

Splitting Messages in the Dark— Rate-Splitting Multiple Access for FDD Massive MIMO without CSI Feedback

Namhyun Kim, Ian P. Roberts, and Jeonghun Park

Abstract—A critical hindrance in realizing frequency division duplex (FDD) massive multi-input multi-output (MIMO) systems is the overhead associated with the downlink (DL) channel state information at the transmitter (CSIT) acquisition. To address this, we propose a novel framework that eliminates the need for CSI feedback, while achieving robust sum spectral efficiency (SE). Specifically, by leveraging partial frequency invariance of channel parameters, we reconstruct the DL CSIT using uplink (UL) pilots with the 2D-Newtonized orthogonal matching pursuit (2D-NOMP) algorithm. Due to discrepancies between the two disjoint bands, however, perfect DL CSIT acquisition is infeasible; resulting in multi-user interference (MUI). To account for this, we reformulate the sum SE maximization problem using the reconstructed channel and its error covariance matrix (ECM). Then, we propose an ECM estimation method based on the observed Fisher information matrix and introduce a precoder optimization technique with rate-splitting multiple access (RSMA). Our simulation results verify the validity of the proposed framework in the practical FDD massive MIMO scenarios, highlighting the essential role of ECM estimation in mitigating MUI to attain RSMA gains.

Index Terms—FDD massive MIMO, Error covariance estimation, Rate-splitting multiple access, Generalized power iteration.

I. INTRODUCTION

Massive multiple-input multiple-output (MIMO) systems are key to the future of wireless communications, effectively increasing spectral efficiency (SE) and reducing costs [1], [2]. To fully realize these benefits, accurate channel state information at the transmitter (CSIT) is critical. In this context, time division duplexing (TDD)-based channel estimation has been preferred for its efficiency from channel reciprocity [3].

Nonetheless, there have been persistent research efforts aimed at realizing the advantages of massive MIMO for frequency division duplexing (FDD) systems as well [4], [5]. The enduring importance of these studies is usually underscored by the following considerations. i) The majority of the bandwidth is allocated to the FDD mode in sub-6 GHz, which still dominates the cellular market today [6]. ii) FDD typically offers superior coverage performance compared to TDD [3], allows simultaneous uplink (UL) and downlink

(DL) transmissions, catering to delay-sensitive applications [7]. Despite these benefits, FDD massive MIMO systems face significant challenges as the number of base station (BS) antennas grows, primarily due to the prohibitive complexity of CSIT acquisition from the lack of channel reciprocity [1], [8]. In this paper, we take a step toward addressing this problem while maximizing the sum SE performance.

A. Prior Work

We categorize the prior work on realizing FDD massive MIMO systems into two primary approaches: those aimed at reducing CSIT overhead and those aimed at mitigating multi-user interference (MUI).

1) *CSIT overhead reduction approaches*: One of the main principles of these studies is that the channel characteristics in FDD MIMO can be compressed in certain domains, which is helpful to reduce the overhead [9]. For instance, a distributed DL CSIT compression method was developed by [8], which harnessed the shared sparsity structures from local scatterers. In [10], a super-resolution compressed sensing (CS) method was proposed in the presence of the beam squint effect in mmWave wideband MIMO-OFDM systems. Recently, methods not relying on a specific compressible basis but rather on data-driven approaches have also been studied. For example, in [11], CS-ReNet was devised, which jointly exploits deep learning techniques and CS-based compression. For example, [12] proposed a multiple-measurement-vectors learned approximate message passing (MMV-LAMP) network to enable efficient channel reconstruction of the spatial-frequency channel matrix, using channels' structured sparsity. In [13], it was proposed to use deep learning to map channels from one set of antennas and frequency band to channels of another set, significantly reducing the training overhead. An extensive overview of deep learning based CSIT acquisition was provided in [14].

In a different line of research, [15] presented an intriguing idea: the DL channel in FDD can be fully reconstructed by UL training without CSI feedback. Specifically, DL channel parameters, such as the angle of departure (AoD), delay, and path gains, can be obtained from the UL channel, assuming that the frequency difference of UL and DL is small enough relative to the carrier frequency. Extensions of this concept were discussed in [4], [5], [16], [17], with [18] supported by measurement campaigns. In [19], channel

N. Kim and J. Park are with the School of Electrical and Electronic Engineering, Yonsei University, Seoul 03722, South Korea. (e-mail: namhyun@yonsei.ac.kr, jhpark@yonsei.ac.kr). I. P. Roberts is with the Department of Electrical and Computer Engineering, University of California, Los Angeles, CA 90095 USA. (e-mail: ianroberts@ucla.edu).

quality indicator (CQI) feedback was omitted by exploiting a machine learning (ML) based technique for reducing delay. However, the accuracy of the reconstructed DL channel is sometimes reported to be limited due to discrepancies between the UL and DL channels, as shown in [4], [16]. It is important to note that this imperfection can significantly undermine the SE performance in DL due to increased MUI.

2) *MUI mitigation approaches*: In the presence of imperfect CSIT, MUI mitigation strategies have been explored as another approach to unleash the gains of FDD massive MIMO systems. In [20], an interference alignment scheme was developed to manage MUI caused from delayed CSIT. In [21], it was shown that when users are sufficiently spatially separated, their channel covariance matrices have non-overlapping subspaces, leading to reduction of MUI.

Recently, an unorthodox type of multiple access, referred as rate-splitting multiple access (RSMA), has also been actively explored to handle the MUI problem [22]. In RSMA, users partially decode MUI through the introduction of a common message and successive interference cancellation (SIC), leading to interference decoding gains. Building on this, [23] presented a linear precoder design based on the weighted minimum mean square error (WMMSE) method, using the sample average approximation (SAA) technique to account for imperfect CSIT in RSMA. Beyond this, [24] proposed a generalized power iteration (GPI)-based precoder optimization, assuming that the channel covariance and DL channel error covariance are given in TDD MIMO. In addition, a bilinear precoding method in FDD was introduced in [25], considering that the knowledge of channel covariance is provided. A fractional programming (FP) based RSMA precoding method was developed in [26], wherein the optimal beamforming structure of RSMA is incorporated.

Notice that the two approaches—CSIT overhead reduction and MUI mitigation—are closely related; yet most previous studies have largely existed in isolation, focusing on one approach while overlooking the other. For instance, efficient linear precoding methods for RSMA were explored in several prior work by considering various CSIT acquisition models; such as FDD MIMO with limited feedback [26] or FDD MIMO with analog feedback [25]. However, no prior work has explored the case where CSI feedback is entirely omitted in FDD MIMO. A key challenge in this case is that, since the DL CSIT cannot be directly obtained, it must be reconstructed from side information, such as UL reference signals [15], [18]. Nonetheless, the impact of the DL CSIT reconstruction algorithm on RSMA performance is not clear [15], [18], thus it is not yet known how to incorporate such DL CSIT reconstruction into RSMA precoder design. For this reason, jointly tackling the DL channel reconstruction and the RSMA precoder design remains an open problem.

B. Contributions

In this paper, we put forth a comprehensive approach to reduce the CSIT acquisition overhead and also efficiently

mitigate MUI. Specifically, we consider an FDD massive MIMO system without direct CSI feedback, and instead estimate the DL channel based on UL pilots. To accomplish this, we employ the 2D-Newtonized orthogonal matching pursuit (NOMP) algorithm that extracts key parameters of UL channels, such as angle of arrival (AoA), delay, and path gain [18]. With this, we rebuild the DL channel by leveraging partial frequency invariance between UL and DL channels [4], [16], [18]. However, due to the inherent discrepancies between the UL and DL channels, CSIT errors are unavoidable, which results in MUI that undermines the SE gains. To resolve this, we employ RSMA.

Nonetheless, realizing the full potential of RSMA is not straightforward. This is mainly because it is difficult to capture the impact of imperfect CSIT reconstruction on RSMA performance, making it challenging to efficiently design the precoder that accounts for the CSIT errors. To resolve this, we characterize the achievable SE as a function of the precoding vectors, the reconstructed CSIT, and the CSIT error covariance matrix (ECM). In our characterization, we find that the ECM plays a crucial role in designing precoders that effectively mitigate the MUI. Since the ECM cannot be obtained in a conventional way due to non-linearity of 2D-NOMP, we develop a novel ECM estimation method. Our rationale is built on the Cramér-Rao lower bound (CRLB) of DL channel reconstruction, where its trace provides a lower bound for MSE. However, computing the CRLB requires true channel parameters, which is not feasible in practice. To address this, we use the observed Fisher information matrix (O-FIM) [27]–[29], which requires only the observed UL reference signals. This enables us to compute the CRLB without the ground truth value or CSI error distribution, thereby improving the practicability of our method.

Subsequently, harnessing the estimated ECM, we propose a precoder optimization method to solve the sum SE maximization problem. Specifically, we derive the Karush-Kuhn-Tucker (KKT) condition for maximizing the achievable sum SE. This can then be interpreted as an eigenvector-dependent nonlinear eigenvalue problem. Accordingly, by finding the leading eigenvector of the derived KKT condition, we reach a locally optimal solution that maximizes the sum SE.

Through numerical simulations, we demonstrate that the proposed method offers significant improvement in terms of sum SE compared to existing methods. For instance, we observe up to a 22.7% increase in sum SE performance compared to other baseline spatial division multiple access (SDMA) methods when no CSI feedback is used. Crucially, we reveal that these gains primarily stem from our ECM estimation. By accurately estimating the ECM and incorporating this into the precoder optimization, we properly account for the CSIT errors and the resultant MUI. Specifically, when incorporating the estimated ECM into the precoder design, we observe up to a 21.1% improvement in sum SE performance compared to a case that does not rely on the estimated ECM. In addition, we also examine a case where the number of DL channel paths is reduced compared to the UL.

We observe that the proposed method still provides robust performance with our appropriate ECM estimation. Further, we present that our framework unlocks new possibilities for enabling low-latency FDD massive MIMO transmission. Specifically, we analyze the latency overhead associated with the typical 5G-NR process of CSI feedback. Based on this, we compare the achievable sum SE between the perfect CSIT case (with feedback) and the no CSI feedback case, offering insights into the trade-off between latency overhead and achievable sum SE. This trade-off highlights how our framework provides an effective alternative for conventional FDD massive MIMO transmission, significantly reducing latency with only a marginal sacrifice in SE performance.

Notation: The superscripts $(\cdot)^T$, $(\cdot)^H$, $(\cdot)^{-1}$, $(\cdot)^\dagger$ denote the transpose, Hermitian, matrix inversion, and Moore-Penrose pseudo-inverse respectively. \mathbf{I}_N is the identity matrix of size $N \times N$. Assuming that $\mathbf{A}_1, \dots, \mathbf{A}_N \in \mathbb{C}^{K \times K}$, $\mathbf{A} = \text{diag}(\mathbf{A}_1, \dots, \mathbf{A}_n, \dots, \mathbf{A}_N)$ is a block-diagonal matrix concatenating $\mathbf{A}_1, \dots, \mathbf{A}_N$. We use $\lfloor \cdot \rfloor$, $\lceil \cdot \rceil$ to denote rounding a decimal number to its nearest lower and higher integers, respectively. We define $[K]$ as the set of natural numbers less than or equal to K . $\mathbf{A} \circ \mathbf{B}$ denotes the Hadamard product of the two matrices, \mathbf{A} and \mathbf{B} .

II. SYSTEM MODEL

We consider the multi-user FDD massive MIMO system, where the BS is equipped with N antennas and serves K single-antenna users. We also assume that the channel remains time-invariant from the transmission of the UL pilots to the DL precoding [5], [18].

A. Channel Model

With orthogonal frequency division multiplexing (OFDM) employed, we denote the number of sub-carriers used for pilots as M and the spacing between these sub-carriers as Δf which implicitly defines the bandwidth of each resource block. That is, these M pilots are uniformly spaced, and the total bandwidth they span is $B = M\Delta f$. Following the multipath channel model in [15], [18], we represent the UL channel on the m -th sub-carrier of user $k \in [K]$ as

$$\mathbf{h}_k^{\text{ul}}[m] = \sum_{\ell=1}^{L_k^{\text{ul}}} \alpha_{k,\ell}^{\text{ul}} \mathbf{a}(\theta_{k,\ell}^{\text{ul}}; \lambda_m^{\text{ul}}) e^{-j2\pi m \Delta f \tau_{k,\ell}^{\text{ul}}} \in \mathbb{C}^{N \times 1}, \quad (1)$$

where $\alpha_{k,\ell}^{\text{ul}}$ is the complex path gain of the ℓ -th path, and $\mathbf{a}(\theta_{k,\ell}^{\text{ul}}; \lambda_m^{\text{ul}})$ is the ULA array response defined sub-carrier specifically as

$$\mathbf{a}(\theta_{k,\ell}^{\text{ul}}; \lambda_m^{\text{ul}}) = \left[1, e^{-j2\pi \frac{d^{\text{ul}}}{\lambda_m^{\text{ul}}} \sin \theta_{k,\ell}^{\text{ul}}}, \dots, e^{-j2\pi(N-1) \frac{d^{\text{ul}}}{\lambda_m^{\text{ul}}} \sin \theta_{k,\ell}^{\text{ul}}} \right]^T, \quad (2)$$

where we denote the AoA as $\theta_{k,\ell}^{\text{ul}}$, the wavelength of the m -th UL sub-carrier as λ_m^{ul} , and the antenna spacing as d^{ul} with $d^{\text{ul}} = \lambda_c^{\text{ul}}/2$, where λ_c^{ul} is the wavelength of carrier frequency

of UL. Further, we denote the propagation delay as $\tau_{k,\ell}^{\text{ul}}$ with $0 < \tau_{k,\ell}^{\text{ul}} < 1/\Delta f$ and the number of paths as L_k^{ul} . We also assume that the baseband UL signal occupies the interval $[-B/2, B/2]^1$. Correspondingly, index m satisfies $\lfloor \frac{-M}{2} \rfloor \leq m \leq \lceil \frac{M}{2} \rceil - 1, m \in \mathbb{Z}$.

To model the distribution of the complex path gain $\alpha_{k,\ell}^{\text{ul}}$, we use the Rician distribution, i.e. $\alpha_{k,\ell}^{\text{ul}} \sim \mathcal{CN}\left(\sqrt{\frac{\kappa_k \sigma_{\text{path},k}^2}{\kappa_k + 1}}, \frac{\sigma_{\text{path},k}^2}{\kappa_k + 1}\right), \forall \ell$. The parameter κ_k denotes the Rician factor of user k and represents the relative strength of the line-of-sight (LoS) component compared to the non-line-of-sight (NLoS) component. The parameter $\sigma_{\text{path},k}^2$ denotes the average channel power of each path, i.e. $\mathbb{E}[|\alpha_{k,\ell}^{\text{ul}}|^2] = \sigma_{\text{path},k}^2, \forall \ell$, where we define $\sigma_{\text{path},k}^2 = 1/(NL_k^{\text{ul}})$ to normalize the channel in (1) to unit variance. We note that $\alpha_{k,\ell}, \theta_{k,\ell}, \tau_{k,\ell}$ are consistent across the sub-carriers within operating bandwidth, under the assumption that the bandwidth is much smaller than the carrier frequency [5], [18].

Notice that in (1), we describe the UL channel vector ($\mathbf{h}_k^{\text{ul}}[m] \in \mathbb{C}^{N \times 1}$) with respect to the OFDM sub-carrier index m . This signal representation in the frequency domain is crucial for the joint estimation of angles ($\theta_{k,\ell}^{\text{ul}}$) and delays ($\tau_{k,\ell}^{\text{ul}}$) during the UL training phase (as explained in Section III) as demonstrated in [15]. In contrast, in other parts—such as defining performance metric and designing precoders—the sub-carrier notation becomes irrelevant. Once reconstructed DL channel is obtained during UL training, the performance metric, i.e., SE, depends solely on the composite DL channel. In our approach, the channel estimate for each narrowband resource block is readily available after identifying the DL channel parameters [18], [23], [26]. Consequently, the precoder design algorithm is applied independently to each resource block, making it unnecessary to retain the sub-carrier index m to define the DL channel. Accordingly, assuming the difference between the UL carrier frequency and the target DL resource block is f , the DL channel as a function of f is given by

$$\mathbf{h}_k(f) = \sum_{\ell=1}^{L_k^{\text{dl}}} \alpha_{k,\ell}^{\text{dl}} \mathbf{a}(\theta_{k,\ell}^{\text{dl}}; \lambda^{\text{dl}}) e^{-j2\pi f \tau_{k,\ell}^{\text{dl}}} \in \mathbb{C}^{N \times 1}, \quad (3)$$

where $\alpha_{k,\ell}^{\text{dl}}, \theta_{k,\ell}^{\text{dl}}, \tau_{k,\ell}^{\text{dl}}, L_k^{\text{dl}}$ are defined in the same way as for the UL channel, and λ^{dl} denotes the wavelength of the corresponding resource block.

As presented in [15], [17] and also supported by an actual measurement campaign [4], the UL and DL channels share several frequency-invariant parameters. We summarize the partial reciprocity relationship for all (k, ℓ) as follows:

- Number of channel paths: $L_k^{\text{ul}} = L_k^{\text{dl}} \triangleq L_k$,
- AoA (or AoD): $\theta_{k,\ell}^{\text{ul}} = \theta_{k,\ell}^{\text{dl}}$,
- Propagation delay: $\tau_{k,\ell}^{\text{ul}} = \tau_{k,\ell}^{\text{dl}}$,
- Complex path gain: $\alpha_{k,\ell}^{\text{ul}} \approx \alpha_{k,\ell}^{\text{dl}}$.

¹This follows the convention where the carrier frequency of the UL band is set to zero [5]. Accordingly, we can represent the DL channel as a function of the frequency difference between the DL and UL.

We emphasize the critical role of frequency-invariant parameters in reconstructing the DL channel from UL training. However, the last assumption of frequency-invariant complex path gain remains controversial. Although [15], [17] claimed that each path gain is also frequency-invariant, i.e. $\alpha_{k,\ell}^{\text{ul}} = \alpha_{k,\ell}^{\text{dl}}, \forall(k, \ell)$, the real-world measurement campaign [16], [18] reported that the instantaneous channel gains may differ. Instead, they tend to have a highly correlated second moment [16]. To account for this, we adopt the DL channel gain model using the first-order Gauss-Markov model:

$$\alpha_{k,\ell}^{\text{dl}} = \eta_{k,\ell} \alpha_{k,\ell}^{\text{ul}} + \sqrt{1 - \eta_{k,\ell}^2} \beta_{k,\ell}, \beta_{k,\ell} \sim \mathcal{CN}(0, \sigma_{\text{path},k}^2), \forall \ell, \quad (4)$$

where $\eta_{k,\ell}$ is the correlation between ℓ -th path of the DL and UL channels for user k . For example, if $\eta_{k,\ell} = 1, \forall(k, \ell)$, our model (4) reduces to the case of perfect reciprocity, while $\eta_{k,\ell} = 0, \forall(k, \ell)$ indicates that UL/DL gains are independent, where our model includes [15], [17] as an edge case and will allow us to investigate performance as a function of path correlation.

B. Signal Model

We consider RSMA, which is a robust multiple access technique against imperfect CSIT [24], [25]. In this paper, we employ 1-layer RSMA so that only one layer of common messages for the all users is used [30]. Specifically, the intended message W_k for user k is split into a common part $W_{c,k}$ and a private part $W_{p,k}$. Then, the common parts of entire users' messages are combined to yield a common message W_c . Subsequently, the $K+1$ messages (the common message W_c and the K private messages, $W_{p,1}, \dots, W_{p,K}$) are encoded into the symbols $s_c, s_{p,1}, \dots, s_{p,K}$ respectively. Note that the common message s_c is designed to be decodable to the all users, which implies that the common rate is not larger than the minimum common rate of entire users.

The transmit signal at the BS is then formed by combining the messages and linear precoding vectors as

$$\mathbf{x} = \mathbf{f}_c s_c + \sum_{k=1}^K \mathbf{f}_k s_k, \quad (5)$$

where $\mathbf{f}_c, \mathbf{f}_k \in \mathbb{C}^{N \times 1}$ are precoding vectors for the common and private stream for user k , respectively. The transmit power constraint is $\|\mathbf{f}_c\|^2 + \sum_{k=1}^K \|\mathbf{f}_k\|^2 \leq 1$. In addition, we assume Gaussian signaling with power constraint P , i.e., $s_c, s_k \sim \mathcal{CN}(0, P)$. At user $k \in [K]$, the received signal is

$$r_k(f) = \mathbf{h}_k^H(f) \mathbf{f}_c s_c + \mathbf{h}_k^H(f) \mathbf{f}_k s_k + \sum_{p=1, p \neq k}^K \mathbf{h}_k^H(f) \mathbf{f}_p s_p + z_k, \quad (6)$$

where $z_k \sim \mathcal{CN}(0, \sigma^2)$ is additive white Gaussian noise (AWGN).

III. DL CHANNEL RECONSTRUCTION AND PERFORMANCE CHARACTERIZATION

In this section, we explain a DL channel reconstruction method using the UL reference signal without CSI feedback. We then derive the achievable SE with the used DL channel reconstruction method.

A. DL Channel Reconstruction

To begin, we define a vector $\mathbf{u}(\tau, \theta) \in \mathbb{C}^{MN \times 1}$ as the received UL training signal across M sub-carriers and N antennas as follows:

$$[\mathbf{u}(\tau, \theta)]_{m,n} = e^{-j2\pi \frac{nd}{\lambda_m^{\text{ul}}} \sin \theta_{k,\ell}^{\text{ul}}} e^{-j2\pi (\lfloor \frac{M}{2} \rfloor + m - 1) \Delta f \tau_{k,\ell}}, \quad (7)$$

where $[\cdot]_{m,n}$ denotes the element of the vector associated with the m -th sub-carrier and n -th antenna. Assuming an all-ones UL sounding reference signal is used, the received signal is given by

$$\mathbf{y}_k = \sum_{\ell=1}^{L_k} \alpha_{k,\ell}^{\text{ul}} \mathbf{u}(\tau_{k,\ell}^{\text{ul}}, \theta_{k,\ell}^{\text{ul}}) + \mathbf{w}_k \in \mathbb{C}^{MN \times 1}, \quad (8)$$

where \mathbf{w}_k is AWGN. Our objective is to predict the DL channel $\mathbf{h}_k(f)$ using \mathbf{y}_k . To this end, we first obtain the UL triplet path parameters, i.e., $\left\{ \left(\hat{\alpha}_{k,\ell}^{\text{ul}}, \hat{\tau}_{k,\ell}^{\text{ul}}, \hat{\theta}_{k,\ell}^{\text{ul}} \right) \right\}_{\ell=1, \dots, L_k}$ from \mathbf{y}_k and rebuild the DL channel $\hat{\mathbf{h}}_k(f)$ by incorporating the frequency invariance. As demonstrated in [31], the number of channel paths in FDD massive MIMO is significantly smaller than the number of BS antennas, allowing us to formulate the UL channel parameter extraction problem as a CS problem. For solving this, we employ a variant of orthogonal matching pursuit, called 2D-NOMP. As demonstrated in [18], the 2D-NOMP algorithm has shown to be suitable for real-time hardware implementations. We explain the basic principle and detailed process of the 2D-NOMP as follows.

For every iteration, the 2D-NOMP algorithm finds the maximum likelihood (ML) estimate of the angle, delay, and channel gain of each path in UL, say $(\tilde{\alpha}_{k,\ell}, \tilde{\tau}_{k,\ell}, \tilde{\theta}_{k,\ell})$. With the estimate, the residual signal at the end of the i -th iteration is computed by

$$\mathbf{y}_r = \mathbf{y}_k - \sum_{\ell=1}^i \tilde{\alpha}_{k,\ell} \mathbf{u}(\tilde{\tau}_{k,\ell}, \tilde{\theta}_{k,\ell}), \quad (9)$$

and \mathbf{y}_r is initialized to \mathbf{y}_k . The ML estimate of the parameters is attained by minimizing the residual power $\|\mathbf{y}_r - \tilde{\alpha}_{k,i} \mathbf{u}(\tilde{\tau}_{k,i}, \tilde{\theta}_{k,i})\|^2$, which is equivalent to maximizing

$$f_{\mathbf{y}_r}(\alpha_{k,\ell}, \tau_{k,\ell}, \theta_{k,\ell}) \triangleq 2 \text{Re} \{ \mathbf{y}_r^H \alpha_{k,\ell} \mathbf{u}(\tau_{k,\ell}, \theta_{k,\ell}) \} - |\alpha_{k,\ell}|^2 \|\mathbf{u}(\tau_{k,\ell}, \theta_{k,\ell})\|^2. \quad (10)$$

Based on this, the three following steps are performed during the i -th iteration until a stopping criterion is met.

1. *New detection*: To maximize (10), the estimate for i -th path for user k , i.e., $(\tilde{\tau}_{k,i}, \tilde{\theta}_{k,i})$ is chosen from a predefined parameter grid, Ω as

$$(\tilde{\tau}_{k,i}, \tilde{\theta}_{k,i}) = \arg \max_{(\tau, \theta) \in \Omega} \frac{|\mathbf{u}^H(\tau, \theta) \mathbf{y}_r|^2}{\|\mathbf{u}(\tau, \theta)\|^2}. \quad (11)$$

With $(\tilde{\tau}_{k,i}, \tilde{\theta}_{k,i})$, we obtain an estimate of the channel gain as

$$\tilde{\alpha}_{k,i} = \frac{\mathbf{u}^H(\tilde{\tau}_{k,i}, \tilde{\theta}_{k,i})\mathbf{y}_r}{\|\mathbf{u}(\tilde{\tau}_{k,i}, \tilde{\theta}_{k,i})\|^2}. \quad (12)$$

2. *Newton refinement*: Since the parameter is chosen within a finite grid, Ω , an off-grid error may occur as a result. Then, in pursuit of reducing the off-grid error, we define the Newton step and gradient step as

$$\mathbf{n}_{y_r}(\tilde{\alpha}, \tilde{\tau}, \tilde{\theta}) = -\nabla_{(\tau, \theta)}^2 f_{y_r}^{-1}(\tilde{\alpha}, \tilde{\tau}, \tilde{\theta}) \nabla_{(\tau, \theta)} f_{y_r}(\tilde{\alpha}, \tilde{\tau}, \tilde{\theta}), \quad (13)$$

$$\mathbf{g}_{y_r}(\tilde{\alpha}, \tilde{\tau}, \tilde{\theta}) = \nabla_{(\tau, \theta)} f_{y_r}(\tilde{\alpha}, \tilde{\tau}, \tilde{\theta}). \quad (14)$$

We then refine $(\tilde{\tau}_{k,i}, \tilde{\theta}_{k,i})$ obtained in (11) over the continuum via Newton's method [32], i.e.,

$$\begin{aligned} (\tilde{\tau}_{k,i}, \tilde{\theta}_{k,i}) &= (\tilde{\tau}_{k,i}, \tilde{\theta}_{k,i}) \\ &+ \begin{cases} \mathbf{n}_{y_r}(\tilde{\alpha}_{k,i}, \tilde{\tau}_{k,i}, \tilde{\theta}_{k,i}), & \text{if } \nabla_{(\tau, \theta)}^2 f_{y_r}(\tilde{\alpha}_{k,i}, \tilde{\tau}_{k,i}, \tilde{\theta}_{k,i}) < \mathbf{0}, \\ \mu \mathbf{g}_{y_r}(\tilde{\alpha}, \tilde{\tau}_{k,i}, \tilde{\theta}_{k,i}), & \text{else,} \end{cases} \end{aligned} \quad (15)$$

where μ is the learning rate. After this refinement, the corresponding gain of the i -th path is recalculated as in (12). We cyclically repeat this refinement process R_c times for all the paths found up to this point.

3. *Update of gains*: In this step, by using the estimates $(\tilde{\tau}_{k,\ell}, \tilde{\theta}_{k,\ell}), \forall \ell \in [i]$ obtained in the previous step, we lastly calibrate the gains of all the paths with least squares (LS):

$$[\tilde{\alpha}_{k,1}, \tilde{\alpha}_{k,2}, \dots, \tilde{\alpha}_{k,i}] = \tilde{\mathbf{U}}^\dagger \mathbf{y}_k, \quad (16)$$

where $\tilde{\mathbf{U}} = [\mathbf{u}(\tilde{\tau}_{k,1}, \tilde{\theta}_{k,1}), \mathbf{u}(\tilde{\tau}_{k,2}, \tilde{\theta}_{k,2}), \dots, \mathbf{u}(\tilde{\tau}_{k,i}, \tilde{\theta}_{k,i})]$.

The algorithm terminates once a stopping criterion is met, often based on the false alarm rate. We denote the final algorithm output as $\{\hat{\alpha}_{k,\ell}^{\text{ul}}, \hat{\tau}_{k,\ell}^{\text{ul}}, \hat{\theta}_{k,\ell}^{\text{ul}}\}_{\ell=1, \dots, L_k}$.

Building on this, we let $\{\hat{\alpha}_{k,\ell}^{\text{dl}}, \hat{\tau}_{k,\ell}^{\text{dl}}, \hat{\theta}_{k,\ell}^{\text{dl}}\}_{\ell=1, \dots, L_k} = \{\eta_{k,\ell} \hat{\alpha}_{k,\ell}^{\text{ul}}, \hat{\tau}_{k,\ell}^{\text{ul}}, \hat{\theta}_{k,\ell}^{\text{ul}}\}_{\ell=1, \dots, L_k}$ by incorporating the UL/DL frequency invariance property and leveraging the channel gain model in (4). As shown [18], the 2D-NOMP algorithm achieves near-optimal mean squared error (MSE) performance, approaching the CRLB level.

Finally, we reconstruct the DL channel as

$$\hat{\mathbf{h}}_k(f) = \sum_{\ell=1}^{L_k} \hat{\alpha}_{k,\ell}^{\text{dl}} \mathbf{a}(\hat{\theta}_{k,\ell}^{\text{dl}}; \lambda^{\text{dl}}) e^{-j2\pi f \hat{\tau}_{k,\ell}^{\text{dl}}}. \quad (17)$$

For conciseness, \mathbf{h}_k and $\hat{\mathbf{h}}_k$ will represent the $\mathbf{h}_k(f)$ and $\hat{\mathbf{h}}_k(f)$, respectively, and r_k will represent the $r_k(f)$ hereafter. Before formulating the problem with the estimated channel, we briefly explain user grouping in the following remark.

B. SE Characterization and Problem Formulation

In this subsection, we characterize the SE under the assumption that the frequency difference between UL and DL is f . Then we formulate the main problem of this paper. As shown in [23], [24], the exact form of the instantaneous

SE for both the common and private messages in each fading block, given imperfect CSIT, cannot be expressed in closed form. To resolve this, we derive a useful lower bound on the SE. Let us model the relationship between the true and reconstructed channel as $\mathbf{h}_k = \hat{\mathbf{h}}_k + \mathbf{e}_k$, where \mathbf{e}_k represents the DL channel reconstruction error. Then, we can express the average power of the received signal at user k in (6), normalized by P , (i.e., $\frac{\mathbb{E}\{|r_k|^2\}}{P}$) as

$$\underbrace{|\hat{\mathbf{h}}_k^H \mathbf{f}_c|^2 + |\hat{\mathbf{h}}_k^H \mathbf{f}_k|^2}_{S_c} + \underbrace{|\hat{\mathbf{h}}_k^H \mathbf{f}_k|^2}_{S_k} + \underbrace{\sum_{p \neq k}^K |\hat{\mathbf{h}}_k^H \mathbf{f}_p|^2 + |\mathbf{e}_k^H \mathbf{f}_c|^2 + \sum_{p=1}^K |\mathbf{e}_k^H \mathbf{f}_p|^2}_{I_c} + \frac{\sigma^2}{P}. \quad (18)$$

Subsequently, by treating \mathbf{e}_k as independent Gaussian noise, we characterize a lower bound on the instantaneous SE of s_c as

$$R_c^{\text{ins.}}(k) \stackrel{(a)}{\geq} \mathbb{E}_{\{\mathbf{e}_k | \hat{\psi}_k\}} \left[\log_2 \left(1 + S_c I_c^{-1} \right) \right] \quad (19)$$

$$\stackrel{(b)}{\geq} \log_2 \left(1 + \frac{|\hat{\mathbf{h}}_k^H \mathbf{f}_c|^2}{\left\{ \sum_{p=1}^K |\hat{\mathbf{h}}_k^H \mathbf{f}_p|^2 + \mathbf{f}_c^H \mathbb{E}[\mathbf{e}_k \mathbf{e}_k^H] \mathbf{f}_c + \sum_{p=1}^K \mathbf{f}_p^H \mathbb{E}[\mathbf{e}_k \mathbf{e}_k^H] \mathbf{f}_p + \frac{\sigma^2}{P} \right\}} \right) \quad (20)$$

$$\stackrel{(c)}{=} \log_2 \left(1 + \frac{|\hat{\mathbf{h}}_k^H \mathbf{f}_c|^2}{\left\{ \sum_{p=1}^K |\hat{\mathbf{h}}_k^H \mathbf{f}_p|^2 + \mathbf{f}_c^H \mathbf{\Phi}_k \mathbf{f}_c + \sum_{p=1}^K \mathbf{f}_p^H \mathbf{\Phi}_k \mathbf{f}_p + \frac{\sigma^2}{P} \right\}} \right) \quad (21)$$

$$\triangleq R_c^{\text{lb}}(k).$$

where (a) follows from treating the DL channel reconstruction error as independent Gaussian noise which gives the worst case mutual information [24], (b) follows Jensen's inequality, and (c) comes from $\mathbb{E}_{\{\mathbf{e}_k | \hat{\psi}_k\}}[\mathbf{e}_k \mathbf{e}_k^H] = \mathbf{\Phi}_k$, where we omit the notation of $\{\mathbf{e}_k | \hat{\psi}_k\}$ due to the space limitation. To guarantee the decodability of s_c for the users in $[K]$, the code rate of s_c is determined as $\min_{k \in [K]} \{R_c^{\text{lb}}(k)\}$. Provided that the proper code rate is used, it is guaranteed that s_c is successfully decoded and eliminated with SIC. After SIC, we derive a lower bound on the instantaneous SE of s_k using the similar process as in (21):

$$R_k^{\text{ins.}} \geq \mathbb{E}_{\{\mathbf{e}_k | \hat{\psi}_k\}} \left[\log_2 \left(1 + S_k I_k^{-1} \right) \right] \quad (22)$$

$$\geq \log_2 \left(1 + \frac{|\hat{\mathbf{h}}_k^H \mathbf{f}_k|^2}{\left\{ \sum_{p=1, p \neq k}^K |\hat{\mathbf{h}}_k^H \mathbf{f}_p|^2 + \sum_{p=1}^K \mathbf{f}_p^H \mathbf{\Phi}_k \mathbf{f}_p + \frac{\sigma^2}{P} \right\}} \right) \quad (23)$$

$$\triangleq R_k^{\text{lb}}.$$

Leveraging the derived SE expressions, we formulate the sum SE maximization problem. As explained above, to ensure the decodability of s_c , the code rate of s_c is set

as $\min_{k \in [K]} \{R_c^{\text{lb}}(k)\}$. With this, the sum SE maximization problem is given by

$$\mathcal{P}_1 : \underset{\mathbf{f}_c, \mathbf{f}_1, \dots, \mathbf{f}_K}{\text{maximize}} \quad \min_{k \in [K]} \{R_c^{\text{lb}}(k)\} + \sum_{k=1}^K R_k^{\text{lb}} \quad (24)$$

$$\text{subject to} \quad \|\mathbf{f}_c\|^2 + \sum_{k=1}^K \|\mathbf{f}_k\|^2 \leq 1. \quad (25)$$

Unfortunately, it is infeasible to directly solve (24) due non-smoothness and non-convexity entailed in the objective function. In the next subsection, we present our approach for reformulating (24) into a tractable form.

C. Problem Reformulation

To tackle our main problem of \mathcal{P}_1 , we first transform the objective function in (24) into a matrix form. To do this, we first define an unified precoding vector by stacking all precoders, $\mathbf{f}_c, \mathbf{f}_1, \dots, \mathbf{f}_K$ as

$$\bar{\mathbf{f}} = [\mathbf{f}_c^T, \mathbf{f}_1^T, \dots, \mathbf{f}_K^T]^T \in \mathbb{C}^{N(K+1) \times 1}. \quad (26)$$

With $\bar{\mathbf{f}}$, we express the SINR of s_c as a form of Rayleigh quotient, i.e. $\bar{\mathbf{f}}^H \mathbf{A}_c(k) \bar{\mathbf{f}} / \bar{\mathbf{f}}^H \mathbf{B}_c(k) \bar{\mathbf{f}}$, where

$$\begin{aligned} \mathbf{A}_c(k) &= \text{diag}(\overbrace{(\hat{\mathbf{h}}_k \hat{\mathbf{h}}_k^H + \mathbf{\Phi}_k), \dots, (\hat{\mathbf{h}}_k \hat{\mathbf{h}}_k^H + \mathbf{\Phi}_k)}^{(K+1) \text{ blocks}} \\ &\quad + \mathbf{I}_{N(K+1)} \frac{\sigma^2}{P}, \end{aligned} \quad (27)$$

$$\begin{aligned} \mathbf{B}_c(k) &= \text{diag}(\overbrace{(\mathbf{\Phi}_k, (\hat{\mathbf{h}}_k \hat{\mathbf{h}}_k^H + \mathbf{\Phi}_k), \dots, (\hat{\mathbf{h}}_k \hat{\mathbf{h}}_k^H + \mathbf{\Phi}_k))}^{K \text{ blocks}} \\ &\quad + \mathbf{I}_{N(K+1)} \frac{\sigma^2}{P}. \end{aligned} \quad (28)$$

Similarly, the SINR of s_k is written as $\bar{\mathbf{f}}^H \mathbf{A}_k \bar{\mathbf{f}} / \bar{\mathbf{f}}^H \mathbf{B}_k \bar{\mathbf{f}}$, where

$$\begin{aligned} \mathbf{A}_k &= \text{diag}(\overbrace{(\mathbf{0}, (\hat{\mathbf{h}}_k \hat{\mathbf{h}}_k^H + \mathbf{\Phi}_k), \dots, (\hat{\mathbf{h}}_k \hat{\mathbf{h}}_k^H + \mathbf{\Phi}_k))}^{K \text{ blocks}} \\ &\quad + \mathbf{I}_{N(K+1)} \frac{\sigma^2}{P}, \end{aligned} \quad (29)$$

$$\begin{aligned} \mathbf{B}_k &= \text{diag}(\overbrace{(\mathbf{0}, (\hat{\mathbf{h}}_k \hat{\mathbf{h}}_k^H + \mathbf{\Phi}_k), \dots, (\hat{\mathbf{h}}_k \hat{\mathbf{h}}_k^H + \mathbf{\Phi}_k))}^{(k-1) \text{ blocks}}, \\ &\quad \overbrace{(\mathbf{\Phi}_k, (\hat{\mathbf{h}}_k \hat{\mathbf{h}}_k^H + \mathbf{\Phi}_k), \dots, (\hat{\mathbf{h}}_k \hat{\mathbf{h}}_k^H + \mathbf{\Phi}_k))}^{(K-k) \text{ blocks}}, \\ &\quad + \mathbf{I}_{N(K+1)} \frac{\sigma^2}{P}. \end{aligned} \quad (30)$$

With this reformulation, we exploit the LogSumExp technique to resolve the non-smoothness involved in (24). We approximate the non-smooth minimum function as

$$\min_{i=1, \dots, N} \{x_i\} \approx -\alpha \log \left(\frac{1}{N} \sum_{i=1}^N \exp \left(\frac{x_i}{-\alpha} \right) \right) \quad (31)$$

$$\triangleq g(\{x_i\}_{i \in [N]}), \quad (32)$$

where α determines the accuracy of the approximation. As $\alpha \rightarrow 0$, the approximation becomes tight. Leveraging (31), we approximate the SE of s_c as

$$\min_{k \in [K]} \{R_c^{\text{lb}}(k)\} \approx g \left(\left\{ \frac{\bar{\mathbf{f}}^H \mathbf{A}_c(k) \bar{\mathbf{f}}}{\bar{\mathbf{f}}^H \mathbf{B}_c(k) \bar{\mathbf{f}}} \right\}_{k \in [K]} \right). \quad (33)$$

Using this, we reformulate the original problem \mathcal{P}_1 defined in (24) as

$$\mathcal{P}_2 : \underset{\bar{\mathbf{f}}}{\text{maximize}} \quad g \left(\left\{ \frac{\bar{\mathbf{f}}^H \mathbf{A}_c(k) \bar{\mathbf{f}}}{\bar{\mathbf{f}}^H \mathbf{B}_c(k) \bar{\mathbf{f}}} \right\}_{k \in [K]} \right) + \sum_{k=1}^K \log_2 \left(\frac{\bar{\mathbf{f}}^H \mathbf{A}_k \bar{\mathbf{f}}}{\bar{\mathbf{f}}^H \mathbf{B}_k \bar{\mathbf{f}}} \right). \quad (34)$$

In \mathcal{P}_2 , we drop the constraint (25) since the objective function of \mathcal{P}_2 is scale invariant, which does not affect optimality [24]. Even with this reformulation, however, it is still infeasible to solve \mathcal{P}_2 , since the ECM $\mathbf{\Phi}_k$, $k \in [K]$ has not yet been determined. To address this, we estimate the ECM $\mathbf{\Phi}_k$, $k \in [K]$ in the next section.

IV. ERROR COVARIANCE MATRIX ESTIMATION

To understand the difficulty of obtaining an ECM without CSI feedback, we first summarize the conventional approaches. For simplicity, we let $L_k = 1, \kappa_k = 0$, and $\sigma_{\text{path},k}^2 = 1$ in the following explanation.

- **Bayesian approach:** In this approach, we calculate the ECM by using known channel statistics. Assuming that the statistics of AoD, denoted by θ_k^{dl} , is known at the BS, then we obtain the channel covariance $\mathbb{E}[\mathbf{h}_k \mathbf{h}_k^H] = \mathbf{a}(\theta_k^{\text{dl}}, \lambda^{\text{dl}}) \cdot \mathbf{a}^H(\theta_k^{\text{dl}}, \lambda^{\text{dl}}) = \mathbf{R}_k$. Assuming that linear MMSE estimation [26] is used, the ECM is given as

$$\hat{\mathbf{\Phi}}_k = \mathbf{R}_k - \mathbf{R}_k \left(\mathbf{R}_k + \rho_{\text{pilot}}^2 \mathbf{I}_N \right)^{-1} \mathbf{R}_k, \quad (35)$$

where ρ_{pilot}^2 represents the inverse of the effective SNR of the pilot signals.

- **Frequentist approach:** In this case, instead of using the channel statistics, we assume that the observed error samples are given. Denoting the i -th DL channel error sample as $\mathbf{e}_k(i)$ with $i \in [T]$, the sample ECM is obtained as

$$\hat{\mathbf{\Phi}}_k = \frac{1}{T} \sum_{i=1}^T \mathbf{e}_k(i) \mathbf{e}_k^H(i) + \rho \mathbf{I}, \quad (36)$$

where ρ is a regularization parameter to well-condition $\hat{\mathbf{\Phi}}_k$ [33].

Unfortunately, neither approach is suitable in our case. First, we assume that no channel statistics are known at the BS. Further, it is difficult to analyze the ECM by using 2D-NOMP due to its non-linearity. Additionally, the sample ECM cannot be obtained in our scenario [17].

To address this challenge, we turn our attention to 2D-NOMP's near-CRLB MSE performance as shown in [32]. This result implies that the ECM can be tightly estimated by leveraging the CRLB [5], whose trace gives a lower bound on the MSE. Considering the relationship between the UL

and DL channels and assuming reciprocal path gains in (4), we derive the CRLB [5], [34] as

$$\mathbf{\Phi}_k \geq \mathbf{J}_k^H(f) \mathbf{I}^{-1}(\psi_k) \mathbf{J}_k(f) = \mathbf{C}(f), \quad (37)$$

where ψ_k is the UL channel parameter vector defined as

$$\psi_k = (\bar{\psi}_{k,1}^\top, \dots, \bar{\psi}_{k,L_k}^\top)^\top \in \mathbb{R}^{4L_k \times 1}, \quad (38)$$

$$\bar{\psi}_{k,\ell} = (\theta_{k,\ell}^{\text{ul}}, \tau_{k,\ell}^{\text{ul}}, \text{Re}\{\alpha_{k,\ell}^{\text{ul}}\}, \text{Im}\{\alpha_{k,\ell}^{\text{ul}}\})^\top \in \mathbb{R}^{4 \times 1}. \quad (39)$$

Here, $\mathbf{I}(\psi_k) \in \mathbb{C}^{4L_k \times 4L_k}$ indicates the FIM of user k and $\mathbf{J}_k(f) \in \mathbb{C}^{4L_k \times N}$ denotes the Jacobian matrix of user k , evaluated at the frequency difference f defined as

$$\mathbf{J}_k(f) \triangleq \frac{\partial \mathbf{h}_k^\top(f)}{\partial \psi_k}, \quad (40)$$

where $\mathbf{h}_k(f)$ is in (3).

An implication of (37) is that each diagonal element follows $[\mathbf{\Phi}_k]_{n,n} \geq [\mathbf{C}(f)]_{n,n}, \forall n$. Given that 2D-NOMP exhibits the near-optimal MSE performance [32], the MSE of 2D-NOMP can be closely approximated by the trace of CRLB, i.e., $\mathbb{E}[\|\mathbf{h}_k(f) - \hat{\mathbf{h}}_k(f)\|^2] = \text{tr}\{\mathbf{\Phi}_k\} \approx \text{tr}\{\mathbf{C}(f)\}$. Building on this, we approximate the ECM as a diagonal matrix by using the CRLB, i.e.,

$$\mathbf{\Phi}_k \approx \mathbf{C}(f) \circ \mathbf{I}_N. \quad (41)$$

To use (41) as the estimated ECM², however, we note that the Jacobian and FIM, i.e., $\mathbf{J}_k(f), \mathbf{I}(\psi_k)$ should even be evaluated using the true UL channel parameter ψ_k . This is typically not available in practice [27]. One maybe tempted to use the estimated parameter $\hat{\psi}_k$ to calculate (41), yet this compromises the interpretation of the FIM. To address this, we propose to use the O-FIM [28]. The relation of the O-FIM to the FIM and derivation of the O-FIM are provided in the following remark and lemma, respectively.

Remark 1 (O-FIM and FIM). By definition, the FIM is

$$\mathbf{I}(\psi_k) = \mathbb{E} \left[-\frac{\partial^2 \log f(\mathbf{y}|\psi_k)}{\partial \psi_k \partial \psi_k^\top} \right], \quad (42)$$

where the expectation is taken with respect to the observation \mathbf{y} conditioned on the true parameter ψ_k . On the contrary, the O-FIM is defined as an instantaneous observation of the FIM. This observed information can be interpreted as a sampled version of (42), given by

$$\tilde{\mathbf{I}}(\hat{\psi}_k) = -\frac{\partial^2 \log f(\mathbf{y}|\psi_k)}{\partial \psi_k \partial \psi_k^\top} \Bigg|_{\psi_k = \hat{\psi}_k}, \quad (43)$$

which can be computed straightforwardly with the estimated parameters.

Lemma 1. *Assuming that the likelihood of the observed signal \mathbf{y} follows Gaussian distribution, each element of the*

²We note that our ECM estimation depends only on the diagonal components since its purpose is based on the 2D NOMP algorithm's near-CRLB performance. However, in Section VI, we will show that this approach is still very useful, as the DL channel reconstruction errors are not negligible.

O-FIM in (43), calculated only using the UL estimated parameter, $\hat{\psi}_k$, is

$$[\tilde{\mathbf{I}}(\hat{\psi}_k)]_{u,v} = \frac{2}{\sigma^2} \text{Re} \left\{ \sum_{n=1}^N \sum_{m=1}^M \left(\frac{\partial \bar{\mathbf{y}}_{n,m}^*}{\partial \psi_u} \frac{\partial \bar{\mathbf{y}}_{n,m}}{\partial \psi_v} \right) - \sum_{n=1}^N \sum_{m=1}^M (\mathbf{y}_{n,m} - \bar{\mathbf{y}}_{n,m})^* \frac{\partial^2 \bar{\mathbf{y}}_{n,m}}{\partial \psi_u \partial \psi_v} \right\} \Bigg|_{\psi_k = \hat{\psi}_k}, \quad (44)$$

where $\bar{\mathbf{y}}$ is the reconstructed UL channel with the parameter set ψ_k defined as

$$\bar{\mathbf{y}} \triangleq \sum_{\ell=1}^{L_k} \alpha_{k,\ell}^{\text{ul}} \mathbf{u}(\tau_{k,\ell}^{\text{ul}}, \theta_{k,\ell}^{\text{ul}}) \in \mathbb{C}^{MN \times 1}. \quad (45)$$

Proof. See Appendix A. \square

In [27], [28], it was found that the inverse of the O-FIM serves as a sampled version of the CRLB and is optimal for estimating the realized squared error. That is to say, the inverse of the O-FIM provides the best prediction of the realized error based on the given received signal. Namely, with the Jacobian transformation [5], we have

$$\begin{aligned} & \tilde{\mathbf{J}}_k^H(f) \tilde{\mathbf{I}}^{-1}(\hat{\psi}_k) \tilde{\mathbf{J}}_k(f) \\ &= \arg \min_{\hat{\mathbf{\Phi}}_k} \mathbb{E}_{\{\mathbf{e}_k | \hat{\psi}_k\}} \left[\sum_{m=1}^N \left[\mathbf{e}_k \mathbf{e}_k^H - \hat{\mathbf{\Phi}}_k \right]_{m,m}^2 \right], \end{aligned} \quad (46)$$

where $\tilde{\mathbf{J}}(f)$ is the Jacobian matrix with ψ_k at f and O-FIM $\tilde{\mathbf{I}}(\hat{\psi}_k)$ is obtained from Lemma 1. Applying Jensen's inequality on the right-hand side of (46) and omitting the notation $\{\mathbf{e}_k | \hat{\psi}_k\}$, we obtain

$$\sum_{m=1}^N \left[\mathbb{E} \left[\mathbf{e}_k \mathbf{e}_k^H \right] - \hat{\mathbf{\Phi}}_k \right]_{m,m}^2 \leq \mathbb{E} \left[\sum_{m=1}^N \left[\mathbf{e}_k \mathbf{e}_k^H - \hat{\mathbf{\Phi}}_k \right]_{m,m}^2 \right]. \quad (47)$$

Note that our aim is to estimate the ECM, i.e., $\mathbb{E}[\mathbf{e}_k \mathbf{e}_k^H]$, using the matrix $\hat{\mathbf{\Phi}}_k$. The sum of its diagonal squared errors is represented on the left-hand side of (47). As shown in (47), our solution derived from (46) in fact minimizes an upper bound on the sum of the squared errors in the diagonal elements for the ECM estimation. This supports the use of the O-FIM to estimate diagonal elements of the ECM. We summarize this in the following proposition.

Proposition 1. *The estimated ECM represented in a diagonal form, which minimizes the upper bound of the MSE, can be derived using the O-FIM as follows:*

$$\hat{\mathbf{\Phi}}_k = \left(\tilde{\mathbf{J}}_k^H(f) \tilde{\mathbf{I}}^{-1}(\hat{\psi}_k) \tilde{\mathbf{J}}_k(f) \right) \circ \mathbf{I}_N \triangleq \tilde{\mathbf{C}}(f). \quad (48)$$

We note that $\hat{\mathbf{\Phi}}_k$ in Proposition 1 can be calculated solely by the estimated UL channel parameter $\hat{\psi}_k$, which is valuable in practical scenarios. However, it is worth mentioning that $\hat{\mathbf{\Phi}}_k$ in Proposition 1 assumes a reciprocal path gain, i.e., $\eta_{k,\ell} = 1, \forall k, \ell$ [5]. It is necessary to further calibrate the estimated ECM for a case where $0 \leq \eta_{k,\ell} \leq 1, \forall k, \ell$. We obtain this in the following corollary.

Corollary 1. For general $\eta_{k,\ell}$, the estimated ECM can be generalized to

$$\hat{\Phi}_k = \left(\sum_{\ell=1}^{L_k} \frac{\eta_{k,\ell}^2}{L_k} \right) \tilde{\mathbf{C}}(f) + \left(\sum_{\ell=1}^{L_k} \frac{1 - \eta_{k,\ell}^2}{L_k} \right) \mathbf{I}_n, \quad (49)$$

where $\tilde{\mathbf{C}}(f)$ is given by (48).

Proof. See Appendix B. \square

Using $\hat{\Phi}_k$ in (49), we eventually reach \mathcal{P}_3 as follows:

$$\mathcal{P}_3 : \underset{\bar{\mathbf{f}}}{\text{maximize}} \quad g \left(\left\{ \frac{\bar{\mathbf{f}}^H \hat{\mathbf{A}}_c(k) \bar{\mathbf{f}}}{\bar{\mathbf{f}}^H \hat{\mathbf{B}}_c(k) \bar{\mathbf{f}}} \right\}_{k \in [K]} \right) + \sum_{k=1}^K \log_2 \left(\frac{\bar{\mathbf{f}}^H \hat{\mathbf{A}}_k \bar{\mathbf{f}}}{\bar{\mathbf{f}}^H \hat{\mathbf{B}}_k \bar{\mathbf{f}}} \right), \quad (50)$$

where we define $\hat{\mathbf{A}}_c(k)$, $\hat{\mathbf{B}}_c(k)$, $\hat{\mathbf{A}}_k$, and $\hat{\mathbf{B}}_k$ by replacing every occurrence of Φ_k in (27) through (30) with $\hat{\Phi}_k$. Note that every element in (50) can be configured during UL training. Now, we are ready to develop an algorithm to solve (50).

V. PRECODER OPTIMIZATION

In this section, we propose a precoding optimization method to solve problem (50). Taking inspiration from [24], [35], the proposed method is centered on the following first-order optimality condition.

Theorem 1. For problem (50), the first-order optimality condition is satisfied when the following holds:

$$\mathbf{B}_{\text{KKT}}^{-1}(\bar{\mathbf{f}}) \mathbf{A}_{\text{KKT}}(\bar{\mathbf{f}}) \bar{\mathbf{f}} = \lambda(\bar{\mathbf{f}}) \bar{\mathbf{f}}, \quad (51)$$

where each matrix is shown below.

$$\begin{aligned} \mathbf{A}_{\text{KKT}}(\bar{\mathbf{f}}) &= \lambda_{\text{num}}(\bar{\mathbf{f}}) \times \left[\sum_{k=1}^K \frac{\hat{\mathbf{A}}_k}{\bar{\mathbf{f}}^H \hat{\mathbf{A}}_k \bar{\mathbf{f}}} \right. \\ &\quad \left. + \sum_{k \in [K]} \left(\frac{\exp\left(\frac{1}{-\alpha} \frac{\bar{\mathbf{f}}^H \hat{\mathbf{A}}_c(k) \bar{\mathbf{f}}}{\bar{\mathbf{f}}^H \hat{\mathbf{B}}_c(k) \bar{\mathbf{f}}}\right)}{\sum_{j \in [K]} \exp\left(\frac{1}{-\alpha} \log_2 \left(\frac{\bar{\mathbf{f}}^H \hat{\mathbf{A}}_c(j) \bar{\mathbf{f}}}{\bar{\mathbf{f}}^H \hat{\mathbf{B}}_c(j) \bar{\mathbf{f}}} \right)\right)} \right) \frac{\hat{\mathbf{A}}_c(k)}{\bar{\mathbf{f}}^H \hat{\mathbf{A}}_c(k) \bar{\mathbf{f}}} \right], \end{aligned} \quad (52)$$

$$\begin{aligned} \mathbf{B}_{\text{KKT}}(\bar{\mathbf{f}}) &= \lambda_{\text{den}}(\bar{\mathbf{f}}) \times \left[\sum_{k=1}^K \frac{\hat{\mathbf{B}}_k}{\bar{\mathbf{f}}^H \hat{\mathbf{B}}_k \bar{\mathbf{f}}} \right. \\ &\quad \left. + \sum_{k \in [K]} \left(\frac{\exp\left(\frac{1}{-\alpha} \frac{\bar{\mathbf{f}}^H \hat{\mathbf{A}}_c(k) \bar{\mathbf{f}}}{\bar{\mathbf{f}}^H \hat{\mathbf{B}}_c(k) \bar{\mathbf{f}}}\right)}{\sum_{j \in [K]} \exp\left(\frac{1}{-\alpha} \log_2 \left(\frac{\bar{\mathbf{f}}^H \hat{\mathbf{A}}_c(j) \bar{\mathbf{f}}}{\bar{\mathbf{f}}^H \hat{\mathbf{B}}_c(j) \bar{\mathbf{f}}} \right)\right)} \right) \frac{\hat{\mathbf{B}}_c(k)}{\bar{\mathbf{f}}^H \hat{\mathbf{B}}_c(k) \bar{\mathbf{f}}} \right], \end{aligned} \quad (53)$$

$$\begin{aligned} \lambda(\bar{\mathbf{f}}) &= \prod_{k=1}^K \left(\frac{\bar{\mathbf{f}}^H \hat{\mathbf{A}}_k \bar{\mathbf{f}}}{\bar{\mathbf{f}}^H \hat{\mathbf{B}}_k \bar{\mathbf{f}}} \right) \\ &\quad \times \left\{ \frac{1}{K} \sum_{k \in [K]} \exp\left(\log_2 \left(\frac{\bar{\mathbf{f}}^H \hat{\mathbf{A}}_c(k) \bar{\mathbf{f}}}{\bar{\mathbf{f}}^H \hat{\mathbf{B}}_c(k) \bar{\mathbf{f}}} \right) \right)^{-\frac{1}{\alpha}} \right\}^{-\frac{\alpha}{\log_2 e}} \\ &= \frac{\lambda_{\text{num}}(\bar{\mathbf{f}})}{\lambda_{\text{den}}(\bar{\mathbf{f}})}. \end{aligned} \quad (54)$$

Proof. See Appendix C. \square

We observe that (51) is interpreted as a form of nonlinear eigenvector-dependent eigenvalue problem. The matrix $\mathbf{B}_{\text{KKT}}^{-1}(\bar{\mathbf{f}}) \mathbf{A}_{\text{KKT}}(\bar{\mathbf{f}})$ is a nonlinear matrix function of $\bar{\mathbf{f}}$, and its eigenvalue $\lambda(\bar{\mathbf{f}})$ is set as the objective function of our main problem (50). Crucially, if we find the leading eigenvector, denoted as $\bar{\mathbf{f}}^*$, it naturally maximizes our objective function of (50) while satisfying the condition (51).

To obtain $\bar{\mathbf{f}}^*$, we propose a GPI-based algorithm. To be specific, denoting that $\bar{\mathbf{f}}_{(t)}$ as a precoding vector obtained in t -th iteration, we iteratively update the precoding vector by

$$\bar{\mathbf{f}}_{(t)} \leftarrow \frac{\mathbf{B}_{\text{KKT}}^{-1}(\bar{\mathbf{f}}_{(t-1)}) \mathbf{A}_{\text{KKT}}(\bar{\mathbf{f}}_{(t-1)}) \bar{\mathbf{f}}_{(t-1)}}{\|\mathbf{B}_{\text{KKT}}^{-1}(\bar{\mathbf{f}}_{(t-1)}) \mathbf{A}_{\text{KKT}}(\bar{\mathbf{f}}_{(t-1)}) \bar{\mathbf{f}}_{(t-1)}\|}, \quad (55)$$

where $\mathbf{A}_{\text{KKT}}(\bar{\mathbf{f}}_{(t-1)})$, $\mathbf{B}_{\text{KKT}}(\bar{\mathbf{f}}_{(t-1)})$ are calculated as (52) and (53). Note that this process is repeated until the update of precoder is bounded by predefined parameter, which is $\|\bar{\mathbf{f}}_{(t)} - \bar{\mathbf{f}}_{(t-1)}\| < \epsilon$.

Remark 2. (Rationales for RSMA and extension to NOMA) While we mainly consider RSMA, non-orthogonal multiple access (NOMA) is another multiple access technique robust with MUI. Nonetheless, in MIMO-NOMA, it is difficult to find the optimal SIC order because each user's SE is determined not only by its channel strength, but also by a complicated function of channel direction, transmit power, and precoding vectors [22]. For this reason, as shown in [36], MIMO-NOMA achieves mediocre SE, even lower than that of conventional SDMA. In contrast, RSMA achieves high SE, particularly in imperfect CSIT scenarios. This is thanks to its unique message construction, by which SIC ordering is simplified [22]. This is a primary reason that we choose RSMA instead of NOMA to mitigate MUI. However, we note that our approach also can be extended to MIMO-NOMA, which presents interesting future work.

VI. SIMULATION RESULTS

In this section, we numerically examine the actual MSE of 2D-NOMP in DL channel reconstruction, compare it with the expected MSE from our ECM estimation. We also present the ergodic sum SE performance comparison between the proposed method and other baseline methods in various environments.

A. MSE of DL Channel Reconstruction

We first define the MSE of estimated DL channel of user k , i.e. $\hat{\mathbf{h}}_k(f)$ as

$$\text{MSE} = \mathbb{E} [\|\mathbf{h}_k(f) - \hat{\mathbf{h}}_k(f)\|^2]. \quad (56)$$

Based on this, we plot the actual MSE (56) of the 2D-NOMP algorithm as a function of the frequency difference f in Fig. 1. We then compare the analytical MSE obtained by the CRLB and by our ECM estimation in (48). To yield the MSE from the CRLB and estimated ECM, we compute $\text{tr}\{\mathbf{C}(f)\}$ and $\text{tr}\{\hat{\Phi}_k\}$ by following [5].

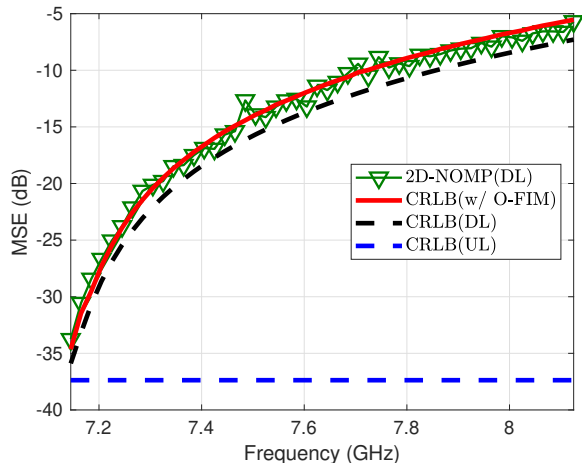


Fig. 1: MSE over DL carrier frequency. The UL carrier frequency is 7.15 GHz and UL SNR is 10 dB with $M = 128$.

In Fig. 1, we first observe that the actual MSE of 2D-NOMP is close to the level of CRLB, i.e., $\text{tr}\{\mathbf{C}(f)\}$. This indicates that 2D-NOMP achieves near-optimal MSE performance in DL channel reconstruction, which is consistent with findings in [32]. Notably, we find that the expected MSE from our proposed ECM estimation $\text{tr}\{\hat{\Phi}_k\}$ is closer to the actual MSE of the 2D NOMP estimator than the original CRLB case. This result justifies the use of O-FIM for ECM estimation, not only because of its practical applicability but also its accuracy.

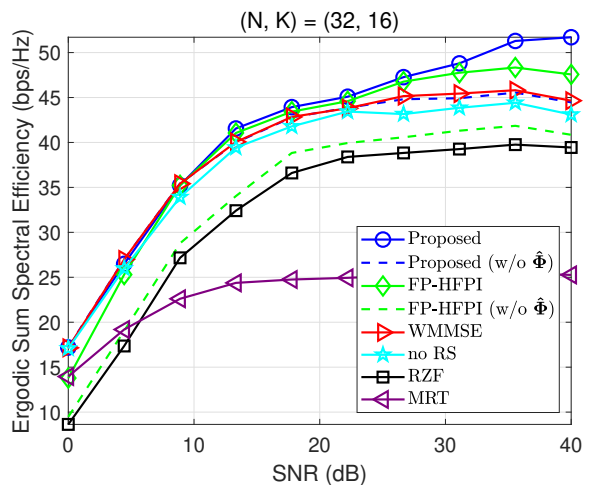
B. Ergodic Sum SE

We now investigate the ergodic sum SE of the proposed method through numerical simulations. For the simulation environments, we follow the system model laid forth in Section II. As baseline methods, we consider the following. We assume that the reconstructed DL channel (17) is used for each baseline precoding method.

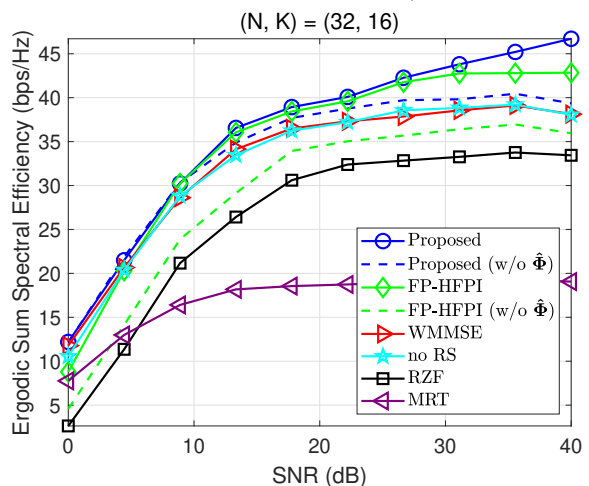
- **FP-HFPI**: As a state-of-the-art RSMA precoding method, we consider the FP-HFPI method, wherein the precoder optimization is performed based on FP [26]. For a comprehensive comparison, we incorporate our ECM estimation $\hat{\Phi}_k$ into the FP-HFPI method.
- **WMMSE**: As a state-of-the-art SDMA precoding optimization method, we consider WMMSE [37].
- **no RS**: We consider the precoding method that maximizes the sum SE by adopting classical SDMA [35].
- **MRT**: The precoders are aligned by the estimated channel vector, i.e. $\mathbf{f}_k = \hat{\mathbf{h}}_k$, $k \in \mathcal{K}$.
- **RZF**: The precoders are designed in a regularized zero-forcing (RZF) fashion as

$$\mathbf{f}_k = \left(\hat{\mathbf{H}}\hat{\mathbf{H}}^H + \frac{\sigma^2}{P}\mathbf{I} \right)^{-1} \hat{\mathbf{h}}_k^H, \quad (57)$$

where $\hat{\mathbf{H}}$ denotes the matrix formed by stacking all the users' estimated channels [38].



(a) Reciprocal channel gain ($\eta_{k,\ell}^2 = 1, \forall k, \ell$).



(b) Non-reciprocal channel gain ($\eta_{k,\ell}^2 \sim \text{Unif}(\mu, 1), \forall k, \ell$).

Fig. 2: Comparison of the ergodic sum SE over transmit SNR, i.e., P/σ^2 with $(N, K) = (32, 16)$ and $M = 128$. The UL carrier frequency is 7.25 GHz, and DL carrier frequency is 7.75 GHz. We use $\epsilon = 0.1$, $\alpha = 0.1$ and $\mu = 0.9$. For the cases without $\hat{\Phi}$, we assume $\hat{\Phi} = \mathbf{0}$.

In Fig. 2, we show the ergodic sum SE for the proposed method and the baseline methods, assuming $\eta_{k,\ell}^2 = 1, \forall k, \ell$ for Fig. 2-(a) and $\eta_{k,\ell}^2 \sim \text{Unif}(\mu, 1), \forall k, \ell$ for Fig. 2-(b). First, we observe that the RSMA approaches offer significant SE gains over SDMA approaches, especially in mid-to-high SNR regimes. For instance, with reciprocal channel gain, the proposed method improves SE by up to 15.8% over the SDMA method in [35]. In the non-reciprocal channel gain case, which causes inaccurate CSIT and typically reduces SE performance, the performance gap between our method and [35] increases up to 22.7% at SNR of 40 dB, demonstrating the effective MUI mitigation capability of RSMA.

Compared to FP-HFPI, our method achieves better sum SE performance at low and high SNR regimes. Especially,

TABLE I: Relative performance compared with perfect CSIT

(a) $(N, K) = (64, 32)$				
Number of paths	1	4	7	10
Proposed (%)	96.01	93.00	89.00	83.63
no RS (%)	95.12	85.50	76.12	68.84
(b) $(N, K) = (16, 8)$				
Number of paths	1	4	7	10
Proposed (%)	86.95	79.63	76.01	73.73
no RS (%)	86.51	65.36	58.34	50.98

Percentage of the sum SE performance compared to WMMSE under the assumption of perfect CSIT [37], for different numbers of channel paths at an SNR of 20 dB.

at SNR of 40 dB in Fig. 2-(b), the proposed method attains around 9% improvement in sum SE over FP-HFPI. We interpret these gains as stemming from the proposed method's ability to capture the impacts of the ECM. That is to say, our method exploits the SINR expression in its original form, while FP-HFPI alters it via a quadratic transform [26]. This may allow the proposed method to more suitably reflect the impacts of the ECM into the precoder design. It is crucial to note that, this level of SE performance for FP-HFPI is only achievable when our ECM estimation is properly incorporated. If FP-HFPI does not rely on our ECM estimation, Fig. 2-(b) shows that the performance gap between the proposed method and FP-HFPI increases to 23.2% at SNR of 40dB. At other SNR levels in Fig. 2-(a), (b), FP-HFPI suffers from significant SE degradation when the ECM estimation is not used.

These observations indicate that the ECM is crucial for achieving robust SE. For example, in Fig. 2-(b), incorporating the ECM into the proposed precoder design results in 21.1% improvement in terms of sum SE at the SNR of 40 dB. This highlights the need for careful handling of CSIT errors to fully realize RSMA's potential.

C. Robustness in Various Channel Environments

We explore the robustness of the proposed method in various settings. In Table I, we compare the sum SE achieved by our approach versus WMMSE with perfect CSIT [37]. Since our approach does not rely on any CSI feedback, Table I implies the inevitable performance loss resulting from its absence. In general, the performance gap decreases as (N, K) increases and the number of paths decreases. This is reasonable because, as N increases, it enables the accurate parameter estimation during UL training, which is effective in mitigating MUI. Similarly, as the number of channel paths decreases, the channels from different users can become nearly orthogonal, which reduces the MUI and thereby increases the sum SE. Given that the number of dominant DL paths is typically small (~ 4) as demonstrated in [31], and also perfect CSIT is not feasible even by using CSI feedback, the actual performance gap in practice would be much less than 7% for $(N, K) = (64, 32)$.

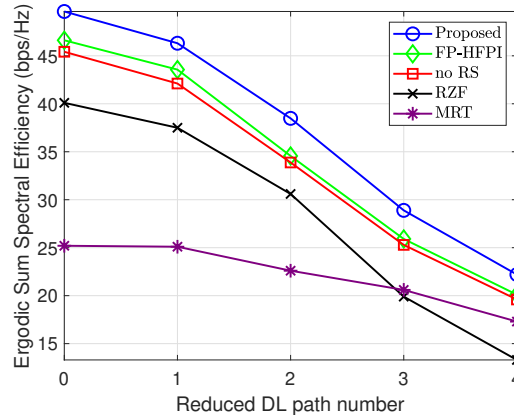


Fig. 3: Comparison of the ergodic sum SE over reduced DL path number with SNR = 30 dB. It is assumed that the UL channel consists of 5 paths and other simulation settings are identical to those of Fig. 2-(b).

Through the comparison, we provide a trade-off guideline between the latency reduction from saving CSI feedback and the corresponding loss of sum SE. In conventional DL transmission within FDD systems, the users estimate the DL channel using reference signals (e.g., CSI-RS) and send feedback (e.g., PMI, RI, SINR) to the BS. In this process, the primary latency arises from CSI report computation on the user side, which can take up to 6 ms due to limited computational resources ([39], Table 5.4-2). This is much longer than other steps, such as CSI-RS scheduling (~ 1 ms), CSI feedback (~ 1 ms), and decoding (~ 1 ms). Using our approach, we shift this CSI processing overhead from the user side to the BS (see Fig. 4). This can save approximately 6-7 ms, while achieving more than 93% of the sum SE in $(N, K) = (64, 32)$ assuming 4 dominant paths [31]. As mentioned before, since perfect CSIT cannot be guaranteed even with conventional CSI feedback [14], the de facto performance loss of the proposed approach in sum SE is well below 7%. As a result, our approach offers a low-latency MIMO transmission framework that significantly reduces latency with only a slight impact on sum SE. We illustrate the comparison of CSI acquisition processes in Fig. 4.

Fig. 3 shows the sum SE performance when the number of DL channel path reduces, which occurs due to a higher carrier frequency of the DL band in FDD [3]. This discrepancy between actual channel path numbers between UL and DL naturally leads to performance degradation, as it would result in inaccurate channel reconstruction. For all the cases of reduced DL path, the proposed method provides robust SE performance compared to other SDMA methods. Similar to the above observation, this robustness of our method is attributed to the proper ECM estimation.

VII. CONCLUSION

In this paper, we have proposed a novel method to achieve robust SE performance in FDD massive MIMO systems without the need for CSI feedback. In order to effectively

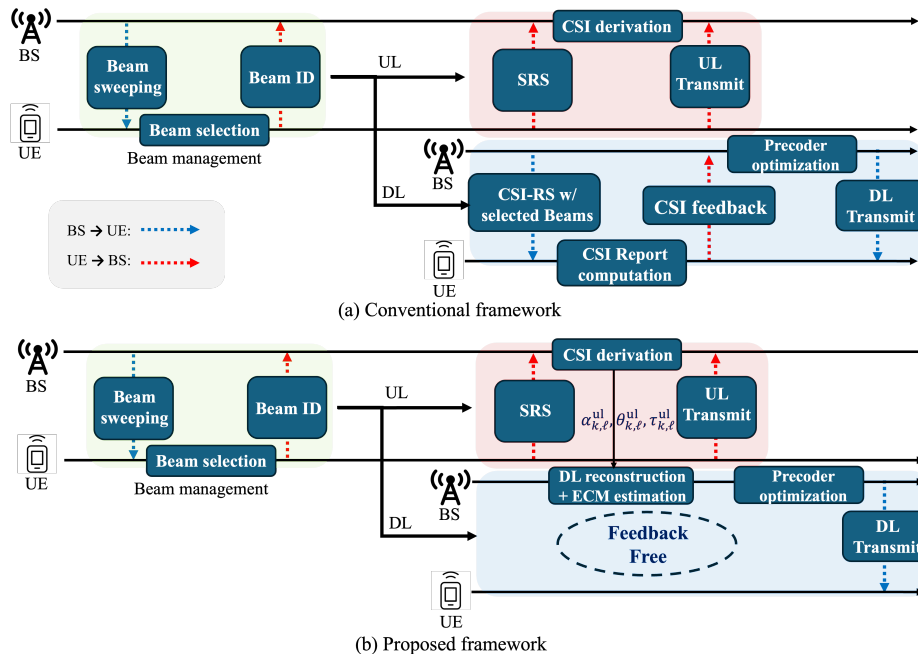


Fig. 4: Comparison of conventional FDD MU-MIMO DL CSI acquisition process vs. the proposed CSI acquisition process.

manage MUI caused by imperfect DL channel reconstruction, our method uses RSMA. This approach requires carefully integrating the CSIT error into the precoder design, with the proposed ECM estimation being crucial. Our major findings are as follows. i) The proposed ECM estimation using the O-FIM accurately captures DL CSIT reconstruction errors. ii) The proposed method significantly improves the sum SE performance over the existing methods. In particular, accurate ECM estimation is crucial for obtaining these gains. iii) Our approach offers a low-latency FDD massive MIMO transmission framework, achieving a favorable trade-off between significant latency reduction and a slight loss in sum SE. As future work, it is promising to extend our approach by considering various system environments, such as energy efficiency maximization, max-min fairness, and integrated sensing and communications.

APPENDIX A

PROOF OF LEMMA 1

Let us assume each measurement $\mathbf{y}_{n,m}$ at each antenna and sub-carrier ($n \in [N], m \in [M]$) is drawn under independent and identically distributed with likelihood $f(\mathbf{y}_{n,m}|\psi_k)$, and its distribution follows Gaussian distribution with covariance $\sigma^2\mathbf{I}$ and mean $\bar{\mathbf{y}}$. Then, the log-likelihood of the parameters ψ_k given measured signal can be represented by

$$\mathcal{L}(\psi_k|\mathbf{y}_{n,m}, n \in [N], m \in [M]) \triangleq \sum_{n=1}^N \sum_{m=1}^M \log f(\mathbf{y}_{n,m}|\psi_k) \quad (58)$$

$$= -\frac{1}{\sigma^2}(\mathbf{y} - \bar{\mathbf{y}})^H(\mathbf{y} - \bar{\mathbf{y}}) + C, \quad (59)$$

where C is a constant. Then, the observed information defined as the negative Hessian of log-likelihood is evaluated at channel parameters $\hat{\psi}_k$, as follows [28]:

$$\mathbf{I}(\hat{\psi}_k) = -\nabla\nabla^H \mathcal{L}(\psi_k)|_{\psi_k=\hat{\psi}_k} \quad (60)$$

$$= - \begin{bmatrix} \frac{\partial^2}{\partial \psi_{1,1}^2} & \frac{\partial^2}{\partial \psi_{1,1} \partial \psi_{1,2}} & \cdots & \frac{\partial^2}{\partial \psi_{1,1} \partial \psi_{L,4}} \\ \frac{\partial^2}{\partial \psi_{1,2} \partial \psi_{1,1}} & \frac{\partial^2}{\partial \psi_{1,2}^2} & \cdots & \frac{\partial^2}{\partial \psi_{1,2} \partial \psi_{L,4}} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial^2}{\partial \psi_{L,4} \partial \psi_{1,1}} & \frac{\partial^2}{\partial \psi_{L,4} \partial \psi_{1,2}} & \cdots & \frac{\partial^2}{\partial \psi_{L,4}^2} \end{bmatrix} \mathcal{L}(\psi_k) \Big|_{\psi_k=\hat{\psi}_k}, \quad (61)$$

where $\mathcal{L}(\psi_k)$ denotes $\mathcal{L}(\psi_k|\mathbf{y}_{n,m}, n \in [N], m \in [M])$ for simplicity and $\mathcal{L}(\psi_k)|_{\psi_k=\hat{\psi}_k}$ follows the normal distribution as well. To evaluate (u, v) component of $\mathbf{I}(\hat{\psi}_k)$, we compute

$$\begin{aligned} \frac{\partial \mathcal{L}(\psi_k)}{\partial \psi_v} &= \frac{1}{\sigma^2} \sum_{n=1}^N \sum_{m=1}^M \left(\frac{\partial \bar{\mathbf{y}}_{n,m}}{\partial \psi_v} \right)^* (\mathbf{y}_{n,m} - \bar{\mathbf{y}}_{n,m}) \\ &\quad + \frac{1}{\sigma^2} \sum_{n=1}^N \sum_{m=1}^M (\mathbf{y}_{n,m} - \bar{\mathbf{y}}_{n,m})^* \left(\frac{\partial \bar{\mathbf{y}}_{n,m}}{\partial \psi_v} \right), \\ &= \frac{2}{\sigma^2} \text{Re} \left\{ \sum_{n=1}^N \sum_{m=1}^M (\mathbf{y}_{n,m} - \bar{\mathbf{y}}_{n,m})^* \left(\frac{\partial \bar{\mathbf{y}}_{n,m}}{\partial \psi_v} \right) \right\}, \end{aligned}$$

where $\bar{\mathbf{y}}_{n,m}$ denotes the reconstructed signal at n -th antenna and s -th sub-carrier. Then, the derivative with respect to ψ_u

can be found as follows

$$\begin{aligned} \frac{\partial}{\partial \psi_u} \left(\frac{\partial \mathcal{L}(\psi_k)}{\partial \psi_v} \right) &= \frac{2}{\sigma^2} \text{Re} \left\{ \sum_{n=1}^N \sum_{m=1}^M \left(-\frac{\partial \bar{\mathbf{y}}_{n,m}^*}{\partial \psi_u} \frac{\partial \bar{\mathbf{y}}_{n,m}}{\partial \psi_v} \right) \right. \\ &\quad \left. + \sum_{n=1}^N \sum_{m=1}^M (\mathbf{y}_{n,m} - \bar{\mathbf{y}}_{n,m})^* \frac{\partial^2 \bar{\mathbf{y}}_{n,m}}{\partial \psi_u \partial \psi_v} \right\}. \end{aligned} \quad (62)$$

Substituting (62) into (61) completes the proof.

APPENDIX B

PROOF OF COROLLARY 1

Consider first the basic case of $\alpha_{k,\ell}^{\text{dl}} = \alpha_{k,\ell}^{\text{ul}}, \forall k, \ell$, then we would represent the ECM of reconstructed channel in (3) as

$$\begin{aligned} \Phi_k \triangleq \mathbb{E}_{\{\mathbf{e}_k, \hat{\psi}_k\}} \left[\left(\sum_{\ell=1}^{L_k} \alpha_{k,\ell}^{\text{dl}} \bar{\mathbf{u}}_{k,\ell} - \sum_{\ell=1}^{L_k} \hat{\alpha}_{k,\ell}^{\text{dl}} \hat{\mathbf{u}}_{k,\ell} \right) \right. \\ \left. \times \left(\sum_{\ell=1}^{L_k} \alpha_{k,\ell}^{\text{dl}} \bar{\mathbf{u}}_{k,\ell} - \sum_{\ell=1}^{L_k} \hat{\alpha}_{k,\ell}^{\text{dl}} \hat{\mathbf{u}}_{k,\ell} \right)^{\text{H}} \right], \end{aligned} \quad (64)$$

where we define $\bar{\mathbf{u}}_{k,\ell}$ as

$$\bar{\mathbf{u}}_{k,\ell} = \mathbf{a} \left(\theta_{k,\ell}^{\text{dl}}; \lambda^{\text{dl}} \right) e^{-j2\pi f \tau_{k,\ell}} \in \mathbb{C}^{N \times 1}, \quad (65)$$

which represents the sampled signature of $\mathbf{u}_{k,\ell}$ from (7) at frequency f , and let $\hat{\mathbf{u}}$ denote its estimate. Recall that this can be obtainable via (48). Based on this, it can be extended to the general case: $\alpha_{k,\ell}^{\text{dl}} = \eta_{k,\ell} \alpha_{k,\ell}^{\text{ul}} + \sqrt{1 - \eta_{k,\ell}^2} \beta_{k,\ell}$, where $\beta_{k,\ell} \sim \mathcal{CN}(0, \sigma_{\text{path},k}^2)$ and independent of $\alpha_{k,\ell}^{\text{ul}}, \hat{\alpha}_{k,\ell}^{\text{ul}}$. Then, we possibly develop the new ECM as follows:

$$\begin{aligned} \mathbb{E} \left[\left(\sum_{\ell=1}^{L_k} (\eta_{k,\ell} \alpha_{k,\ell}^{\text{ul}} + \sqrt{1 - \eta_{k,\ell}^2} \beta_{k,\ell}) \bar{\mathbf{u}}_{k,\ell} - \sum_{\ell=1}^{L_k} \eta_{k,\ell} \hat{\alpha}_{k,\ell}^{\text{ul}} \hat{\mathbf{u}}_{k,\ell} \right) \right. \\ \left. \times \left(\sum_{\ell=1}^{L_k} (\eta_{k,\ell} \alpha_{k,\ell}^{\text{ul}} + \sqrt{1 - \eta_{k,\ell}^2} \beta_{k,\ell}) \bar{\mathbf{u}}_{k,\ell} - \sum_{\ell=1}^{L_k} \eta_{k,\ell} \hat{\alpha}_{k,\ell}^{\text{ul}} \hat{\mathbf{u}}_{k,\ell} \right)^{\text{H}} \right], \end{aligned} \quad (66)$$

where we assumed that $\hat{\alpha}^{\text{dl}} = \eta_{k,\ell} \hat{\alpha}_{k,\ell}^{\text{ul}}$ using the correlation knowledge. To gain further insight, we can consider the following assumptions by examining (66), i.e.,

$$\begin{aligned} \mathbb{E} \left[\left(\alpha_{k,\ell}^{\text{ul}} \bar{\mathbf{u}}_{k,\ell} - \hat{\alpha}_{k,\ell}^{\text{ul}} \hat{\mathbf{u}}_{k,\ell} \right) \left(\alpha_{k',\ell'}^{\text{ul}} \bar{\mathbf{u}}_{k',\ell'} - \hat{\alpha}_{k',\ell'}^{\text{ul}} \hat{\mathbf{u}}_{k',\ell'} \right)^{\text{H}} \right] \approx 0, \\ \forall k \neq k' \text{ or } \ell \neq \ell', \end{aligned} \quad (67)$$

which implies no error correlation between different user or path. In addition, by leveraging the relationship between $\beta_{k,\ell}$ and $\alpha_{k,\ell}, \hat{\alpha}_{k,\ell}$, (66) can be developed by

$$\begin{aligned} \mathbb{E} \left[\sum_{\ell=1}^{L_k} \eta_{k,\ell}^2 (\alpha_{k,\ell}^{\text{ul}} \bar{\mathbf{u}}_{k,\ell} - \hat{\alpha}_{k,\ell}^{\text{ul}} \hat{\mathbf{u}}_{k,\ell}) (\alpha_{k,\ell}^{\text{ul}} \bar{\mathbf{u}}_{k,\ell} - \hat{\alpha}_{k,\ell}^{\text{ul}} \hat{\mathbf{u}}_{k,\ell})^{\text{H}} \right. \\ \left. + \sum_{\ell=1}^{L_k} (1 - \eta_{k,\ell}^2) \beta_{k,\ell} \beta_{k,\ell}^{\text{H}} \bar{\mathbf{u}}_{k,\ell} \bar{\mathbf{u}}_{k,\ell}^{\text{H}} \right] \\ = \left(\sum_{\ell=1}^{L_k} \frac{\eta_{k,\ell}^2}{L_k} \right) \Phi_k + \left(\sum_{\ell=1}^{L_k} \frac{1 - \eta_{k,\ell}^2}{L_k} \right) \mathbf{I}_n, \end{aligned} \quad (68)$$

where $\sigma_{\text{path},k}^2 = \frac{1}{NL_k}$ by normalization. Since the underlying ECM given channel model follows (68), we replace Φ_k with $\tilde{\mathbf{C}}(f)$ to estimate the ECM. This completes the proof.

APPENDIX C

PROOF OF THEOREM 1

We first explore the KKT condition of our main problem (50). In this, we develop the Lagrangian function as

$$\begin{aligned} L(\bar{\mathbf{f}}) = \log \left(\frac{1}{K} \sum_{k \in [K]} \exp \left(\log_2 \left(\frac{\bar{\mathbf{f}}^{\text{H}} \mathbf{A}_c(k) \bar{\mathbf{f}}}{\bar{\mathbf{f}}^{\text{H}} \mathbf{B}_c(k) \bar{\mathbf{f}}} \right) \right)^{-\frac{1}{\alpha}} \right)^{-\alpha} \\ + \sum_{k=1}^K \log_2 \left(\frac{\bar{\mathbf{f}}^{\text{H}} \mathbf{A}_k \bar{\mathbf{f}}}{\bar{\mathbf{f}}^{\text{H}} \mathbf{B}_k \bar{\mathbf{f}}} \right). \end{aligned} \quad (69)$$

Then, we take the partial derivatives of $L(\bar{\mathbf{f}})$ in terms of $\bar{\mathbf{f}}$ to find a stationary point, where we set it as zero. For notational simplicity, we denote the first term of (69) as $L_1(\bar{\mathbf{f}})$ and second term as $L_2(\bar{\mathbf{f}})$ respectively, and each partial derivative is obtained respectively, as follows

$$\begin{aligned} \frac{\partial L_1(\bar{\mathbf{f}})}{\partial \bar{\mathbf{f}}^{\text{H}}} = \sum_{k \in [K]} \left(\frac{\exp \left(\frac{1}{-\alpha} \frac{\bar{\mathbf{f}}^{\text{H}} \mathbf{A}_c(k) \bar{\mathbf{f}}}{\bar{\mathbf{f}}^{\text{H}} \mathbf{B}_c(k) \bar{\mathbf{f}}} \right)}{\sum_{j \in [K]} \exp \left(\frac{1}{-\alpha} \log_2 \left(\frac{\bar{\mathbf{f}}^{\text{H}} \mathbf{A}_c(j) \bar{\mathbf{f}}}{\bar{\mathbf{f}}^{\text{H}} \mathbf{B}_c(j) \bar{\mathbf{f}}} \right) \right)} \right) \\ \times \partial \left(\log_2 \left(\frac{\bar{\mathbf{f}}^{\text{H}} \mathbf{A}_c(k) \bar{\mathbf{f}}}{\bar{\mathbf{f}}^{\text{H}} \mathbf{B}_c(k) \bar{\mathbf{f}}} \right) \right) / \partial \bar{\mathbf{f}}^{\text{H}} \end{aligned} \quad (70)$$

$$\begin{aligned} = \frac{1}{\log 2} \sum_{i=1}^G \left\{ \sum_{k \in [K]} \left(\frac{\exp \left(\frac{1}{-\alpha} \frac{\bar{\mathbf{f}}^{\text{H}} \mathbf{A}_c(k) \bar{\mathbf{f}}}{\bar{\mathbf{f}}^{\text{H}} \mathbf{B}_c(k) \bar{\mathbf{f}}} \right)}{\sum_{j \in [K]} \exp \left(\frac{1}{-\alpha} \log_2 \left(\frac{\bar{\mathbf{f}}^{\text{H}} \mathbf{A}_c(j) \bar{\mathbf{f}}}{\bar{\mathbf{f}}^{\text{H}} \mathbf{B}_c(j) \bar{\mathbf{f}}} \right) \right)} \right) \right\} \\ \times \left\{ \frac{\mathbf{A}_c(k) \bar{\mathbf{f}}}{\bar{\mathbf{f}}^{\text{H}} \mathbf{A}_c(k) \bar{\mathbf{f}}} - \frac{\mathbf{B}_c(k) \bar{\mathbf{f}}}{\bar{\mathbf{f}}^{\text{H}} \mathbf{B}_c(k) \bar{\mathbf{f}}} \right\}, \end{aligned} \quad (71)$$

$$\frac{\partial L_2(\bar{\mathbf{f}})}{\partial \bar{\mathbf{f}}^{\text{H}}} = \frac{1}{\log 2} \sum_{k=1}^K \left[\frac{\mathbf{A}_k \bar{\mathbf{f}}}{\bar{\mathbf{f}}^{\text{H}} \mathbf{A}_k \bar{\mathbf{f}}} - \frac{\mathbf{B}_k \bar{\mathbf{f}}}{\bar{\mathbf{f}}^{\text{H}} \mathbf{B}_k \bar{\mathbf{f}}} \right]. \quad (72)$$

It is obvious that when the sum of both terms is zero, stationarity for first-order KKT condition is met, which is

$$\frac{\partial L_1(\bar{\mathbf{f}})}{\partial \bar{\mathbf{f}}^{\text{H}}} + \frac{\partial L_2(\bar{\mathbf{f}})}{\partial \bar{\mathbf{f}}^{\text{H}}} = 0 \quad (73)$$

$$\Leftrightarrow \sum_{k \in [K]} \left(\frac{\exp \left(\frac{1}{-\alpha} \frac{\bar{\mathbf{f}}^{\text{H}} \mathbf{A}_c(k) \bar{\mathbf{f}}}{\bar{\mathbf{f}}^{\text{H}} \mathbf{B}_c(k) \bar{\mathbf{f}}} \right)}{\sum_{j \in [K]} \exp \left(\frac{1}{-\alpha} \log_2 \left(\frac{\bar{\mathbf{f}}^{\text{H}} \mathbf{A}_c(j) \bar{\mathbf{f}}}{\bar{\mathbf{f}}^{\text{H}} \mathbf{B}_c(j) \bar{\mathbf{f}}} \right) \right)} \right) \quad (74)$$

$$\times \left\{ \frac{\mathbf{A}_c(k) \bar{\mathbf{f}}}{\bar{\mathbf{f}}^{\text{H}} \mathbf{A}_c(k) \bar{\mathbf{f}}} - \frac{\mathbf{B}_c(k) \bar{\mathbf{f}}}{\bar{\mathbf{f}}^{\text{H}} \mathbf{B}_c(k) \bar{\mathbf{f}}} \right\} \quad (75)$$

$$+ \sum_{k=1}^K \left[\frac{\mathbf{A}_k \bar{\mathbf{f}}}{\bar{\mathbf{f}}^{\text{H}} \mathbf{A}_k \bar{\mathbf{f}}} - \frac{\mathbf{B}_k \bar{\mathbf{f}}}{\bar{\mathbf{f}}^{\text{H}} \mathbf{B}_k \bar{\mathbf{f}}} \right] = 0. \quad (76)$$

With equation (52), (53), (54), we can arrange the first-order KKT condition as

$$\mathbf{A}_{\text{KKT}}(\bar{\mathbf{f}})\bar{\mathbf{f}} = \lambda(\bar{\mathbf{f}})\mathbf{B}_{\text{KKT}}(\bar{\mathbf{f}})\bar{\mathbf{f}} \Leftrightarrow \mathbf{B}_{\text{KKT}}(\bar{\mathbf{f}})^{-1}\mathbf{A}_{\text{KKT}}(\bar{\mathbf{f}})\bar{\mathbf{f}} \quad (77)$$

$$= \lambda(\bar{\mathbf{f}})\bar{\mathbf{f}}. \quad (78)$$

This completes the proof.

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, 2014.
- [2] H. Viswanathan and P. E. Mogensen, "Communications in the 6G era," *IEEE Access*, vol. 8, pp. 57 063–57 074, 2020.
- [3] J. Flordelis, F. Rusek, F. Tufvesson, E. G. Larsson, and O. Edfors, "Massive MIMO performance—TDD versus FDD: What do measurements say?" *IEEE Trans. Wireless Commun.*, vol. 17, no. 4, pp. 2247–2261, 2018.
- [4] T. Choi, F. Rottenberg, J. Gómez-Ponce, A. Ramesh, P. Luo, C. J. Zhang, and A. F. Molisch, "Experimental investigation of frequency domain channel extrapolation in massive MIMO systems for zero-feedback FDD," *IEEE Trans. Wireless Commun.*, vol. 20, no. 1, pp. 710–725, 2021.
- [5] F. Rottenberg, T. Choi, P. Luo, C. J. Zhang, and A. F. Molisch, "Performance analysis of channel extrapolation in FDD massive MIMO systems," *IEEE Trans. Wireless Commun.*, vol. 19, no. 4, pp. 2728–2741, Apr. 2020.
- [6] J. Choi, D. J. Love, and P. Bidigare, "Downlink training techniques for FDD massive MIMO systems: Open-loop and closed-loop training with memory," *IEEE J. Sel. Topics Signal Process.*, vol. 8, no. 5, pp. 802–814, 2014.
- [7] B. Lee, J. Choi, J.-Y. Seol, D. J. Love, and B. Shim, "Antenna grouping based feedback compression for FDD-based massive MIMO systems," *IEEE Trans. Commun.*, vol. 63, no. 9, pp. 3261–3274, 2015.
- [8] X. Rao and V. K. N. Lau, "Distributed compressive CSIT estimation and feedback for FDD multi-user massive MIMO systems," *IEEE Trans. Signal Process.*, vol. 62, no. 12, pp. 3261–3271, 2014.
- [9] Y. Han, J. Lee, and D. J. Love, "Compressed sensing-aided downlink channel training for FDD massive MIMO systems," *IEEE Trans. Commun.*, vol. 65, no. 7, pp. 2852–2862, 2017.
- [10] B. Wang, M. Jian, F. Gao, G. Y. Li, and H. Lin, "Beam squint and channel estimation for wideband mmWave massive MIMO-OFDM systems," *IEEE Trans. Signal Process.*, vol. 67, no. 23, pp. 5893–5908, 2019.
- [11] P. Liang, J. Fan, W. Shen, Z. Qin, and G. Y. Li, "Deep learning and compressive sensing-based CSI feedback in FDD massive MIMO systems," *IEEE Trans. Veh. Technol.*, vol. 69, no. 8, pp. 9217–9222, 2020.
- [12] F. G. X. Ma, Z. Gao and M. D. Renzo, "Model-driven deep learning based channel estimation and feedback for millimeter-wave massive hybrid MIMO systems," *IEEE J. Sel. Areas Commun.*, vol. 39, no. 8, pp. 2388–2406, 2021.
- [13] M. Alrabeiah and A. Alkhateeb, "Deep learning for TDD and FDD massive MIMO: Mapping channels in space and frequency," in *Proc. of Asilomar Conf. on Sign., Syst. and Computers*, 2019, pp. 1465–1470.
- [14] J. Guo, C.-K. Wen, S. Jin, and G. Y. Li, "Overview of deep learning-based CSI feedback in massive MIMO systems," *IEEE Trans. Commun.*, vol. 70, no. 12, pp. 8017–8045, 2022.
- [15] D. Vasishth, S. Kumar, H. Rahul, and D. Katabi, "Eliminating channel feedback in next-generation cellular networks," in *Proc. of the ACM SIGCOMM Conf.*, 2016, pp. 398–411.
- [16] Z. Zhong, L. Fan, and S. Ge, "FDD massive MIMO uplink and downlink channel reciprocity properties: Full or partial reciprocity?" *Proc. IEEE Glob. Comm. Conf.*, Dec. 2020.
- [17] D. Han, J. Park, and N. Lee, "FDD massive MIMO without CSI feedback," *IEEE Trans. Wireless Commun.*, vol. 23, no. 5, pp. 4518–4530, May 2024.
- [18] Y. Han, T. H. Hsu, C. K. Wen, K. K. Wong, and S. Jin, "Efficient downlink channel reconstruction for FDD multi-antenna systems," *IEEE Trans. Wireless Commun.*, vol. 18, no. 6, pp. 3161–3176, June 2019.
- [19] Q. An, M. Zafari, C. Dick, S. Segarra, A. Sabharwal, and R. Doost-Mohammady, "ML-based feedback-free adaptive MCS selection for massive multi-user MIMO," in *Proc. of Asilomar Conf. on Sign., Syst. and Computers*, 2023, pp. 157–161.
- [20] M. A. Maddah-Ali and D. Tse, "Completely stale transmitter channel state information is still very useful," *IEEE Trans. Inf. Theory*, vol. 58, no. 7, pp. 4418–4431, 2012.
- [21] A. Adhikary, J. Nam, J.-Y. Ahn, and G. Caire, "Joint spatial division and multiplexing—the large-scale array regime," *IEEE Trans. Inf. Theory*, vol. 59, no. 10, pp. 6441–6463, 2013.
- [22] J. Park, B. Lee, J. Choi, H. Lee, N. Lee, S.-H. Park, K.-J. Lee, J. Choi, S. H. Chae, S.-W. Jeon, K. S. Kwak, B. Clerckx, and W. Shin, "Rate-splitting multiple access for 6G networks: Ten promising scenarios and applications," *IEEE Network*, pp. 1–1, 2023.
- [23] H. Joudeh and B. Clerckx, "Sum-rate maximization for linearly precoded downlink multiuser MISO systems with partial CSIT: A rate-splitting approach," *IEEE Trans. Commun.*, vol. 64, no. 11, pp. 4847–4861, 2016.
- [24] J. Park, J. Choi, N. Lee, W. Shin, and H. V. Poor, "Rate-splitting multiple access for downlink MIMO: A generalized power iteration approach," *IEEE Trans. Wireless Commun.*, vol. 22, no. 3, pp. 1588–1603, Mar. 2023.
- [25] D. B. Amor, M. Joham, and W. Utschick, "Rate splitting in FDD massive MIMO systems based on the second order statistics of transmission channels," *IEEE J. Sel. Areas Commun.*, vol. 41, no. 5, pp. 1351–1365, May 2023.
- [26] T. Fang and Y. Mao, "Rate splitting multiple access: Optimal beamforming structure and efficient optimization algorithms," *IEEE Trans. Wireless Commun.*, vol. 23, no. 10, pp. 15 642–15 657, Oct. 2024.
- [27] B. Efron and D. V. Hinkley, "Assessing the accuracy of the maximum likelihood estimator: observed versus expected Fisher information," *Biometrika*, vol. 65, no. 3, pp. 457–483, 1978.
- [28] B. G. Lindsay and B. Li, "On second-order optimality of the observed Fisher information," *The Annals of Statistics*, vol. 25, no. 5, pp. 2172–2199, 1997.
- [29] X. Cao and J. C. Spall, "Relative performance of expected and observed Fisher information in covariance estimation for maximum likelihood estimates," in *American Control Conference (ACC)*. IEEE, 2012, pp. 1871–1876.
- [30] B. Clerckx, Y. Mao, E. A. Jorswieck, J. Yuan, D. J. Love, E. Erkip, and D. Niyato, "A primer on rate-splitting multiple access: Tutorial, myths, and frequently asked questions," *IEEE J. Sel. Areas Commun.*, 2023.
- [31] L. Z. X. Zhang and A. Sabharwal, "Directional training for FDD massive MIMO," *IEEE Trans. Wireless Commun.*, vol. 17, no. 8, pp. 5183–5197, 2018.
- [32] B. Mamandipoor, D. Ramasamy, and U. Madhoo, "Newtonized orthogonal matching pursuit: Frequency estimation over the continuum," *IEEE Trans. Signal Process.*, vol. 64, no. 19, pp. 5066–5081, Oct. 2016.
- [33] E. Björnson, L. Sanguinetti, and M. Debbah, "Massive MIMO with imperfect channel covariance information," in *Proc. of Asilomar Conf. on Sign., Syst. and Computers*, 2016, pp. 974–978.
- [34] S. M. Kay, "Statistical signal processing: estimation theory," *Prentice Hall*, vol. 1, pp. Chapter–3, 1993.
- [35] J. Choi, N. Lee, S.-N. Hong, and G. Caire, "Joint user selection, power allocation, and precoding design with imperfect CSIT for multi-cell MU-MIMO downlink systems," *IEEE Trans. Wireless Commun.*, vol. 19, no. 1, pp. 162–176, 2020.
- [36] B. Clerckx, Y. Mao, R. Schober, E. A. Jorswieck, D. J. Love, J. Yuan, L. Hanzo, G. Y. Li, E. G. Larsson, and G. Caire, "Is NOMA efficient in multi-antenna networks? A critical look at next generation multiple access techniques," *IEEE Open J. of the Commun. Society*, vol. 2, pp. 1310–1343, 2021.
- [37] S. S. Christensen, R. Agarwal, E. De Carvalho, and J. M. Cioffi, "Weighted sum-rate maximization using weighted MMSE for MIMO-BC beamforming design," *IEEE Trans. Wireless Commun.*, vol. 7, no. 12, pp. 4792–4799, 2008.
- [38] R. W. Heath Jr and A. Lozano, "Foundations of MIMO communication," *Cambridge University Press*, 2018.
- [39] 3GPP Technical Specification 38.214, "Physical layer procedures for data (Release 15)," 2018.