{H}}_k^H}{N_k}+\mathbf{Z}_k+\rho_k\mathbf{I}_{N_k}\right)^{-1},\\&\mathbf{L}=\operatorname{diag}(N_k^{-1};k\in[K]),\ \mathbf{D}_{h,0}=\operatorname{diag}(\widehat{\mathbf{h}}_{0k};k\in[K])\\&\mathbf{F}=\operatorname{diag}\left(\left(1+\frac{1}{N_k}\widehat{\mathbf{h}}_{0k}^H\mathbf{Q}_k\widehat{\mathbf{h}}_{0k}\right)^{-1};k\in[K]\right).\end{aligned} \tag{63}$$

We first analyze the LFSC case. By setting $a=1$, $\mathbf{m}=\mathbf{D}_r^H\widehat{\mathbf{h}}_0$, and $\mathbf{M}=\mathbf{D}_r^H(\widehat{\mathbf{H}}\widehat{\mathbf{H}}^H+\mathbf{D}_W+\sigma^2\mathbf{I}_N)\mathbf{D}_r$ in Lemma 1 and using (62), we can obtain

$$\gamma^{\text{lfsc}}=\frac{|\mathbf{g}^H\mathbf{G}_{\Delta_I}^{-1}\widetilde{\mathbf{g}}|^2}{\mathbf{g}^H\mathbf{G}_{\Delta_I}^{-1}\mathbf{G}_{\Delta}\mathbf{G}_{\Delta_I}^{-1}\mathbf{g}}, \tag{64}$$

where

$$\begin{aligned}\mathbf{g}^H&=\widehat{\mathbf{h}}_0^H\mathbf{Q}\mathbf{D}_{h,0}\mathbf{L},\ \widetilde{\mathbf{g}}^H=\widetilde{\mathbf{h}}_0^H\mathbf{Q}\mathbf{D}_{h,0}\mathbf{L},\\\mathbf{G}_{\Delta_I}&=\mathbf{L}\mathbf{D}_{h,0}^H\mathbf{Q}(\widehat{\mathbf{H}}\widehat{\mathbf{H}}^H+\mathbf{D}_W+\sigma^2\mathbf{I}_N)\mathbf{Q}\mathbf{D}_{h,0}\mathbf{L},\\\mathbf{G}_{\Delta}&=\mathbf{L}\mathbf{D}_{h,0}^H\mathbf{Q}(\widetilde{\mathbf{H}}\widetilde{\mathbf{H}}^H+\mathbf{W}+\sigma^2\mathbf{I}_N)\mathbf{Q}\mathbf{D}_{h,0}\mathbf{L}.\end{aligned} \tag{65}$$

It can be proved that $\|\mathbf{Q}\|=\sup_k\|\mathbf{Q}_k\|\leq\frac{1}{\min_k\rho_k}$ for $\rho_k>0$. According to Lemmas 2 and 3, we have

$$\begin{aligned}[\mathbf{g}]_k-\frac{\operatorname{Tr}[\mathbf{\Phi}_0]_{[k,k]}\mathbf{\Theta}_k}{N_k}&=\frac{\widehat{\mathbf{h}}_{0k}^H\mathbf{Q}_k\widehat{\mathbf{h}}_{0k}}{N_k}-\frac{\operatorname{Tr}[\mathbf{\Phi}_0]_{[k,k]}\mathbf{\Theta}_k}{N_k}\\&=\frac{\mathbf{q}_0^H(\mathbf{\Phi}_{0k}^{\frac{1}{2}})^H\mathbf{Q}_k\mathbf{\Phi}_{0k}^{\frac{1}{2}}\mathbf{q}_0}{N_k}-\frac{\operatorname{Tr}(\mathbf{\Phi}_{0k}^{\frac{1}{2}})^H\mathbf{Q}_k\mathbf{\Phi}_{0k}^{\frac{1}{2}}}{N_k}\end{aligned}$$

$$\begin{aligned}\overline{\mathbf{\Pi}}_{kl,B}(\boldsymbol{\mathcal{A}})=&\left[\widetilde{\boldsymbol{\lambda}}_{B,B,kl}(\boldsymbol{\mathcal{A}})-\widetilde{\boldsymbol{\lambda}}_{B,A,kl}(\boldsymbol{\mathcal{A}})\mathbf{D}_{A,B,l}\widetilde{\mathbf{F}}_l-\widetilde{\boldsymbol{\lambda}}_{A,B,kl}(\boldsymbol{\mathcal{A}})\mathbf{D}_{B,A,k}\widetilde{\mathbf{F}}_k\right]\mathbf{1}_M\\&+\widetilde{\boldsymbol{\lambda}}_{kl}(\boldsymbol{\mathcal{A}})\widetilde{\mathbf{F}}_k\widetilde{\mathbf{F}}_l\mathbf{\Xi}_{kl}\left[\mathbf{\Lambda}_{B,B,kl}-\mathbf{\Lambda}_{B,A,kl}\mathbf{D}_{A,B,l}\widetilde{\mathbf{F}}_l-\mathbf{\Lambda}_{A,B,kl}\mathbf{D}_{B,A,k}\widetilde{\mathbf{F}}_k+\mathbf{D}_{B,A,k}\mathbf{D}_{A,B,l}\right]\mathbf{1}_M.\end{aligned} \tag{56}$$

$$+\frac{\mathrm{Tr}[\mathbf{\Phi}_0]_{[k,k]}\mathbf{Q}_k}{N_k}-\frac{\mathrm{Tr}[\mathbf{\Phi}_0]_{[k,k]}\mathbf{\Theta}_k}{N_k}\xrightarrow[N\xrightarrow{c_1,\ldots,c_K}+\infty]{a.s.}0, \quad (66)$$

where $\mathbf{q}_0\sim\mathcal{CN}(0,\mathbf{I}_N)$ and $\mathbf{\Phi}_{0k}^{\frac{1}{2}}=\mathbf{\Phi}_0^{\frac{1}{2}}[\varsigma_k,:]\in\mathbb{C}^{N_k\times N}$. Since K is a given number, it holds true that $[\mathbf{g}]_k-[\mathbf{v}]_k$ converges to 0 almost surely for each $k\in[K]$. Similarly, we write

$$[\mathbf{G}_\Delta]_{k,l}=\frac{\widehat{\mathbf{h}}_{0k}^H\mathbf{Q}_k([\mathbf{W}]_{[k,l]}+\sigma^2\mathbb{I}_{\{k=l\}}\mathbf{I}_{N_k})\mathbf{Q}_l\widehat{\mathbf{h}}_{0l}}{N_kN_l}$$
$$+\frac{\widehat{\mathbf{h}}_{0k}^H\mathbf{Q}_k\widetilde{\mathbf{H}}_k\widetilde{\mathbf{H}}_l^H\mathbf{Q}_l\widehat{\mathbf{h}}_{0l}}{N_kN_l}=G_{kl,1}+G_{kl,2}. \quad (67)$$

According to [34, Theorem 2], we can conclude that the spectral norm of matrix $\frac{1}{\sqrt{N_k}}\widetilde{\mathbf{H}}_k$ is bounded with probability 1 for each $k\in[K]$ such that the following holds almost surely

$$\sup_{N\geq1}\|\frac{\mathbf{Q}_k\widetilde{\mathbf{H}}_k\widetilde{\mathbf{H}}_l^H\mathbf{Q}_l}{\sqrt{N_kN_l}}\|\leq\sup_{N\geq1}\frac{\|\frac{\widetilde{\mathbf{H}}_k}{\sqrt{N_k}}\|\|\frac{\widetilde{\mathbf{H}}_l}{\sqrt{N_l}}\|}{\min_k\rho_k^2}<+\infty, \quad (68)$$

for each $k,l\in[K]$. By Lemmas 2 and 3, the difference $\mathcal{D}_{G_{kl,2}}=G_{kl,2}-\frac{1}{\sqrt{N_kN_l}}\overline{\Pi}_{B,kl}([\mathbf{\Phi}_0]_{[l,k]})$ can be written as

$$\mathcal{D}_{G_{kl,2}}=G_{kl,2}-\frac{\mathrm{Tr}\mathbf{Q}_k\widetilde{\mathbf{H}}_k\widetilde{\mathbf{H}}_l^H\mathbf{Q}_l[\mathbf{\Phi}_0]_{[l,k]}}{N_kN_l}$$
$$+\frac{\mathrm{Tr}\mathbf{Q}_k\widetilde{\mathbf{H}}_k\widetilde{\mathbf{H}}_l^H\mathbf{Q}_l[\mathbf{\Phi}_0]_{[l,k]}}{N_kN_l}-\frac{\overline{\Pi}_{B,kl}([\mathbf{\Phi}_0]_{[l,k]})}{\sqrt{N_kN_l}}, \quad (69)$$

which converges to 0 almost surely. Following a similar derivation, we can prove $G_{kl,1}$ converges almost surely to $\frac{1}{N_kN_l}\overline{\Upsilon}_{kl}([\mathbf{\Phi}_0]_{[l,k]},[\mathbf{W}+\sigma^2\mathbf{I}_N]_{[k,l]})$. Moreover, it can be shown that,

$$[\widetilde{\mathbf{g}}]_k-[\mathbf{v}]_k\xrightarrow[N\xrightarrow{c_1,\ldots,c_K}+\infty]{a.s.}0, \quad (70)$$

$$[\mathbf{G}_{\Delta_I}]_{k,l}-[\mathbf{\Delta}_I]_{k,l}\xrightarrow[N\xrightarrow{c_1,\ldots,c_K}+\infty]{a.s.}0, \quad (71)$$

for each $k,l\in[K]$. According to the continuous mapping theorem [35], we have

$$\gamma^{\mathrm{lfsc}}\xrightarrow[N\xrightarrow{c_1,\ldots,c_K}+\infty]{a.s.}\overline{\gamma}^{\mathrm{lfsc}}, \quad (72)$$

which concludes (28b). The proof for (28a) an (28c) can be obtained similarly and is omitted here. □

B. Proof of the Approximation for the Mean

We analyze the LFOC case here and the other two cases are similar. By using the resolvent identity $\mathbf{A}-\mathbf{B}=\mathbf{A}(\mathbf{B}^{-1}-\mathbf{A}^{-1})\mathbf{B}$, we have

$$\mathbf{G}_\Delta^{-1}=\mathbf{\Delta}^{-1}+\mathbf{\Delta}^{-1}(\mathbf{\Delta}-\mathbf{G}_\Delta)\mathbf{\Delta}^{-1}$$
$$+\mathbf{G}_\Delta^{-1}(\mathbf{\Delta}-\mathbf{G}_\Delta)\mathbf{\Delta}^{-1}(\mathbf{\Delta}-\mathbf{G}_\Delta)\mathbf{\Delta}^{-1}. \quad (73)$$

According to the definition, γ^{lfoc} can be rewritten as (74) shown at the bottom of this page. It can be proved that $|\mathbb{E}X_j|$, $j=2,3,\ldots9$ is of order $\mathcal{O}(N^{-1})$ and we will evaluate $|\mathbb{E}X_6|$ in the following, as the remaining terms can be addressed using a similar method. To this end, we first prove the following bounds

$$\mathbb{E}|[\widetilde{\mathbf{g}}-\mathbf{v}]_k|^2=\mathcal{O}\left(\frac{1}{N}\right),\forall k\in[K] \quad (75)$$

$$\mathbb{E}|[\mathbf{G}_\Delta-\mathbf{\Delta}]_{k,l}|^2=\mathcal{O}\left(\frac{1}{N}\right),\forall k,l\in[K]. \quad (76)$$

Denote $\mathscr{F}_M$ as the σ-algebra generated by $\mathbf{h}_1,\ldots,\mathbf{h}_M$ and

$$\underline{\mathbf{g}}_k=\mathbb{E}([\widetilde{\mathbf{g}}]_k|\mathscr{F}_M)=\frac{1}{N_k}\mathrm{Tr}[\mathbf{\Phi}_0]_{[k,k]}\mathbf{Q}_k. \quad (77)$$

According to the triangle inequality, we have

$$\mathbb{E}|[\widetilde{\mathbf{g}}-\mathbf{v}]_k|^2\leq C(\mathbb{E}|[\widetilde{\mathbf{g}}]_k-\underline{\mathbf{g}}_k|^2+\mathbb{E}|\underline{\mathbf{g}}_k-\mathbb{E}[\underline{\mathbf{g}}]_k|^2$$
$$+|\mathbb{E}[\underline{\mathbf{g}}]_k-[\mathbf{v}]_k|^2):=C(W_1+W_2+W_3), \quad (78)$$

where C is a constant independent of N. By applying the trace lemma [36, Lemma 2.7], we have

$$W_1=\frac{\mathbb{E}\left[\mathbb{E}\left(|\widehat{\mathbf{h}}_{0k}^H\mathbf{Q}_k\widetilde{\mathbf{h}}_{0k}-\mathrm{Tr}\mathbf{Q}_k[\mathbf{\Phi}_0]_{[k,k]}|^2\big|\mathscr{F}_M\right)\right]}{N_k^2}$$
$$\leq\frac{C_1}{N_k}\mathbb{E}\|\mathbf{Q}_k\|^2=\mathcal{O}\left(\frac{1}{N}\right), \quad (79)$$

where C_1 is a constant. By using Poincaré-Nash inequality [37, Eq. (18)], we can get

$$W_2=\mathrm{Var}(\underline{\mathbf{g}}_k)\leq\mathcal{O}\left(\frac{1}{N^2}\right). \quad (80)$$

Finally, by Lemma 3, we have $W_3\leq\mathcal{O}(\frac{1}{N^3})$. Therefore, the dominant term on the RHS of (78) is $\mathcal{O}(N^{-1})$ which proves (75). The bound (76) can be proved in a similar manner. Next, we evaluate the term $|\mathbb{E}X_6|$. By applying (75) and (76), bound (81) shown at the bottom of the next page can be derived, where step (a) follows from Cauchy-Schwartz inequality $\mathbb{E}|XY|\leq\mathbb{E}^{\frac{1}{2}}|X|^2\mathbb{E}^{\frac{1}{2}}|Y|^2$. Therefore, Theorem 1 is proved. □

$$\gamma^{\mathrm{lfoc}}=\mathbf{v}^H\mathbf{\Delta}^{-1}\mathbf{v}+\mathbf{v}^H\mathbf{\Delta}^{-1}(\widetilde{\mathbf{g}}-\mathbf{v})+(\widetilde{\mathbf{g}}-\mathbf{v})^H\mathbf{\Delta}^{-1}\mathbf{v}+(\widetilde{\mathbf{g}}-\mathbf{v})^H\mathbf{\Delta}^{-1}(\widetilde{\mathbf{g}}-\mathbf{v})+\mathbf{v}^H\mathbf{\Delta}^{-1}(\mathbf{\Delta}-\mathbf{G}_\Delta)\mathbf{\Delta}^{-1}\mathbf{v}$$
$$+\mathbf{v}^H\mathbf{\Delta}^{-1}(\mathbf{\Delta}-\mathbf{G}_\Delta)\mathbf{\Delta}^{-1}(\widetilde{\mathbf{g}}-\mathbf{v})+(\widetilde{\mathbf{g}}-\mathbf{v})^H\mathbf{\Delta}^{-1}(\mathbf{\Delta}-\mathbf{G}_\Delta)\mathbf{\Delta}^{-1}\mathbf{v}+(\widetilde{\mathbf{g}}-\mathbf{v})^H\mathbf{\Delta}^{-1}(\mathbf{\Delta}-\mathbf{G}_\Delta)\mathbf{\Delta}^{-1}(\widetilde{\mathbf{g}}-\mathbf{v})$$
$$+\widetilde{\mathbf{g}}^H\mathbf{G}_\Delta^{-1}(\mathbf{\Delta}-\mathbf{G}_\Delta)\mathbf{\Delta}^{-1}(\mathbf{\Delta}-\mathbf{G}_\Delta)\mathbf{\Delta}^{-1}\widetilde{\mathbf{g}}:=\sum_{j=1}^{9}X_j. \quad (74)$$

APPENDIX D
PROOF OF COROLLARY 1

Define $\mathbf{R}_j^{\mathcal{K}_g}$, $\mathbf{D}_{\widetilde{\sigma}}^{\mathcal{K}_g}$, $\mathbf{D}_{R,j}^{\mathcal{K}_g}$, $\mathbf{T}_j^{\mathcal{K}_g}$, $\mathbf{D}_{T,j}^{\mathcal{K}_g}$, $\mathbf{\Phi}_j^{\mathcal{K}_g}$, $\mathbf{V}_j^{\mathcal{K}_g}$, $\mathbf{W}_j^{\mathcal{K}_g}$, and $\mathbf{W}^{\mathcal{K}_g}$ as the matrices obtained by considering the channel $\mathbf{h}_j^{\mathcal{K}_g}$ in (11) and (12) for each $j \in [M]_0$. Denote the asymptotic regime $\widetilde{\sigma}_e^2 \to +\infty$ as $\widetilde{\sigma}_k^2 \to +\infty$ for each $k \in \mathcal{K}_e$ with equal value $\widetilde{\sigma}_k^2 = \widetilde{\sigma}_e^2$. To prove Corollary 1, it is sufficient to show

1) $[\mathbf{R}^{\mathcal{K}_g}]_{[a,b]} \to [\mathbf{R}]_{[k_a,k_b]}$, for $a, b \in [G]$ as $\widetilde{\sigma}_e^2 \to +\infty$. Here, $\mathbf{R}$ is one of

$$\left\{\mathbf{\Phi}_j, \mathbf{V}_j\mathbf{\Phi}_j, \mathbf{V}_j\mathbf{\Phi}_j\mathbf{V}_j^H\right\}_{j=0}^{M} \cup \{\mathbf{W}, \mathbf{D}_W\}. \tag{82}$$

2) $[\mathbf{\Phi}_j]_{[x,y]} \to 0$ as $\widetilde{\sigma}_e^2 \to +\infty$, for one of x or y that belongs to $\mathcal{K}_e$.

Here, $[\mathbf{R}_j^{\mathcal{K}_g}]_{[a,b]}$ is the submatrix obtained by partitioning the rows and columns based on $\mathbf{h}_j^{\mathcal{K}_g}$, with a size of $k_a \times k_b$. Thus, we have $[\mathbf{R}_j^{\mathcal{K}_g}]_{[a,b]} = [\mathbf{R}_j]_{[k_a,k_b]}$ and $[\mathbf{D}_{\widetilde{\sigma}}^{\mathcal{K}_g}]_{[a,a]} = \widetilde{\sigma}_{k_a}^2\mathbf{I}_{N_{k_a}}$ for each $a, b \in [G]$.

We prove item 1) first. We only prove the case with $\mathbf{R} = \mathbf{W}$ as the proofs for the other terms are similar. For $a, b \in [G]$, we have

$$\begin{aligned}[\mathbf{W}_j]_{[k_a,k_b]} &= [\mathbf{R}_j(\mathbf{D}_{\widetilde{\sigma}} + \mathbf{R}_j)^{-1}\mathbf{D}_{\widetilde{\sigma}}]_{[k_a,k_b]} \\ &= \widetilde{\sigma}_{k_b}^2[\mathbf{R}_j(\mathbf{D}_{\widetilde{\sigma}} + \mathbf{R}_j)^{-1}]_{[k_a,k_b]}.\end{aligned} \tag{83}$$

Without loss of generality, assume $k_a = a$ for $a \in [G]$ and write

$$\mathbf{R}_j = \begin{bmatrix}\mathbf{R}_j^{\mathcal{K}_g} & \mathbf{R}_{j,1} \\ \mathbf{R}_{j,2} & \mathbf{R}_{j,3}\end{bmatrix}, \mathbf{D}_{\widetilde{\sigma}} = \begin{bmatrix}\mathbf{D}_{\widetilde{\sigma}}^{\mathcal{K}_g} & \mathbf{0} \\ \mathbf{0} & \mathbf{D}_{\widetilde{\sigma},1}\end{bmatrix}. \tag{84}$$

According to the block matrix inversion formula [38], we can get

$$(\mathbf{D}_{\widetilde{\sigma}} + \mathbf{R}_j)^{-1} = \begin{bmatrix}\mathbf{P}_1 & \mathbf{P}_2 \\ \mathbf{P}_3 & \mathbf{P}_4\end{bmatrix}, \tag{85}$$

where terms $\mathbf{P}_t$, $t = 2, 3, 4$ approach $\mathbf{0}$ as $\widetilde{\sigma}_e^2 \to +\infty$, and the term

$$\begin{aligned}\mathbf{P}_1 &= \left(\mathbf{R}_j^{\mathcal{K}_g} + \mathbf{D}_{\widetilde{\sigma}}^{\mathcal{K}_g} - \mathbf{R}_{j,2}(\mathbf{R}_{j,3} + \mathbf{D}_{\widetilde{\sigma},1})^{-1}\mathbf{R}_{j,1}\right)^{-1} \\ &\to \left(\mathbf{R}_j^{\mathcal{K}_g} + \mathbf{D}_{\widetilde{\sigma}}^{\mathcal{K}_g}\right)^{-1}, \text{ as } \widetilde{\sigma}_e^2 \to +\infty.\end{aligned} \tag{86}$$

Hence, by (83), the following holds

$$\lim_{\widetilde{\sigma}_e^2 \to +\infty}[\mathbf{W}_j]_{[k_a,k_b]} = \widetilde{\sigma}_{k_b}^2\left[\begin{bmatrix}\mathbf{R}_j^{\mathcal{K}_g}(\mathbf{R}_j^{\mathcal{K}_g} + \mathbf{D}_{\widetilde{\sigma}}^{\mathcal{K}_g})^{-1} & \mathbf{0} \\ \mathbf{R}_{j,2}(\mathbf{R}_j^{\mathcal{K}_g} + \mathbf{D}_{\widetilde{\sigma}}^{\mathcal{K}_g})^{-1} & \mathbf{0}\end{bmatrix}\right]_{[k_a,k_b]}$$

$$= \left[\mathbf{R}_j^{\mathcal{K}_g}(\mathbf{R}_j^{\mathcal{K}_g} + \mathbf{D}_{\widetilde{\sigma}}^{\mathcal{K}_g})^{-1}\right]_{[a,b]}\left[\mathbf{D}_{\widetilde{\sigma}}^{\mathcal{K}_g}\right]_{[b,b]} = \left[\mathbf{W}_j^{\mathcal{K}_g}\right]_{[a,b]}. \tag{87}$$

By summing over the index j in above equation, we get $\lim_{\widetilde{\sigma}_e^2 \to +\infty}[\mathbf{W}]_{[k_a,k_b]} = [\mathbf{W}^{\mathcal{K}_g}]_{[a,b]}$. Therefore, item 1) is proved.

Next, we prove item 2). Without loss of generality, we assume $x \in \mathcal{K}_e$. As a result,

$$[\mathbf{D}_{T,j}]_{[x,x]} = [\mathbf{R}_j]_{[x,x]}(\widetilde{\sigma}_x^2\mathbf{I}_{N_x} + [\mathbf{R}_j]_{[x,x]})^{-1} \to \mathbf{0}, \tag{88}$$

as $\widetilde{\sigma}_e^2 \to +\infty$. Hence, we have

$$[\mathbf{\Phi}_j]_{[x,y]} = [\mathbf{D}_{T,j}]_{[x,x]}[(\mathbf{D}_{\widetilde{\sigma}} + \mathbf{R}_j)\mathbf{D}_{T,j}]_{[x,y]} \to \mathbf{0}. \tag{89}$$

Therefore, we complete the proof of Corollary 1. □

APPENDIX E
PROOF OF LEMMA 3

The proof relies on Gaussian tools [19], [37], [39], [40], [41], which consist of the integration by parts formula [37, Eq. (17)] and the Poincaré-Nash inequality [37, Eq. (18)]. According to the Poincaré–Nash inequality, the variance of functions defined in Table II can be proved of order $\mathcal{O}(1)$ (See [34, Lemma 9] for the case of F_k and the proof of other functions are similar). As a result, for given $\epsilon > 0$, we have

$$\sum_{N=1}^{+\infty}\mathbb{P}\left(\frac{|\mathcal{X} - \mathbb{E}\mathcal{X}|}{N} \geq \epsilon\right) \leq \sum_{N=1}^{+\infty}\frac{\mathrm{Var}(\mathcal{X})}{\epsilon^2 N^2} < +\infty, \tag{90}$$

where $\mathcal{X}$ can be any function in Table II. Thus, according to Borel–Cantelli lemma, we have

$$\frac{\mathcal{X} - \mathbb{E}\mathcal{X}}{N} \xrightarrow[N\xrightarrow{c_1,\ldots,c_K}+\infty]{a.s.} 0. \tag{91}$$

Therefore, to show the stochastic equivalence, we only need to evaluate the means of functions in Table II.

A. Evaluation of $\mathbb{E}F_k$

Before delving into the details of the proof, we first define the following quantities

$$\begin{aligned}&\alpha_{jk} = \frac{1}{N_k}\mathrm{Tr}[\mathbf{\Omega}_j]_{[k,k]}\mathbf{Q}_k, \ \mathbf{D}_{\alpha,k} = \mathrm{diag}(\alpha_{jk}; j \in [M]), \\ &\widetilde{\mathbf{F}}_{\alpha,k} = (\mathbf{I} + \mathbb{E}\mathbf{D}_{\alpha,k})^{-1}, \\ &\mathbf{\Theta}_{\alpha,k} = \left(-z_k\mathbf{I}_{N_k} + \mathbf{S}_k + \frac{1}{N_k}\sum_{j=1}^{M}\frac{[\mathbf{\Omega}_j]_{[k,k]}}{1 + \underline{\alpha}_{jk}}\right)^{-1}.\end{aligned} \tag{92}$$

$$|\mathbb{E}X_6| = |\mathbb{E}\mathbf{v}^H\mathbf{\Delta}^{-1}(\mathbf{\Delta} - \mathbf{G}_\Delta)\mathbf{\Delta}^{-1}(\widetilde{\mathbf{g}} - \mathbf{v})| = |\mathbb{E}\sum_{i_1,i_2,i_3,i_4\in[K]}[\mathbf{v}^H\mathbf{\Delta}^{-1}]_{i_1}[\mathbf{\Delta}^{-1}]_{i_1,i_2}[\mathbf{\Delta} - \mathbf{G}_\Delta]_{i_2,i_3}[\mathbf{\Delta}^{-1}]_{i_3,i_4}[\widetilde{\mathbf{g}} - \mathbf{v}]_{i_4}|$$

$$\overset{(a)}{\leq}\sum_{i_1,i_2,i_3,i_4\in[K]}|[\mathbf{v}^H\mathbf{\Delta}^{-1}]_{i_1}[\mathbf{\Delta}^{-1}]_{i_1,i_2}[\mathbf{\Delta}^{-1}]_{i_3,i_4}|\mathbb{E}^{\frac{1}{2}}|[\mathbf{\Delta} - \mathbf{G}_\Delta]_{i_2,i_3}|^2\mathbb{E}^{\frac{1}{2}}|[\widetilde{\mathbf{g}} - \mathbf{v}]_{i_4}|^2 = \mathcal{O}\left(\frac{1}{N}\right). \tag{81}$$

These quantities will serve as intermediate variables in the proof, i.e., we will show $\mathbf{Q}_k \approx \mathbf{\Theta}_{\alpha,k} \approx \mathbf{\Theta}_k$.

Applying integration by parts formula to $\mathbb{E}[\mathbf{Q}_k]_{i,p}[\mathbf{x}_s]_a[\mathbf{Y}_l]^*_{j,s}$ and taking $[\mathbf{x}_s]_a$ as the variable, we have

$$\begin{aligned}\mathbb{E}[\mathbf{Q}_k]_{i,p}[\mathbf{x}_s]_a[\mathbf{Y}_l]^*_{j,s} &= \mathbb{E}\frac{\partial[\mathbf{Q}_k]_{i,p}[\mathbf{Y}_l]^*_{j,s}}{\partial[\mathbf{x}_s]^*_a} \\ = \mathbb{E} - \frac{1}{\sqrt{N_k}}&[\mathbf{Q}_k\mathbf{X}_k]_{i,s}\left[\mathbf{A}^H_{sk}\mathbf{Q}_k\right]_{a,p}[\mathbf{Y}_l]^*_{j,s} \\ + \mathbb{E}\frac{1}{\sqrt{N_l}}&\left[\mathbf{Q}_k\right]_{i,p}[\mathbf{B}^H_{sl}]_{a,j}. \end{aligned} \tag{93}$$

Multiplying $\frac{1}{\sqrt{N_k}}[\mathbf{A}_{sk}]_{p,a}$ at both sides of (93) and summing over the subscripts p and a, the following holds

$$\begin{aligned}\mathbb{E}[\mathbf{Q}_k\mathbf{X}_k]_{i,s}[\mathbf{Y}_l]^*_{j,s} &= \mathbb{E}\frac{[\mathbf{Q}_k\mathbf{A}_{sk}\mathbf{B}^H_{sl}]_{i,j}}{\sqrt{N_kN_l}} \\ - \mathbb{E}\alpha_{sk}&[\mathbf{Q}_k\mathbf{X}_k]_{i,s}[\mathbf{Y}_l]^*_{j,s}. \end{aligned} \tag{94}$$

By writing the random variable α_{sk} as the sum of its mean and its centered version, i.e., $\alpha_{sk} = \underline{\alpha}_{sk} + \mathring{\alpha}_{sk}$ and solving (94) w.r.t. $\mathbb{E}[\mathbf{Q}_k\mathbf{X}_k]_{i,s}[\mathbf{Y}_l]^*_{j,s}$, we get

$$\begin{aligned}\mathbb{E}[\mathbf{Q}_k\mathbf{X}_k]_{i,s}[\mathbf{Y}_l]^*_{j,s} &= \mathbb{E}\frac{[\mathbf{Q}_k\mathbf{A}_{sk}\mathbf{B}^H_{sl}]_{i,j}}{\sqrt{N_kN_l}(1+\underline{\alpha}_{sk})} \\ - \mathbb{E}&\frac{\mathring{\alpha}_{sk}}{1+\underline{\alpha}_{sk}}[\mathbf{Q}_k\mathbf{X}_k]_{i,s}[\mathbf{Y}_l]^*_{j,s}. \end{aligned} \tag{95}$$

Equation (95) will also be used for the evaluation of $\mathbb{E}\Phi_{B,kl}$ later. To obtain the approximation of the trace of the resolvent, we set $l = k$ and $\mathbf{Y}_k = \mathbf{X}_k$, i.e., $\mathbf{B}_s = \mathbf{A}_s$, $\forall s \in [M]$. By summing over the subscript s, we can obtain

$$\begin{aligned}\mathbb{E}\left[\mathbf{Q}_k\mathbf{X}_k\mathbf{X}^H_k\right]_{i,j} &= \mathbb{E}\Big[\mathbf{Q}_k\sum_{s=1}^{M}\frac{[\mathbf{\Omega}_s]_{[k,k]}}{N_k(1+\underline{\alpha}_{sk})}\Big]_{i,j} \\ - \mathbb{E}&\left[\mathbf{Q}_k\mathbf{X}_k\mathring{\mathbf{D}}_{\alpha,k}\widetilde{\mathbf{F}}_{\alpha,k}\mathbf{X}^H_k\right]_{i,j}. \end{aligned} \tag{96}$$

By using the resolvent identity $\mathbf{Q}_k\mathbf{X}_k\mathbf{X}^H_k = z_k\mathbf{Q}_k - \mathbf{Q}_k\mathbf{S}_k + \mathbf{I}_{N_k}$ on $\mathbb{E}[\mathbf{Q}_k\mathbf{X}_k\mathbf{X}^H_k]_{i,j}$, multiplying both sides $[\mathbf{\Theta}_{\alpha,k}\boldsymbol{\mathcal{A}}]_{j,i}$, and summing over the subscripts i and j, we can get

$$\begin{aligned}&\mathbb{E}\mathrm{Tr}\boldsymbol{\mathcal{A}}\mathbf{Q}_k - \mathrm{Tr}\boldsymbol{\mathcal{A}}\mathbf{\Theta}_{\alpha,k} \\ &= \mathbb{E}\mathrm{Tr}\mathbf{Q}_k\mathbf{X}_k\mathring{\mathbf{D}}_{\alpha,k}\widetilde{\mathbf{F}}_{\alpha,k}\mathbf{X}^H_k\mathbf{\Theta}_{\alpha,k}\boldsymbol{\mathcal{A}} := \varepsilon_k(\boldsymbol{\mathcal{A}}). \end{aligned} \tag{97}$$

We can write the residual term $\varepsilon_k(\boldsymbol{\mathcal{A}})$ as

$$\begin{aligned}\varepsilon_k(\boldsymbol{\mathcal{A}}) &= \sum_{j=1}^{M}\mathbb{E}\mathring{\alpha}_{jk}\underbrace{[\widetilde{\mathbf{F}}_{\alpha,k}\mathbf{X}^H_k\mathbf{\Theta}_{\alpha,k}\boldsymbol{\mathcal{A}}\mathbf{Q}_k\mathbf{X}_k]_{j,j}}_{:=\chi_{jk}} \\ &\overset{(a)}{\le} \sum_{j=1}^{M}\mathrm{Var}^{\frac{1}{2}}(\alpha_{jk})\mathrm{Var}^{\frac{1}{2}}(\chi_{jk}) \overset{(b)}{=} \mathcal{O}\left(\frac{1}{N^{\frac{1}{2}}}\right), \end{aligned} \tag{98}$$

where step (a) follows from Cauchy-Schwartz inequality. Step (b) follows from the fact $\mathrm{Var}(\alpha_{jk}) = \mathcal{O}(\frac{1}{N^2})$ and $\mathrm{Var}(\chi_{jk}) = \mathcal{O}(\frac{1}{N})$, which can be proved by Poincaré-Nash inequality. Hence, we build the relation between $\mathbf{Q}_k$ and $\mathbf{\Theta}_{\alpha,k}$.

To establish the relationship between $\mathbf{\Theta}_{\alpha,k}$ and $\mathbf{\Theta}_k$, we take the difference between $\frac{1}{N_k}\mathrm{Tr}[\mathbf{\Omega}_j]_{[k,k]}\mathbf{\Theta}_{\alpha,k}$ and δ_{jk} for each $j \in [M]$. Following the same procedure as in the proof of [34, Proposition 3], we can show that $|\mathrm{Tr}\boldsymbol{\mathcal{A}}\mathbf{\Theta}_k - \mathrm{Tr}\boldsymbol{\mathcal{A}}\mathbf{\Theta}_{\alpha,k}| = \mathcal{O}(N^{-\frac{1}{2}})$, which concludes (53).

B. Evaluation of $\mathbb{E}\Phi_{B,kl}$

Multiplying $[\mathbf{b}]_s$ and $[\boldsymbol{\mathcal{A}}]_{j,i}$ at both sides of (95) and summing over the subscripts s, i, and j, we can obtain

$$\begin{aligned}\mathbb{E}\Phi_{B,kl}(\boldsymbol{\mathcal{A}},\mathbf{b}) &= -\mathbb{E}[\mathbf{Q}_k\mathbf{X}_k\mathrm{diag}(\mathbf{b})\mathring{\mathbf{D}}_{\alpha,k}\widetilde{\mathbf{F}}_{\alpha,k}\mathbf{Y}^H_l] \\ &+ \sum_s\frac{[\mathbf{b}]_s\mathrm{Tr}\boldsymbol{\mathcal{A}}(\mathbb{E}\mathbf{Q}_k)\mathbf{A}_{sk}\mathbf{B}^H_{sl}}{\sqrt{N_kN_l}(1+\underline{\alpha}_{sk})} \\ &\overset{(a)}{=} \sum_s\frac{[\mathbf{b}]_s\mathrm{Tr}\boldsymbol{\mathcal{A}}\mathbf{\Theta}_k\mathbf{A}_{sk}\mathbf{B}^H_{sl}}{\sqrt{N_kN_l}(1+\delta_{sk})} + \mathcal{O}(N^{-\frac{1}{2}}), \end{aligned} \tag{99}$$

where step (a) follows from (53) and the variance control.

C. Evaluation of $\mathbb{E}\Upsilon_{kl}$

According to integration by parts formula, we get

$$\begin{aligned}\mathbb{E}[\mathbf{Q}_k\boldsymbol{\mathcal{B}}\mathbf{Q}_l\mathbf{X}_l]_{i,s}[\mathbf{X}_l]^*_{j,s} &= \mathbb{E}\frac{[\mathbf{Q}_k\boldsymbol{\mathcal{B}}\mathbf{Q}_l[\mathbf{\Omega}_s]_{[l,l]}]_{i,j}}{N_l(1+\underline{\alpha}_{sl})} \\ - \mathbb{E}&\frac{\mathring{\alpha}_{sl}}{1+\underline{\alpha}_{sl}}[\mathbf{Q}_k\boldsymbol{\mathcal{B}}\mathbf{Q}_l\mathbf{X}_l]_{i,s}[\mathbf{X}_l]^*_{j,s} \\ - &\frac{\mathbb{E}\Upsilon_{kl}([\mathbf{\Omega}_s]_{[l,k]},\boldsymbol{\mathcal{B}})[\mathbf{Q}_k\mathbf{X}_k]_{i,s}[\mathbf{X}_l]^*_{j,s}}{\sqrt{N_lN_k}(1+\underline{\alpha}_{sl})}. \end{aligned} \tag{100}$$

Summing over the subscript s and using the resolvent identity $\mathbf{Q}_l\mathbf{X}_l\mathbf{X}^H_l = z_l\mathbf{Q}_l - \mathbf{Q}_l\mathbf{S}_l + \mathbf{I}_{N_l}$, the following holds

$$\begin{aligned}\mathbb{E}[\mathbf{Q}_k\boldsymbol{\mathcal{B}}\mathbf{Q}_l\mathbf{\Theta}^{-1}_{\alpha,l}]_{i,j} &= \mathbb{E}[\mathbf{Q}_k\boldsymbol{\mathcal{B}}]_{i,j} \\ + \mathbb{E}[\mathbf{Q}_k\mathbf{X}_k\mathbf{D}_{\Upsilon,kl}(\boldsymbol{\mathcal{B}})\widetilde{\mathbf{F}}_{\alpha,l}\mathbf{X}^H_l]_{i,j} &+ [\boldsymbol{\varepsilon}_{kl}]_{i,j}, \end{aligned} \tag{101}$$

where $\mathbf{D}_{\Upsilon,kl}(\boldsymbol{\mathcal{B}}) = \frac{1}{\sqrt{N_kN_l}}\mathrm{diag}(\Upsilon_{kl}([\mathbf{\Omega}_s]_{[l,k]},\boldsymbol{\mathcal{B}});s\in[M])$ and $\boldsymbol{\varepsilon}_{kl}$ is the residual matrix. It can be proved by Poincaré–Nash inequality that $\mathrm{Tr}\boldsymbol{\varepsilon}\boldsymbol{\mathcal{C}} = \mathcal{O}(N^{-\frac{1}{2}})$ for any $\boldsymbol{\mathcal{C}}$ with bounded spectral norm. Thus, we have

$$\begin{aligned}\mathbb{E}\Upsilon_{kl}(\boldsymbol{\mathcal{A}},\boldsymbol{\mathcal{B}}) &= \mathrm{Tr}\boldsymbol{\mathcal{A}}(\mathbb{E}\mathbf{Q}_k)\boldsymbol{\mathcal{B}}\mathbf{\Theta}_{\alpha,l} \\ + \mathbb{E}\mathrm{Tr}\mathbf{Q}_k\mathbf{X}_k\mathbf{D}_{\Upsilon,kl}(\boldsymbol{\mathcal{B}})\widetilde{\mathbf{F}}_{\alpha,l}\mathbf{X}^H_l\mathbf{\Theta}_{\alpha,l}\boldsymbol{\mathcal{A}} &+ \mathcal{O}(N^{-\frac{1}{2}}). \end{aligned} \tag{102}$$

By replacing $\mathbf{\Theta}_{\alpha,l}$ and $\widetilde{\mathbf{F}}_{\alpha,l}$ with $\mathbf{\Theta}_l$ and $\widetilde{\mathbf{F}}_l$, respectively, setting $\boldsymbol{\mathcal{A}} = [\mathbf{\Omega}_u]_{[l,k]}$ and $\boldsymbol{\mathcal{B}} = [\mathbf{\Omega}_v]_{[k,l]}$, and using the approximation rules of $\Phi_{B,kl}$ in (54), we have

$$\begin{aligned}[\mathbf{\Upsilon}_{\Omega,kl}]_{u,v} &= N_kN_l[\mathbf{\Gamma}_{kl}]_{u,v} \\ + \sum_{j=1}^{M}&[\mathbf{\Upsilon}_{\Omega,kl}]_{j,v}[\widetilde{\mathbf{F}}_k\widetilde{\mathbf{F}}_l]_{j,j}[\mathbf{\Gamma}_{kl}]_{u,j} + \mathcal{O}(N^{-\frac{1}{2}}), \end{aligned} \tag{103}$$

where $[\mathbf{\Upsilon}_{\Omega,kl}]_{u,v} = \mathbb{E}\mathrm{Tr}[\mathbf{\Omega}_u]_{[l,k]}\mathbf{Q}_k[\mathbf{\Omega}_v]_{[k,l]}\mathbf{Q}_l$. Solving (103) w.r.t. $\mathbf{\Upsilon}_{\Omega,kl}$, we get

$$\mathbf{\Upsilon}_{\Omega,kl} = N_kN_l\mathbf{\Xi}_{kl}\mathbf{\Gamma}_{kl} + \mathcal{O}(N^{-\frac{1}{2}})\mathbf{1}_M\mathbf{1}^T_M. \tag{104}$$

$$\mathbb{E}\Pi_{B,kl}(\boldsymbol{\mathcal{A}}) = \sum_{i=1}^{M}\Big\{\Big[\widetilde{\boldsymbol{\lambda}}_{B,B,kl}(\boldsymbol{\mathcal{A}}) + \widetilde{\boldsymbol{\lambda}}_{kl}(\boldsymbol{\mathcal{A}})\widetilde{\mathbf{F}}_k\widetilde{\mathbf{F}}_l\boldsymbol{\Xi}_{kl}\boldsymbol{\Lambda}_{B,B,kl} - [\mathbf{D}_{A,B,l}\widetilde{\mathbf{F}}_l]_{i,i}\left(\widetilde{\boldsymbol{\lambda}}_{B,A,kl}(\boldsymbol{\mathcal{A}}) + \widetilde{\boldsymbol{\lambda}}_{kl}(\boldsymbol{\mathcal{A}})\widetilde{\mathbf{F}}_k\widetilde{\mathbf{F}}_l\boldsymbol{\Xi}_{kl}\boldsymbol{\Lambda}_{B,A,kl}\right)$$
$$-[\mathbf{D}_{B,A,k}\widetilde{\mathbf{F}}_k]_{i,i}\left(\widetilde{\boldsymbol{\lambda}}_{A,B,kl}(\boldsymbol{\mathcal{A}}) + \widetilde{\boldsymbol{\lambda}}_{kl}(\boldsymbol{\mathcal{A}})\widetilde{\mathbf{F}}_k\widetilde{\mathbf{F}}_l\boldsymbol{\Xi}_{kl}\boldsymbol{\Lambda}_{A,B,kl}\right)\Big]\mathbf{e}_i + [\mathbf{D}_{B,A,k}\mathbf{D}_{A,B,l}\widetilde{\mathbf{F}}_k\widetilde{\mathbf{F}}_l]_{i,i}[\widetilde{\boldsymbol{\lambda}}_{kl}(\boldsymbol{\mathcal{A}})\widetilde{\boldsymbol{\Xi}}_{kl}]_i\Big\} + \mathcal{O}(N^{-\frac{1}{2}})$$
$$\overset{(a)}{=}\Big(\widetilde{\boldsymbol{\lambda}}_{B,B,kl}(\boldsymbol{\mathcal{A}}) + \widetilde{\boldsymbol{\lambda}}_{kl}(\boldsymbol{\mathcal{A}})\widetilde{\mathbf{F}}_k\widetilde{\mathbf{F}}_l\boldsymbol{\Xi}_{kl}\boldsymbol{\Lambda}_{B,B,kl} - \widetilde{\boldsymbol{\lambda}}_{B,A,kl}(\boldsymbol{\mathcal{A}})\mathbf{D}_{A,B,l}\widetilde{\mathbf{F}}_l - \widetilde{\boldsymbol{\lambda}}_{kl}(\boldsymbol{\mathcal{A}})\widetilde{\mathbf{F}}_k\widetilde{\mathbf{F}}_l\boldsymbol{\Xi}_{kl}\boldsymbol{\Lambda}_{B,A,kl}\mathbf{D}_{A,B,l}\widetilde{\mathbf{F}}_l$$
$$-\widetilde{\boldsymbol{\lambda}}_{A,B,kl}(\boldsymbol{\mathcal{A}})\mathbf{D}_{B,A,k}\widetilde{\mathbf{F}}_k - \widetilde{\boldsymbol{\lambda}}_{kl}(\boldsymbol{\mathcal{A}})\widetilde{\mathbf{F}}_k\widetilde{\mathbf{F}}_l\boldsymbol{\Xi}_{kl}\boldsymbol{\Lambda}_{A,B,kl}\mathbf{D}_{B,A,k}\widetilde{\mathbf{F}}_k + \widetilde{\boldsymbol{\lambda}}_{kl}(\boldsymbol{\mathcal{A}})\widetilde{\mathbf{F}}_k\widetilde{\mathbf{F}}_l\boldsymbol{\Xi}_{kl}\mathbf{D}_{B,A,k}\mathbf{D}_{A,B,l}\Big)\mathbf{1}_M + \mathcal{O}(N^{-\frac{1}{2}}). \quad (109)$$

By following the similar method as in [34], we can show that $\mathbf{I}_M - \boldsymbol{\Gamma}_{kl}\widetilde{\mathbf{F}}_k\widetilde{\mathbf{F}}_l$ is invertible and $\max_{i\in[M]}\sum_{j=1}^{M}|[\boldsymbol{\Xi}_{kl}]_{i,j}| = \mathcal{O}(1)$. Due to the space limitation, we omit the details. Similar to the evaluation of $\boldsymbol{\Upsilon}_{\Omega,kl}$, we can derive the approximation rules for $\mathbb{E}\Upsilon_{kl}(\boldsymbol{\mathcal{A}}, [\boldsymbol{\Omega}_v]_{[k,l]})$ and $\mathbb{E}\Upsilon_{kl}([\boldsymbol{\Omega}_u]_{[l,k]}, \boldsymbol{\mathcal{B}})$. Plugging these results in (102) yields

$$\mathbb{E}\mathrm{Tr}\boldsymbol{\mathcal{A}}\mathbf{Q}_k\boldsymbol{\mathcal{B}}\mathbf{Q}_l = \mathrm{Tr}\boldsymbol{\mathcal{A}}\boldsymbol{\Theta}_k\boldsymbol{\mathcal{B}}\boldsymbol{\Theta}_l + \widetilde{\boldsymbol{\lambda}}_{kl}(\boldsymbol{\mathcal{A}})\widetilde{\mathbf{F}}_k\widetilde{\mathbf{F}}_l\boldsymbol{\Xi}_{kl}\boldsymbol{\lambda}_{kl}(\boldsymbol{\mathcal{B}}) + \mathcal{O}(N^{-\frac{1}{2}}), \quad (105)$$

which proves (55).

D. Evaluation of $\mathbb{E}\Pi_{B,kl}$

By applying integration by parts formula, we can obtain

$$\mathbb{E}[\mathbf{Y}_l^H\mathbf{Q}_l\boldsymbol{\mathcal{A}}\mathbf{Q}_k]_{i,p}[\mathbf{Y}_k]_{p,i} = \mathbb{E}\left\{\frac{[\mathbf{B}_{ik}\mathbf{B}_{il}^H\mathbf{Q}_l\boldsymbol{\mathcal{A}}\mathbf{Q}_k]_{p,p}}{\sqrt{N_kN_l}}\right.$$
$$-\frac{1}{\sqrt{N_kN_l}}[\mathbf{Y}_l^H\mathbf{Q}_l\mathbf{X}_l]_{i,i}[\mathbf{B}_{ik}\mathbf{A}_{il}^H\mathbf{Q}_l\boldsymbol{\mathcal{A}}\mathbf{Q}_k]_{p,p}$$
$$\left.-\frac{1}{N_k}[\mathbf{Y}_l^H\mathbf{Q}_l\boldsymbol{\mathcal{A}}\mathbf{Q}_k\mathbf{X}_k]_{i,i}[\mathbf{B}_{ik}\mathbf{A}_{ik}^H\mathbf{Q}_k]_{p,p}\right\}. \quad (106)$$

Summing over the subscript p, and using the variance control and the approximation rules for Υ_{kl}, we get

$$\mathbb{E}[\mathbf{Y}_l^H\mathbf{Q}_l\boldsymbol{\mathcal{A}}\mathbf{Q}_k\mathbf{Y}_k]_{i,i} = \frac{\overline{\Upsilon}_{kl}(\boldsymbol{\mathcal{A}}, \mathbf{B}_{ik}\mathbf{B}_{il}^H)}{\sqrt{N_kN_l}}$$
$$-\frac{\overline{\Upsilon}_{kl}(\boldsymbol{\mathcal{A}}, \mathbf{B}_{ik}\mathbf{A}_{il}^H)}{\sqrt{N_kN_l}}\mathbb{E}[\mathbf{Y}_l^H\mathbf{Q}_l\mathbf{X}_l]_{i,i}$$
$$-\mathbb{E}[\mathbf{Y}_l^H\mathbf{Q}_l\boldsymbol{\mathcal{A}}\mathbf{Q}_k\mathbf{X}_k]_{i,i}\frac{\mathrm{Tr}\mathbf{B}_{ik}\mathbf{A}_{ik}^H\boldsymbol{\Theta}_k}{N_k} + \mathcal{O}(N^{-\frac{3}{2}}). \quad (107)$$

Similarly, we can obtain

$$\mathbb{E}[\mathbf{Y}_l^H\mathbf{Q}_l\boldsymbol{\mathcal{A}}\mathbf{Q}_k\mathbf{X}_k]_{i,i} = \frac{\overline{\Upsilon}_{kl}(\boldsymbol{\mathcal{A}}, \mathbf{A}_{ik}\mathbf{B}_{il}^H)}{\sqrt{N_kN_l}(1+\delta_{ik})}$$
$$-\frac{\overline{\Upsilon}_{kl}(\boldsymbol{\mathcal{A}}, [\boldsymbol{\Omega}_i]_{[k,l]})}{\sqrt{N_kN_l}(1+\delta_{ik})}\mathbb{E}[\mathbf{Y}_l^H\mathbf{Q}_l\mathbf{X}_l]_{i,i} + \mathcal{O}(N^{-\frac{3}{2}}). \quad (108)$$

Then, plugging (108) into (107), summing over the subscript i and using the approximation rules for $\Phi_{B,kl}$ in (54), we get (109) at the top of this page, where $\widetilde{\boldsymbol{\Xi}}_{kl} = \mathbf{I}_M + \widetilde{\mathbf{F}}_k\widetilde{\mathbf{F}}_l\boldsymbol{\Xi}_{kl}\boldsymbol{\Gamma}_{kl}$, and step (a) in (109) follows by the identity $\widetilde{\mathbf{F}}_k\widetilde{\mathbf{F}}_l\boldsymbol{\Xi}_{kl} = \widetilde{\boldsymbol{\Xi}}_{kl}\widetilde{\mathbf{F}}_k\widetilde{\mathbf{F}}_l$. Therefore, we complete the proof of Lemma 3. □

REFERENCES

[1] L. Lu, G. Y. Li, A. L. Swindlehurst, A. Ashikhmin, and R. Zhang, "An overview of massive MIMO: Benefits and challenges," *IEEE J. Sel. Topics Signal Process.*, vol. 8, no. 5, pp. 742–758, Oct. 2014.

[2] C. Jeon, K. Li, J. R. Cavallaro, and C. Studer, "Decentralized equalization with feedforward architectures for massive MU-MIMO," *IEEE Trans. Signal Process.*, vol. 67, no. 17, pp. 4418–4432, Sep. 2019.

[3] L. Van der Perre, L. Liu, and E. G. Larsson, "Efficient DSP and circuit architectures for massive MIMO: State of the art and future directions," *IEEE Trans. Signal Process.*, vol. 66, no. 18, pp. 4717–4736, Sep. 2018.

[4] K. Li, R. R. Sharan, Y. Chen, T. Goldstein, J. R. Cavallaro, and C. Studer, "Decentralized baseband processing for massive MU-MIMO systems," *IEEE Trans. Emerg. Sel. Topics Circuits Syst.*, vol. 7, no. 4, pp. 491–507, Dec. 2017.

[5] Y. Xu, E. G. Larsson, E. A. Jorswieck, X. Li, S. Jin, and T.-H. Chang, "Distributed signal processing for extremely large-scale antenna array systems: State-of-the-art and future directions," *IEEE J. Sel. Topics Signal Process. (Early Access)*, pp. 1–526, 2025.

[6] K. Li, Y. Chen, R. Sharan, T. Goldstein, J. R. Cavallaro, and C. Studer, "Decentralized data detection for massive MU-MIMO on a Xeon Phi cluster," in *Proc. Conf. Rec. 50th Asilomar Conf. Signals, Syst. Comput.*, 2016, pp. 468–472.

[7] J. Rodríguez Sánchez, F. Rusek, O. Edfors, M. Sarajlić, and L. Liu, "Decentralized massive MIMO processing exploring daisy-chain architecture and recursive algorithms," *IEEE Trans. Signal Process.*, vol. 68, pp. 687–700, 2020.

[8] K. Li et al., "Design trade-offs for decentralized baseband processing in massive MU-MIMO systems," in *Proc. Conf. Rec. 53rd Asilomar Conf. Signals, Syst. Comput.*, 2019, pp. 906–912.

[9] Y. Dong, H. Li, C. Gong, X. Wang, and X. Dai, "An enhanced fully decentralized detector for the uplink M-MIMO system," *IEEE Trans. Veh. Technol.*, vol. 71, no. 12, pp. 13030–13042, Dec. 2022.

[10] X. Zhou, L. Liang, J. Zhang, C.-K. Wen, and S. Jin, "Mini-batch gradient-based MCMC for decentralized massive MIMO detection," *IEEE Trans. Commun.*, vol. 73, no. 1, pp. 677–692, Jan. 2025.

[11] K. Li, C. Jeon, J. R. Cavallaro, and C. Studer, "Decentralized equalization for massive MU-MIMO on FPGA," in *Proc. Conf. Rec. 51st Asilomar Conf. Signals, Syst. Comput.*, 2017, pp. 1532–1536.

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

[13] H. Q. Ngo, A. Ashikhmin, H. Yang, E. G. Larsson, and T. L. Marzetta, "Cell-free massive MIMO versus small cells," *IEEE Trans. Wireless Commun.*, vol. 16, no. 3, pp. 1834–1850, Mar. 2017.

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

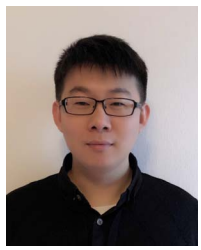
[15] J. Hoydis, S. ten Brink, and M. Debbah, "Massive MIMO in the UL/DL of cellular networks: How many antennas do we need?," *IEEE J. Select. Areas Commun.*, vol. 31, no. 2, pp. 160–171, Feb. 2013.

[16] J. Zhang, C. -K. Wen, S. Jin, X. Gao, and K.-K. Wong, "Large system analysis of cooperative multi-cell downlink transmission via regularized channel inversion with imperfect CSIT," *IEEE Trans. Wireless Commun.*, vol. 12, no. 10, pp. 4801–4813, Oct. 2013.

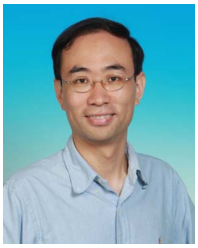
[17] Z. Wei, D. Xu, S. Li, S. Song, D. W. K. Ng, and G. Caire, "Resource allocation design for next-generation multiple access: A tutorial overview," in *Proc. IEEE*, vol. 112, no. 9, pp. 1230–1263, Sep. 2024.
[18] A. Kammoun, M. Kharouf, W. Hachem, and J. Najim, "A central limit theorem for the SINR at the LMMSE estimator output for large-dimensional signals," *IEEE Trans. Inf. Theory*, vol. 55, no. 11, pp. 5048–5063, Nov. 2009.
[19] A. Kammoun et al., "BER and outage probability approximations for LMMSE detectors on correlated MIMO channels," *IEEE Trans. Inf. Theory*, vol. 55, no. 10, pp. 4386–4397, Oct. 2009.
[20] X. Zhang, S. Song, and Y. C. Eldar, "Secrecy analysis for MISO broadcast systems with regularized zero-forcing precoding," 2023, *arXiv:2301.11515*.
[21] X. Zhao, M. Li, Y. Liu, T. -H. Chang, and Q. Shi, "Communication-efficient decentralized linear precoding for massive MU-MIMO systems," *IEEE Trans. Signal Process.*, vol. 71, pp. 4045–4059, 2023.
[22] X. Zhao, M. Li, B. Wang, E. Song, T.-H. Chang, and Q. Shi, "Efficient LMMSE equalization for massive MIMO systems under decentralized baseband processing architecture," *IEEE J. Select. Areas Commun*, vol. 43, no. 3, pp. 736–751, Mar. 2025.
[23] S. Wagner, R. Couillet, M. Debbah, and D. T. M. Slock, "Large system analysis of linear precoding in correlated MISO broadcast channels under limited feedback," *IEEE Trans. Inf. Theory*, vol. 58, no. 7, pp. 4509–4537, Jul. 2012.
[24] J. Hoydis, M. Kobayashi, and M. Debbah, "Asymptotic performance of linear receivers in network MIMO," in *Proc. Conf. Rec. 44th Asilomar Conf. Signals, Syst. Comput.*, 2010, pp. 942–948.
[25] E. Björnson, J. Hoydis, and L. Sanguinetti, *Massive MIMO Networks: Spectral, Energy, and Hardware Efficiency*. London, U.K.: Now Found. Trends, 2017.
[26] B. Picinbono, "Second-order complex random vectors and normal distributions," *IEEE Trans. Signal Process.*, vol. 44, no. 10, pp. 2637–2640, Oct. 1996.
[27] B. Hassibi and B. Hochwald, "How much training is needed in multiple-antenna wireless links?," *IEEE Trans. Inf. Theory*, vol. 49, no. 4, pp. 951–963, Apr. 2003.
[28] A. Abrardo, G. Fodor, M. Moretti, and M. Telek, "MMSE receiver design and SINR calculation in MU-MIMO systems with imperfect CSI," *IEEE Wireless Commun. Lett.*, vol. 8, no. 1, pp. 269–272, Feb. 2019.
[29] D. Guo, S. Shamai, and S. Verdu, "Mutual information and minimum mean-square error in Gaussian channels," *IEEE Trans. Inf. Theory*, vol. 51, no. 4, pp. 1261–1282, Apr. 2005.
[30] A. W. Marshall, I. Olkin, and B. C. Arnold, *Inequalities: Theory of Majorization and its Applications*. New York, NY, USA: Academic Press, 1979.
[31] Z. Zhuang, X. Zhang, D. Xu, S. Song, and Y. C. Eldar, "Decentralized MIMO systems with imperfect CSI using LMMSE receivers," 2025, *arXiv:2408.12811*.
[32] J. Zhang, C.-K. Wen, S. Jin, X. Gao, and K.-K. Wong, "On capacity of large-scale MIMO multiple access channels with distributed sets of correlated antennas," *IEEE J. Select. Areas Commun.*, vol. 31, no. 2, pp. 133–148, Feb. 2013.
[33] X. Zhang and S. Song, "Secrecy analysis for IRS-aided wiretap MIMO communications: Fundamental limits and system design," *IEEE Trans. Inf. Theory*, vol. 70, no. 6, pp. 4140–4159, Jun. 2024.
[34] A. Kammoun and M.-S. Alouini, "No eigenvalues outside the limiting support of generally correlated gaussian matrices," *IEEE Trans. Inf. Theory*, vol. 62, no. 7, pp. 4312–4326, Jul. 2016.
[35] A. W. Van der Vaart, *Asymptotic Statistics*. Cambridge, U.K.: Cambridge Univ. Press, 1998.
[36] Z.-D. Bai and J. W. Silverstein, "No eigenvalues outside the support of the limiting spectral distribution of large-dimensional sample covariance matrices," *Ann. Probab.*, vol. 26, no. 1, pp. 316–345, Jan. 1998.
[37] W. Hachem, O. Khorunzhiy, P. Loubaton, J. Najim, and L. Pastur, "A new approach for mutual information analysis of large dimensional multi-antenna channels," *IEEE Trans. Inf. Theory*, vol. 54, no. 9, pp. 3987–4004, Sep. 2008.
[38] R. A. Horn and C. R. Johnson, *Matrix Analysis*. Cambridge, U.K.: Cambridge Univ. Press, 2012.
[39] L. A. Pastur and M. Shcherbina, *Eigenvalue Distribution of Large Random Matrices*. Providence, RI, USA: Amer. Math. Soc., 2011.
[40] X. Zhang and S. Song, "Asymptotic mutual information analysis for double-scattering MIMO channels: A new approach by Gaussian tools," *IEEE Trans. Inf. Theory*, vol. 69, no. 9, pp. 5497–5527, Sep. 2023.
[41] Z. Zhuang, X. Zhang, D. Xu, and S. Song, "Fundamental limits of two-hop MIMO channels: An asymptotic approach," 2024, *arXiv:2402.03772*.

Zeyan Zhuang (Graduate Student Member, IEEE) received the B.Eng. degree from Nankai University, Tianjin, China, in 2022. He is currently working toward the Ph.D. degree with the Department of Electronic and Computer Engineering, The Hong Kong University of Science and Technology, Hong Kong. His research interests include random matrix theory and its corresponding applications.

Xin Zhang (Member, IEEE) received the B.Eng. degree in information engineering from the Beijing University of Posts and Telecommunications (BUPT) in 2015, the master's degree in electronic engineering from Tsinghua University, Beijing, China, in 2018, and the Ph.D. degree in electronic and computer engineering (ECE) from the Hong Kong University of Science and Technology (HKUST), Hong Kong, in 2024. He is currently a Postdoctoral Research Fellow with HKUST. His research interests include random matrix theory, information theory, and their applications in signal processing, communications, and learning. He was recognized as an Exemplary Reviewer of the IEEE COMMUNICATIONS LETTERS in 2024.



Dongfang Xu (Member, IEEE) received the B.Eng. degree in communication engineering from Shandong University, Jinan, China in July 2014, and the M.Sc. degree (with Distinction) in communications and multimedia engineering and the Ph.D. degree (with Distinction) in electrical engineering from Friedrich-Alexander-Universität Erlangen-Nürnberg (FAU), Erlangen, Germany, in September 2017 and December 2022, respectively. He is currently a Research Assistant Professor with the Hong Kong University of Science and Technology (HKUST), Hong Kong. He was the co-recipient of the IEEE Communications Society Leonard G. Abraham Prize 2023, IEEE Communications Society Stephen O. Rice Prize 2022, and IEEE Global Communications Conference (GLOBECOM) 2019 Best Paper Award. He was also recognized as an Exemplary Reviewer of IEEE TRANSACTIONS ON COMMUNICATIONS in 2020, 2021, and 2022, and an Exemplary Reviewer of the IEEE COMMUNICATIONS LETTERS in 2023. He is also the Editor of IEEE TRANSACTIONS ON MOBILE COMPUTING and IEEE COMMUNICATIONS LETTERS.



Shenghui Song (Senior Member, IEEE) is currently an Assistant Professor jointly appointed by the Division of Integrative Systems and Design (ISD) and the Department of Electronic and Computer Engineering (ECE) with the Hong Kong University of Science and Technology (HKUST), Hong Kong. His research interests include the areas of wireless communications and machine learning with current focus on distributed intelligence, semantic communications, machine learning for communications, integrated sensing and communication, and information theory. He is also interested in the research on Engineering Education. He was named the Exemplary Reviewer for IEEE COMMUNICATIONS LETTER. He was the Tutorial Program Co-Chairs of the 2022 IEEE International Mediterranean Conference on Communications and Networking, and the Technical Program Chairs of the International Conference on 6G Communications Networking and Signal Processing, in 2023 and 2024. He is also an Associate Editor for IEEE TRANSACTIONS ON EDUCATION. He was the recipient of the several teaching awards at HKUST, including the Michael G. Gale Medal for Distinguished Teaching in 2018, Best Ten Lecturers in 2013, 2015, and 2017, School of Engineering Distinguished Teaching Award in 2012, Teachers I Like Award in 2013, 2015, 2016, and 2017, M.Sc. (Telecom) Teaching Excellent Appreciation Award for 2020-21 and 2022-23, and the honorees of the Third Faculty Recognition at HKUST in 2021.

Yonina C. Eldar (Fellow, IEEE) received the B.Sc. degree in physics and the B.Sc. degree in electrical engineering from Tel-Aviv University (TAU), Tel-Aviv, Israel, in 1995 and 1996, respectively, and the Ph.D. degree in electrical engineering and computer science from the Massachusetts Institute of Technology (MIT), Cambridge, MA, USA, in 2002. She is currently a Professor with the Department of Mathematics and Computer Science, Weizmann Institute of Science, Rehovot, Israel, where she holds the Dorothy and Patrick Gorman Professorial Chair and also Heads the Center for Biomedical Engineering. She was a Professor with the Department of Electrical Engineering, Technion, where she held the Edwards Chair in Engineering. She is also a Visiting Professor with MIT, Visiting Scientist with Broad Institute, Visiting Research Collaborator with Princeton, an Adjunct Professor with Duke University, Durham, NC, USA, Advisory Professor with Fudan University, Shanghai, China, Distinguished Visiting Professor with Tsinghua University, Beijing, China, and was Visiting Professor with Stanford. She was a Horev Fellow of the Leaders in Science and Technology Program with Technion and an Alon Fellow. She is author of the book *Sampling Theory: Beyond Bandlimited Systems* and co-author of ten other books and also a highly cited Researcher. Her research interests include the broad areas of statistical signal processing, sampling theory, and compressed sensing, learning and optimization methods, and their applications to biology, medical imaging, and optics. Dr. Eldar was the recipient of the many awards for excellence in research and teaching, including the IEEE Signal Processing Society Technical Achievement Award (2013), IEEE/AESS Fred Nathanson Memorial Radar Award (2014), IEEE Kiyo Tomiyasu Award (2016), Michael Bruno Memorial Award from the Rothschild Foundation, Weizmann Prize for Exact Sciences, Wolf Foundation Krill Prize for Excellence in Scientific Research, Henry Taub Prize for Excellence in Research (twice), Hershel Rich Innovation Award (three times), Award for Women with Distinguished Contributions, Andre and Bella Meyer Lectureship, Career Development Chair at the Technion, Muriel David Jacknow Award for Excellence in Teaching, Technion's Award for Excellence in Teaching (two times), several Best Paper awards and Best Demo awards together with her research students and colleagues including the SIAM outstanding Paper Prize, UFFC Outstanding Paper Award, Signal Processing Society Best Paper Award and the IET Circuits, Devices and Systems Premium Award, was selected as one of the 50 most influential women in Israel and in Asia. She is the Editor in Chief of *Foundations and Trends in Signal Processing*. She was a Signal Processing Society Distinguished Lecturer, member of the IEEE Signal Processing Theory and Methods and Bio Imaging Signal Processing Technical Committees. She was an Associate Editor for the IEEE TRANSACTIONS ON SIGNAL PROCESSING, *EURASIP Journal of Signal Processing*, *SIAM Journal on Matrix Analysis and Applications*, and *SIAM Journal on Imaging Sciences*. She was Co-Chair and Technical Co-Chair of several international conferences and workshops. She was a member of the Young Israel Academy of Science and Humanities and the Israel Committee for Higher Education. She is a member with the Israel Academy of Sciences and Humanities (elected 2017) and with the Academia Europaea (elected 2023), EURASIP Fellow, Fellow of the Asia-Pacific Artificial Intelligence Association, and Fellow of the 8400 Health Network. She is also a member of the IEEE Sensor Array and Multichannel Technical Committee and serves on several other IEEE committees.