**QUEEN'S
UNIVERSITY
BELFAST**

Hybrid Precoding for Massive MIMO with Low Rank Channels: A Two-Stage User Scheduling Approach

Almradi, A., Matthaiou, M., Xiao, P., & Fusco, V. F. (2020). Hybrid Precoding for Massive MIMO with Low Rank Channels: A Two-Stage User Scheduling Approach. *IEEE Transactions on Communications*, 68(8), 4816-4831. <https://doi.org/10.1109/TCOMM.2020.2991805>

Published in:
IEEE Transactions on Communications

Document Version:
Peer reviewed version

Queen's University Belfast - Research Portal:
[Link to publication record in Queen's University Belfast Research Portal](#)

Publisher rights
Copyright 2020 IEEE. This work is made available online in accordance with the publisher's policies. Please refer to any applicable terms of use of the publisher.

General rights
Copyright for the publications made accessible via the Queen's University Belfast Research Portal is retained by the author(s) and / or other copyright owners and it is a condition of accessing these publications that users recognise and abide by the legal requirements associated with these rights.

Take down policy
The Research Portal is Queen's institutional repository that provides access to Queen's research output. Every effort has been made to ensure that content in the Research Portal does not infringe any person's rights, or applicable UK laws. If you discover content in the Research Portal that you believe breaches copyright or violates any law, please contact openaccess@qub.ac.uk.

Hybrid Precoding for Massive MIMO with Low Rank Channels: A Two-Stage User Scheduling Approach

Ahmed Almradi, Michail Matthaiou, *Senior Member, IEEE*, Pei Xiao, *Senior Member, IEEE*, and Vincent F. Fusco, *Fellow, IEEE*

Abstract—To fully reap the benefits of massive multiple-input multiple-output hybrid analog and digital precoding in frequency division duplexing single-cell systems, a two-stage precoder is developed utilizing the signal-to-leakage-plus-noise ratio metric. The main idea of this technique is to jointly design the analog precoder based only on the long-term channel statistics information at the transmitter, i.e., the channel mean and reconstructed reduced rank covariance statistics, while the digital precoder is designed based on the instantaneous channel state information of the reduced dimensionality effective reconstructed channel. Consequently, we can significantly reduce the downlink training and uplink feedback overhead analogously to the rank of the resultant effective channel. The two extremes of full channel state information at the transmitter (CSIT) and statistical CSIT are also investigated. The performance gap between the full and statistical CSIT corroborates the importance of the proposed two-stage CSIT approach. These precoders are then extended to multi-cell systems. It is shown that the digital baseband precoder design problem reduces to the generalized Rayleigh quotient problem, while the analog precoder design problem reduces to the quotient trace problem, also known as the ratio trace problem. These dimensionality reduction problems are solved via the generalized eigenvalue decomposition method. Finally, in the presence of multiuser diversity where only a subset of the users are scheduled, to considerably alleviate the channel estimation and feedback overhead burden, a low-complexity one-stage and two-stage CSIT joint user scheduler and precoder algorithms are developed.

Index Terms—Hybrid analog and digital precoding, multiuser MIMO systems, sum spectral efficiency, statistical CSIT, signal-to-leakage-plus-noise ratio (SLNR), two-stage precoding, zero-forcing (ZF).

I. INTRODUCTION

MULTIUSER massive multiple-input multiple-output (MIMO) systems have attracted a lot of research interest due to their ability to significantly improve the achievable

throughput [2]–[6]. Over the same time-frequency transmission resources, massive MIMO can simultaneously serve multiple users in a space division multiple access (SDMA) manner. However, such systems require a dedicated radio-frequency (RF) chain for each antenna element, resulting in high cost and power consumption at the base station (BS). To alleviate these issues, hybrid analog and digital precoding transceivers have been recently investigated, where a much smaller number of RF chains are connected to a massive number of antennas via a set of low power and cost analog phase shifters. These kinds of structures divide the precoding and beamforming operations between the digital baseband and analog RF domains [7]–[9].

To fully reap the benefits of multiuser massive MIMO, full channel state information at the transmitter (CSIT) is required in order to accurately serve multiple users with the least possible multiuser interference [10]–[12]. However, the acquisition and feedback overhead of instantaneous CSIT is the bottleneck of multiuser massive MIMO systems in cellular networks, where frequency-division duplexing (FDD) is utilized.¹ In FDD systems, the uplink (UL) and downlink (DL) channels have to be estimated separately as they operate in different frequency bands. Therefore, in order to have instantaneous CSIT with the aid of training symbols, the channel has to be estimated at each user terminal and then fed back to the BS via a low-rate feedback channel. The bottleneck with FDD massive MIMO systems is that the training and feedback overhead resources scale with the number of BS antennas [5], [13]–[15]. Therefore, to reap the benefits of massive MIMO in FDD systems, as in the case of massive MIMO in time-division duplexing (TDD) systems, where the reciprocity of the channel is being leveraged, channel dimensionality reduction techniques have been proposed. In this context, a technique called joint spatial division and multiplexing (JSDM) was proposed in [14] for single-cell massive MIMO systems. In this work, channel dimensionality reduction is performed in two-stages; a pre-beamforming stage, which utilizes the channel second-order statistics to divide users into groups based on the similarities of their correlation structure, followed by a multiuser MIMO precoding stage, that eliminates multiuser interference within each group, designed based on full CSIT of the reduced dimensionality effective channel. The JSDM technique was extended to the multi-cell scenario in [16], where the outer

The work of A. Almradi and V. Fusco was supported by the U.K. Engineering and Physical Sciences Research Council (EPSRC) under Grant EP/P000673/1 and Grant EP/N020391/1. The work of M. Matthaiou was in part supported by a research grant from the Department for the Economy Northern Ireland under the US-Ireland R&D Partnership Programme. The work of P. Xiao was supported by the U.K. Engineering and Physical Sciences Research Council (EPSRC) under Grant EP/N020391/1. Part of this paper will be presented at the IEEE ICC conference, Dublin, Ireland, June 2020 [1]. (Corresponding author: Ahmed Almradi.)

A. Almradi, M. Matthaiou and V. F. Fusco are with ECIT, Queen's University Belfast, Belfast, BT3 9DT, U.K. (emails: {a.almradi, m.matthaiou}@qub.ac.uk, v.fusco@ecit.qub.ac.uk).

P. Xiao is with the 5G Innovation Centre, University of Surrey, Guildford, GU2 7XH, U.K. (email: p.xiao@surrey.ac.uk).

¹Since FDD mode dominates current wireless cellular networks, it is of practical interest to design transceivers that leverage massive MIMO in FDD mode.

precoder was used to alleviate inter-cell interference, relying only on the second order channel statistics², while the inner precoder mitigates intra-cell interference leveraging full CSIT of a reduced dimensionality effective channel.

Most previous works on hybrid precoding design the analog and digital precoders based on the assumption of full CSIT (see e.g., [17]–[20]). However, the acquisition of perfect CSIT in hybrid precoding imposes substantial challenges in both TDD and FDD systems. The hybrid structure makes it infeasible to acquire the entire channel of all antennas since the channel estimator in the baseband pinpoints only to a reduced dimensionality channel through a fewer number of RF chains. In this context, the authors in [2] utilized the covariance channel statistics for the design of the digital precoder, while the analog precoder is designed by beam search based on a beamsteering codebook. We recall that the long-term channel statistics remain constant over a longer period of time compared to instantaneous channel realizations, thereby offering a significant reduction in pilot symbols and feedback overhead. Yet, although the statistical CSIT results in comparable performance with full CSIT at low signal-to-noise ratio (SNR), its performance saturates at high SNR due to the residual multiuser interference caused by the unknown instantaneous variations of the channels. In addition, the authors in [3] designed the analog precoder based on the knowledge of channel covariance statistics, utilizing the averaged signal-to-leakage-plus-noise ratio (SLNR) metric, while the digital precoder was derived by using the regularized zero-forcing (ZF) technique, exploiting full CSIT of the reduced dimensionality effective channel.

Statistical beamforming for multiuser massive MIMO systems have been extensively investigated based on the assumption of correlated Rayleigh fading channels (see e.g., [21], [22]) and cannot be easily generalized to scenarios where we have a dominant line-of-sight (LOS) component. In the context of Rician fading channels, the authors in [4], [23] exploited the knowledge of channel mean information at the transmitter to introduce a sub-optimal precoder for the traditional fully digital structure. Therefore, contrary to prior works, we consider a two-stage hybrid precoding in generalized fading conditions, namely rank deficient correlated Rician fading channels. This generalized model is suitable for dense small cell networks, vehicular networks, indoor applications, and millimeter wave (mmWave) propagation scenarios. Consequently, in the two-stage precoder, the analog precoder is determined by the knowledge of the direct LOS path components and reconstructed reduced rank channel covariance matrices of the users, whereas the digital precoder utilizes perfect CSIT of the reduced dimensionality reconstructed effective channel.

The main contributions of this work are listed as follows:

- 1) A new two-stage CSIT hybrid precoder design is proposed for FDD multiuser massive MIMO systems. In this work, the analog and digital precoders are jointly designed by utilizing the SLNR metric. In contrast to the traditional literature where the analog precoder is de-

signed by utilizing only the second order channel statistics; in particular, the knowledge of the channel mean and reconstructed reduced rank covariance statistics are leveraged to determine a closed-form analog precoder, whereas a closed-form baseband precoder is proposed by utilizing full CSIT of the reduced dimensionality reconstructed effective channel. The reconstructed reduced rank covariance statistics leverages only the dominant eigenmodes defined through the notion of effective rank.

- 2) The case of statistical CSIT is also investigated where closed-form analog and digital precoders are presented. In statistical CSIT, these precoders are designed based on the knowledge of channel mean and reconstructed reduced rank covariance statistics. Meanwhile, for comparison purposes, the case of full CSIT is also considered. The proposed precoders are then generalized to multi-cell systems.
- 3) Based on the derived closed-form precoders, low-complexity leakage based user scheduling algorithms are proposed for the full CSIT, two-stage CSIT, and statistical CSIT scenarios. Motivated by the fact that user scheduling consumes a considerable amount of CSIT acquisition and feedback overhead, and that the scheduling and beamforming stage could get away with coarse CSIT, e.g., statistical CSIT. We emphasize that this is done by leveraging only the first and reconstructed second order channel statistics information at the BS, we schedule users that have statistically semi-orthogonal dominant eigendirections, i.e., spatially well separated users with nearly non-overlapping eigendirections, thereby significantly alleviating the CSIT acquisition overhead burden for the baseband precoder design in the second stage of the two-stage CSIT precoder. Therefore, only full CSIT of the reduced dimensionality reconstructed effective channels of the scheduled users are acquired and fed back to the BS.

A. Paper Organization

Section II introduces the system and channel model. In Section III, the design of baseband and analog precoders in the presence of full CSIT is investigated, whereas Section IV considers the same design problem for the case of statistical CSIT. Section V proposes a two-stage CSIT precoder. In Section VI, three low-complexity user schedulers are introduced. Section VII presents the numerical results, while Section VIII concludes the paper.

B. Notations

Bold lower (upper) case letters denote vectors (matrices). The notations $\mathbf{X}_k$, $\mathbf{X}^T$, $\mathbf{X}^\dagger$, $\mathbf{X}^{-1}$, and $\text{Tr}(\mathbf{X})$ denote the matrix $\mathbf{X}$ that excludes column $\mathbf{x}_k$, the transpose, conjugate transpose, inverse and trace of a matrix $\mathbf{X}$, respectively. Also, $[\mathbf{X}]_{m,n}$ and $\angle [\mathbf{X}]_{m,n}$ denote the $(m, n)^{\text{th}}$ element of the matrix $\mathbf{X}$ and its phase; $\|\cdot\|$ and $|\cdot|$ denote the Euclidean norm and absolute value operators, respectively; $\mathbb{E}(\cdot)$ denotes the expectation operator, whereas $\mathbf{u}_l[\mathbf{X}]$ and $\lambda_l[\mathbf{X}]$ are the l^{th} dominant eigenvector and the corresponding eigenvalue of the

²Paper [16] assumes that each BS serves one cluster of users, thus users in each BS share the same covariance statistics.

matrix $\mathbf{X}$, with $\lambda_{\max}[\mathbf{X}] = \lambda_1[\mathbf{X}]$ and $\mathbf{u}_{\max}[\mathbf{X}] = \mathbf{u}_1[\mathbf{X}]$; $\text{diag}(x_1, x_2, \dots, x_N)$ denotes a diagonal matrix with diagonal elements $\{x_1, x_2, \dots, x_N\}$, whilst $\lceil \cdot \rceil$ is the ceiling operator.

II. SYSTEM AND CHANNEL MODEL

Consider a single-cell multiuser MIMO DL system, where a BS equipped with N_{BS} antennas and N_{RF} RF chains serves $\mathcal{K}$ single antenna users, where $\mathcal{K} \leq N_{RF} \leq N_{BS}$. A hybrid analog RF precoder $\mathbf{F} \in \mathbb{C}^{N_{BS} \times N_{RF}}$ and a digital baseband precoder $\mathbf{W} \in \mathbb{C}^{N_{RF} \times \mathcal{K}}$ are deployed at the BS. As a network of phase shifters is used to represent the analog precoding stage with phase-only control, thus, its entries are restricted to have a constant modulus, i.e., $|\mathbf{F}|_{m,n}| = \frac{1}{\sqrt{N_{BS}}}$.

To explicitly take into account the dominant LOS component, in addition to the low rank property of the channel covariance matrix, the k^{th} user downlink channel model is defined as³

$$\mathbf{h}_k = \sqrt{\frac{K_k}{K_k + 1}} \mathbf{h}_{\text{LOS},k} + \sqrt{\frac{1}{K_k + 1}} \mathbf{U}_k \mathbf{\Lambda}_k^{\frac{1}{2}} \mathbf{h}_{\text{NLOS},k}, \quad (1)$$

where K_k is the ratio between the specular LOS and scattered non-LOS (NLOS) channel components, known as the Rician K-factor for the k^{th} user, $\mathbf{\Sigma}_k = \mathbf{U}_k \mathbf{\Lambda}_k \mathbf{U}_k^\dagger$ is the eigenvalue decomposition for the k^{th} user NLOS channel covariance matrix, $\mathbf{\Lambda}_k$ is an $r_k \times r_k$ diagonal matrix whose entries are the $r_k \leq N_{BS}$ nonzero eigenvalues of $\mathbf{\Sigma}_k$ organized in a descending order, defined as⁴ $r_k = \text{rank}(\mathbf{\Sigma}_k)$, $\mathbf{U}_k \in \mathbb{C}^{N_{BS} \times r_k}$ is a semi-unitary matrix consists of the eigenvectors of $\mathbf{\Sigma}_k$ corresponding to the nonzero eigenvalues. Also, $\mathbf{h}_{\text{NLOS},k} \in \mathbb{C}^{r_k \times 1}$ denotes the random NLOS channel for the k^{th} user, whose entries are independent and identically distributed (i.i.d.) random variables and distributed according to $\mathcal{CN}(0, 1)$, while $\mathbf{h}_{\text{LOS},k} \in \mathbb{C}^{N_{BS} \times 1}$ represents the deterministic (LOS) component, satisfying $\|\mathbf{h}_{\text{LOS},k}\|^2 = N_{BS}$, defined as

$$\mathbf{h}_{\text{LOS},k} = \left[1 e^{-j \frac{2\pi d}{\lambda} \sin(\varphi_k)} \dots e^{-j \frac{2\pi d}{\lambda} (N_{BS}-1) \sin(\varphi_k)} \right]^T, \quad (2)$$

where λ is the wavelength, d is the inter-antenna spacing, and φ_k is the k^{th} user angle-of-departure (AoD), which is uniformly distributed according to $\varphi_k \sim \mathcal{U}(-\frac{\pi}{2}, \frac{\pi}{2})$.

The one-ring local scattering correlation model is used here, where the $(m, n)^{\text{th}}$ entry of $\mathbf{\Sigma}_k$ is given as [25, Eq. (21)]

$$[\mathbf{\Sigma}_k]_{m,n} = \frac{1}{2\Delta_k} \int_{\phi_k - \Delta_k}^{\phi_k + \Delta_k} e^{-j \frac{2\pi d}{\lambda} (n-m) \sin(\theta_k)} d\theta_k, \quad (3)$$

where Δ_k denotes the k^{th} user azimuth angular spread, ϕ_k is the k^{th} user central azimuth angle, and θ_k is the k^{th} user actual AoD, distributed according to $\theta_k \sim \mathcal{U}(\phi_k - \Delta_k, \phi_k + \Delta_k)$.

The composite transmitted signal at the BS is given as

$$\mathbf{x} = \mathbf{F} \mathbf{W} \mathbf{P} \mathbf{s} = \sum_{j=1}^{\mathcal{K}} \sqrt{P_j} \mathbf{F} \mathbf{w}_j s_j, \quad (4)$$

³The Karhunen–Loève correlated channel model representation is extended here for the case where there is a strong LOS component.

⁴The rank of the channel covariance matrix depends on the angular spread, inter-antenna spacing, and wavelength [14], [24].

where $\mathbf{s} \in \mathbb{C}^{\mathcal{K} \times 1} = [s_1 \ s_2 \ \dots \ s_{\mathcal{K}}]^T$ are the payload data symbols intended for the $\mathcal{K}$ scheduled users, with entries that are i.i.d. Gaussian random variables with zero mean and unit variance, and $\mathbf{P} = \text{diag}(\sqrt{P_1}, \sqrt{P_2}, \dots, \sqrt{P_{\mathcal{K}}})$ is a diagonal power allocation matrix that satisfies the power constraint, i.e., the total transmit power $\sum_{j=1}^{\mathcal{K}} P_j = P$.

For ease of exposition, the power is equally allocated amongst users,⁵ i.e., $\mathbf{P} = \sqrt{\frac{P}{\mathcal{K}}} \mathbf{I}_{\mathcal{K}}$, resulting in $\sum_{j=1}^{\mathcal{K}} \|\mathbf{F} \mathbf{w}_j\|^2 = \mathcal{K}$, hence, in this work, we normalize $\mathbf{F}$ and $\mathbf{W}$ such that $\|\mathbf{F} \mathbf{w}_j\|^2 = 1$.

Remark 1. Due to the small angular spread of the incoming/outgoing multipaths at the BS, massive MIMO channels tend to very often have low rank structure with $r_k \ll N_{BS}$ [14], [24], [27]. Furthermore, it is observed that some of the r_k nonzero eigenvalues of $\mathbf{\Sigma}_k$ are extremely small, and as such could be ignored, resulting in an effective rank of $r_{\text{eff},k} \ll r_k \ll N_{BS}$. For instance, from (3), if we set $N_{BS} = 64$, $d = 0.5$, $\phi_k = \frac{\pi}{6}$, and $\Delta_k = 15^\circ$, these settings result in a covariance matrix of rank $r_k = 38$ including some extremely small nonzero eigenvalues. Meanwhile, the effective rank (based on (5) below) is only $r_{\text{eff},k} = 26$. As a consequence, the chosen effective rank will substantially impact the overall system performance. For this reason, we now pursue a detailed analysis of this metric.

Definition 2. The effective rank, which captures the number of dominant eigenmodes of the channel covariance matrix $\mathbf{\Sigma}_k$, is defined as [28]

$$r_{\text{eff},k} = \lceil \exp[\mathcal{H}(\sigma_1, \sigma_2, \dots, \sigma_{r_k})] \rceil \quad (5)$$

where $\sigma_l = \frac{\lambda_l[\mathbf{\Sigma}_k]}{\sum_{l=1}^{r_k} \lambda_l[\mathbf{\Sigma}_k]}$, and $\mathcal{H}(\sigma_1, \sigma_2, \dots, \sigma_{r_k})$ is the Shannon (spectral) entropy,⁶ given by $\mathcal{H}(\sigma_1, \sigma_2, \dots, \sigma_{r_k}) = -\sum_{l=1}^{r_k} \sigma_l \log(\sigma_l)$.

A. Dominant Eigenspace Reconstruction

Motivated by the introduced effective rank in (5) and the work in [5], [29], we henceforth explicitly take into account only the dominant eigenspace of the channel in the design of $\mathbf{F}$ and $\mathbf{W}$, thus, substantially reducing the training sequence and feedback overhead. Hence, $\mathbf{F}$ and $\mathbf{W}$ are designed based on the reconstructed channels and covariance matrices $\hat{\mathbf{h}}_k$, $\mathbf{\Sigma}_{\text{eff},k} \forall k \in \{1, 2, \dots, \mathcal{K}\}$ at the BS. Therefore, the k^{th} user downlink channel model is now represented as

$$\mathbf{h}_k = \hat{\mathbf{h}}_k + \tilde{\mathbf{h}}_k, \quad (6)$$

where $\hat{\mathbf{h}}_k = \sqrt{\frac{K_k}{K_k + 1}} \mathbf{h}_{\text{LOS},k} + \sqrt{\frac{1}{K_k + 1}} \mathbf{U}_{\text{eff},k} \mathbf{\Lambda}_{\text{eff},k}^{\frac{1}{2}} \mathbf{h}_{\text{NLOS,eff},k}$ is the dominant eigenspace reconstructed channel, $\mathbf{U}_{\text{eff},k} = \mathbf{U}_k(:, 1:r_{\text{eff},k})$, the $r_{\text{eff},k}$ dominant eigenvectors corresponding to $\mathbf{\Lambda}_{\text{eff},k}$, with $r_{\text{eff},k}$ given in (5), $\mathbf{\Lambda}_{\text{eff},k} =$

⁵Since the focus of this work is on the analog and digital precoder design, a uniform power allocation is considered herein for the sake of analytical tractability. In our future work, the results of this work will be extended to more advanced power allocation schemes, e.g., the max min weighted SINR formula [26].

⁶The entropy is a measure of the average uncertainty of a random variable, and has been extensively used as a measure of the effective dimensionality [28].

$\mathbf{\Lambda}_k(1:r_{\text{eff},k}, 1:r_{\text{eff},k})$, the $r_{\text{eff},k}$ dominant eigenvalues of $\mathbf{\Sigma}_k$, and $\mathbf{h}_{\text{NLOS,eff},k} \in \mathbb{C}^{r_{\text{eff},k} \times 1}$ with i.i.d. entries distributed according to $\mathcal{CN}(0, 1)$, whereas the residual error is given as $\tilde{\mathbf{h}}_k = \sqrt{\frac{1}{K_k+1}} \mathbf{U}_{\text{res},k} \mathbf{\Lambda}_{\text{res},k}^{\frac{1}{2}} \mathbf{h}_{\text{NLOS,res},k}$, with $\mathbf{U}_{\text{res},k} = \mathbf{U}_k(:, r_{\text{eff},k}+1:r_k)$, the remaining $r_k - r_{\text{eff},k}$ non-dominant eigenvectors corresponding to $\mathbf{\Lambda}_{\text{res},k}$, $\mathbf{\Lambda}_{\text{res},k} = \mathbf{\Lambda}_k(r_{\text{eff},k}+1:r_k, r_{\text{eff},k}+1:r_k)$, the remaining $r_k - r_{\text{eff},k}$ non-dominant eigenvalues of $\mathbf{\Sigma}_k$, and $\mathbf{h}_{\text{NLOS,res},k} \in \mathbb{C}^{(r_k - r_{\text{eff},k}) \times 1}$. Note that for brevity, only the reconstruction error due to reduced rank channel covariance matrix is investigated herein, whilst channel estimation and quantization errors due to training sequences and UL feedback are ignored.

In the presence of reconstructed reduced rank channels at the BS, the received signal at the k^{th} user is written as

$$y_k = \underbrace{\sqrt{P_k} \hat{\mathbf{h}}_k^\dagger \mathbf{F} \mathbf{w}_k s_k}_{\text{desired signal}} + \underbrace{\sqrt{P_k} \tilde{\mathbf{h}}_k^\dagger \mathbf{F} \mathbf{w}_k s_k}_{\text{reconstruction residual}} + \underbrace{\mathbf{h}_k^\dagger \sum_{j \neq k} \sqrt{P_j} \mathbf{F} \mathbf{w}_j s_j}_{\text{multiuser interference}} + \underbrace{n_k}_{\text{AWGN noise}}, \quad (7)$$

where n_k is the additive white Gaussian noise (AWGN) at the k^{th} user, distributed as $n_k \sim \mathcal{CN}(0, \sigma^2)$.

Therefore, conditioned on $\hat{\mathbf{h}}_k$, the signal-to-interference-plus-noise ratio (SINR) at the k^{th} user is written as

$$\text{SINR}_k = \frac{\rho \left| \hat{\mathbf{h}}_k^\dagger \mathbf{F} \mathbf{w}_k \right|^2}{1 + \rho \left| \tilde{\mathbf{h}}_k^\dagger \mathbf{F} \mathbf{w}_k \right|^2 + \rho \sum_{j \neq k} \left| \mathbf{h}_j^\dagger \mathbf{F} \mathbf{w}_j \right|^2}, \quad (8)$$

where $\rho = \frac{P}{K\sigma^2}$ is the signal-to-noise ratio (SNR).

Our objective is to derive closed-form solutions for the $\mathbf{F}$ and $\mathbf{W}$ precoders such that the spectral efficiency is maximized. The achievable sum spectral efficiency of the system is expressed as

$$\mathcal{R}_{\text{sum}} = \sum_{k=1}^K \mathcal{R}_k = \sum_{k=1}^K \mathbb{E} [\log_2 (1 + \text{SINR}_k)]. \quad (9)$$

Finding the optimal $\mathbf{F}$ and $\mathbf{W}$ which maximize the achievable sum spectral efficiency in (9) is a challenging non-convex optimization problem due to the coupling between the baseband precoders $\{\mathbf{w}_k, k = 1, \dots, K\}$ in each user's SINR in (8). As a consequence, maximum ratio transmission (MRT) and ZF schemes are two well known techniques utilized to obtain sub-optimal closed-form precoders (see e.g., [7], [12], [17], [30] among others). The MRT scheme selfishly finds the precoder which maximizes the numerator in (8), $\left| \hat{\mathbf{h}}_k^\dagger \mathbf{F} \mathbf{w}_k \right|^2$, while ignoring the impact of multiuser interference. On the other hand, the ZF scheme completely nulls out the negative effect of multiuser interference in the denominator, i.e., forces $\sum_{j \neq k} \left| \mathbf{h}_j^\dagger \mathbf{F} \mathbf{w}_j \right|^2 = 0$, by ignoring the impact of AWGN noise.

To avoid tackling the coupled optimization problem in (9), we leverage the signal-to-leakage-plus-noise ratio (SLNR) metric, which leads to a near optimal closed-form solution, and has been recently investigated in [2], [3], [10], [11], [29], [31].

The SLNR at the k^{th} user is defined as⁷ [11]

$$\text{SLNR}_k = \frac{\rho \left| \hat{\mathbf{h}}_k^\dagger \mathbf{F} \mathbf{w}_k \right|^2}{1 + \rho \left| \tilde{\mathbf{h}}_k^\dagger \mathbf{F} \mathbf{w}_k \right|^2 + \rho \sum_{j \neq k} \left| \mathbf{h}_j^\dagger \mathbf{F} \mathbf{w}_k \right|^2}. \quad (10)$$

Let $\mathbf{V}$ be the unconstrained analog precoder,⁸ then the constrained analog precoder $\mathbf{F}$ can be found by solving [3]

$$\begin{aligned} \mathbf{F}^* &= \arg \min_{\mathbf{F}} \quad \|\mathbf{V} - \mathbf{F}\|_F^2 \\ \text{s. t.} \quad & \left| [\mathbf{F}]_{m,n} \right| = \frac{1}{\sqrt{N_{BS}}} \end{aligned} \quad (11)$$

A sub-optimal solution to (11) is given by $[\mathbf{F}^*]_{m,n} = \frac{1}{\sqrt{N_{BS}}} \exp(\angle [\mathbf{V}]_{m,n})$ [3].

Therefore, in this paper, the optimal solution which maximizes the SLNR_k in (10) for the unconstrained $\mathbf{V}$ is investigated first, such that the constrained $\mathbf{F}$ can simply be found by (11).

III. FULL CSIT

Due to its fewer number of RF chains than BS antennas, it is infeasible to acquire the entire channel matrix of all antennas at the BS in the hybrid precoding structure for both TDD and FDD systems, in addition to the prohibitive CSIT acquisition and feedback overhead in FDD systems. Therefore, the full CSIT scenario is considered here simply for benchmarking against the proposed two-stage CSIT hybrid precoder. In this section, $\mathbf{V}$ and $\mathbf{W}$ are both designed based on the reconstructed reduced rank channel, which significantly reduces the acquisition and feedback overhead in reconstructing⁹ the channels at the BS for the design of $\mathbf{V}$ and $\mathbf{W}$.

A. Baseband Precoder Design

From (10), conditioned on the knowledge of the reconstructed channels $\hat{\mathbf{h}}_k \forall k$, in the absence of residual errors, the transmitter treats the reconstructed channels as the real channel coefficients, thus, we formulate the baseband precoder that maximizes the SLNR_k as follows

$$\begin{aligned} \mathbf{w}_k^* &= \arg \max_{\mathbf{w}_k} \quad \frac{\rho \left| \hat{\mathbf{h}}_k^\dagger \mathbf{V} \mathbf{w}_k \right|^2}{1 + \rho \sum_{j \neq k} \left| \hat{\mathbf{h}}_j^\dagger \mathbf{V} \mathbf{w}_k \right|^2} \\ \text{s. t.} \quad & \|\mathbf{V} \mathbf{w}_k\|^2 = 1 \end{aligned} \quad (12)$$

⁷The SLNR definition can be alternatively obtained by leveraging the SLNR leakage space formulation of [29, Eq. (25)], by setting the weighting factors as $\eta_k, k = \eta_m, k = \rho$. Note that the SINR_k in (8) and SLNR_k in (10) are observed at the user side, and not to be confused with the ones utilized for precoder design at the BS.

⁸Unlike the constrained analog precoder $\mathbf{F}$ where its entries are restricted to have constant modulus, the unconstrained analog precoder $\mathbf{V}$ entries do not have such constraint and the amplitude and phase of each entry of $\mathbf{V}$ can be controlled.

⁹In limited feedback systems where finite size codebooks are utilized to estimate the channel and feedback its index to the BS, a reduced rank channel results in substantially less quantization error due to the reduced quantization dimensions.

Lemma 3. *The optimal solution to the baseband precoder in (12) is derived as*

$$\mathbf{w}_k^* = \frac{\left(\mathbf{V}^\dagger \mathbf{V} + \rho \sum_{j \neq k} \mathbf{V}^\dagger \hat{\mathbf{h}}_j \hat{\mathbf{h}}_j^\dagger \mathbf{V} \right)^{-1} \mathbf{V}^\dagger \hat{\mathbf{h}}_k}{\left\| \mathbf{V} \left(\mathbf{V}^\dagger \mathbf{V} + \rho \sum_{j \neq k} \mathbf{V}^\dagger \hat{\mathbf{h}}_j \hat{\mathbf{h}}_j^\dagger \mathbf{V} \right)^{-1} \mathbf{V}^\dagger \hat{\mathbf{h}}_k \right\|}. \quad (13)$$

Proof: The proof is given in Appendix A. ■

Given $\mathbf{w}_k^*$ in (13), the resultant SLNR for the k^{th} user from (12) simplifies to

$$\text{SLNR}_k = \rho \hat{\mathbf{h}}_k^\dagger \mathbf{V} \left(\mathbf{V}^\dagger \mathbf{V} + \rho \sum_{j \neq k} \mathbf{V}^\dagger \hat{\mathbf{h}}_j \hat{\mathbf{h}}_j^\dagger \mathbf{V} \right)^{-1} \mathbf{V}^\dagger \hat{\mathbf{h}}_k. \quad (14)$$

1) *Zero-Forcing Precoder:* ZF precoding is used in the baseband to completely eliminate the remaining multiuser interference due to the constant modulus constraint of the analog precoder. The digital precoder for the k^{th} user is designed to lie in the null space of the multiuser interference channels, $\mathbf{w}_k \in \text{null} \left\{ \hat{\mathbf{h}}_k^\dagger \mathbf{V} \mathbf{w}_j \right\} \forall k \neq j$. Therefore, by adding the constraint $\hat{\mathbf{h}}_k^\dagger \mathbf{V} \mathbf{w}_j = 0 \forall k \neq j$, a sub-optimal ZF baseband precoder is formulated as

$$\begin{aligned} \mathbf{w}_k^{\text{ZF}} = \arg \max_{\mathbf{w}_k} & \quad \left| \hat{\mathbf{h}}_k^\dagger \mathbf{V} \mathbf{w}_k \right|^2 \\ \text{s. t.} & \quad \left\| \mathbf{V} \mathbf{w}_k \right\|^2 = 1 \ \& \ \hat{\mathbf{h}}_k^\dagger \mathbf{V} \mathbf{w}_j = 0 \ \forall k \neq j \end{aligned} \quad (15)$$

Corollary 4. *The optimal solution to the constrained optimization problem in (15) is derived as*

$$\mathbf{w}_k^{\text{ZF}} = \frac{\mathbf{P}_k^\perp \mathbf{V}^\dagger \hat{\mathbf{h}}_k}{\left\| \mathbf{V} \mathbf{P}_k^\perp \mathbf{V}^\dagger \hat{\mathbf{h}}_k \right\|}, \quad (16)$$

where $\mathbf{P}_k^\perp = \mathbf{I}_{N_{RF}} - \mathbf{V}^\dagger \hat{\mathbf{H}}_k \left(\hat{\mathbf{H}}_k^\dagger \mathbf{V} \mathbf{V}^\dagger \hat{\mathbf{H}}_k \right)^{-1} \hat{\mathbf{H}}_k^\dagger \mathbf{V}$ is the null space projection matrix of the effective reconstructed channel $\mathbf{V}^\dagger \hat{\mathbf{H}}_k$, and $\hat{\mathbf{H}}_k = [\hat{\mathbf{h}}_1, \hat{\mathbf{h}}_2, \dots, \hat{\mathbf{h}}_{k-1}, \hat{\mathbf{h}}_{k+1}, \dots, \hat{\mathbf{h}}_K]$ is the composite reconstructed channel matrix which excludes $\hat{\mathbf{h}}_k$.

Proof: The proof is given in Appendix B. ■

The optimal baseband precoder in (13) approaches the ZF baseband precoder in (16) at high SNR. This is easily seen from (72) in Appendix B, where $\lim_{\rho \rightarrow \infty} \left(\mathbf{V}^\dagger \mathbf{V} + \rho \sum_{j \neq k} \mathbf{V}^\dagger \hat{\mathbf{h}}_j \hat{\mathbf{h}}_j^\dagger \mathbf{V} \right)^{-1} \mathbf{V}^\dagger \hat{\mathbf{h}}_k = \mathbf{P}_k^\perp \mathbf{V}^\dagger \hat{\mathbf{h}}_k$. With $\mathbf{w}_k^{\text{ZF}}$ in (16), the resultant SINR_k in (8) is simplified as

$$\text{SINR}_k^{\text{ZF}} = \frac{\rho \left| \hat{\mathbf{h}}_k^\dagger \mathbf{V} \mathbf{P}_k^\perp \mathbf{V}^\dagger \hat{\mathbf{h}}_k \right|^2}{\left\| \mathbf{V} \mathbf{P}_k^\perp \mathbf{V}^\dagger \hat{\mathbf{h}}_k \right\|^2 + \rho \left| \hat{\mathbf{h}}_k^\dagger \mathbf{V} \mathbf{P}_k^\perp \mathbf{V}^\dagger \hat{\mathbf{h}}_k \right|^2}. \quad (17)$$

B. Analog Precoder Design

Since $\mathbf{V}$ comprises the precoding vectors of all users, and although several problem formulations could be investigated, for the sake of closed-form solutions, we utilize the sum

SLNR_k of all users to formulate the optimal analog precoder design:

$$\mathbf{V}^* = \arg \max_{\mathbf{V}} \sum_{k=1}^K \text{SLNR}_k \quad (18)$$

where the SLNR_k is given in (14).

Proposition 5. *A sub-optimal solution to the analog precoder in (18) is derived as*

$$\mathbf{V}^* = \mathbf{U}_{[1:N_{RF}]} \left[\sum_{k=1}^K \left(\mathbf{I}_{N_{BS}} + \rho \sum_{j=1}^K \hat{\mathbf{h}}_j \hat{\mathbf{h}}_j^\dagger \right)^{-1} \hat{\mathbf{h}}_k \hat{\mathbf{h}}_k^\dagger \right] \quad (19)$$

where $\mathbf{U}_{[1:N_{RF}]} = \mathbf{U}(:, 1:N_{RF})$ is a $N_{BS} \times N_{RF}$ matrix whose columns are the N_{RF} leading principal components (eigenvectors) corresponding to the N_{RF} strongest eigenvalues of the matrix $\sum_{k=1}^K \left(\mathbf{I}_{N_{BS}} + \rho \sum_{j=1}^K \hat{\mathbf{h}}_j \hat{\mathbf{h}}_j^\dagger \right)^{-1} \hat{\mathbf{h}}_k \hat{\mathbf{h}}_k^\dagger$.

Proof: The proof is given in Appendix C. ■

IV. STATISTICAL CSIT

Statistical CSIT is preferable in FDD massive MIMO and leveraged here as it varies over a longer timescale when compared to the instantaneous channel variations, hence, substantially reducing the required training sequences and feedback overhead. Traditionally, in FDD systems, the acquisition of channel statistics information at the transmitter is acquired through long-term DL training sequences and low-rate UL feedback. However, channel statistics reciprocity is being leveraged for FDD systems more recently [24], [32]–[34]. This key observation eliminates the need for channel statistics estimation and feedback in FDD systems, and results in significant training and feedback overhead reduction. However, a residual amount of multiuser interference remains due to the impact of unknown instantaneous multipath components, resulting in a performance saturation particularly at high SNR. Therefore, in this section, $\mathbf{V}$ and $\mathbf{W}$ are designed based only on the channel mean and reconstructed covariance statistics via the maximization of a lower-bound on the average SLNR.

Proposition 6. *A lower-bound on the average SLNR at the k^{th} user is derived as¹⁰*

$$\mathbb{E}[\text{SLNR}_k] \geq \frac{\rho \mathbf{w}_k^\dagger \mathbf{V}^\dagger \mathbf{R}_{\text{eff},k} \mathbf{V} \mathbf{w}_k}{\mathbf{w}_k^\dagger \mathbf{V}^\dagger \left(\mathbf{I}_{N_{BS}} + \rho \mathbf{R}_{\text{res},k} + \rho \sum_{j \neq k} \mathbf{R}_j \right) \mathbf{V} \mathbf{w}_k}, \quad (20)$$

where $\mathbf{R}_{\text{eff},k} = \frac{K_k}{K_k+1} \mathbf{h}_{\text{LOS},k} \mathbf{h}_{\text{LOS},k}^\dagger + \frac{1}{K_k+1} \mathbf{\Sigma}_{\text{eff},k}$, $\mathbf{\Sigma}_{\text{eff},k} = \mathbf{U}_{\text{eff},k} \mathbf{\Lambda}_{\text{eff},k} \mathbf{U}_{\text{eff},k}^\dagger$, $\mathbf{R}_{\text{res},k} = \frac{1}{K_k+1} \mathbf{\Sigma}_{\text{res},k}$, $\mathbf{\Sigma}_{\text{res},k} = \mathbf{U}_{\text{res},k} \mathbf{\Lambda}_{\text{res},k} \mathbf{U}_{\text{res},k}^\dagger$, and $\mathbf{R}_j = \frac{K_j}{K_j+1} \mathbf{h}_{\text{LOS},j} \mathbf{h}_{\text{LOS},j}^\dagger + \frac{1}{K_j+1} \mathbf{\Sigma}_j$.

Proof: The proof is given in Appendix D. ■

A. Baseband Precoder Design

The transmitter treats the reconstructed channel statistics as the real ones, thus, from (20), the baseband precoder design

¹⁰This lower bound reduces to equality for a large number of antennas [22, Eq. (10)].

problem is formulated as

$$\mathbf{w}_k^* = \arg \max_{\mathbf{w}_k} \frac{\rho \mathbf{w}_k^\dagger \mathbf{V}^\dagger \mathbf{R}_{\text{eff},k} \mathbf{V} \mathbf{w}_k}{\mathbf{w}_k^\dagger (\mathbf{V}^\dagger \mathbf{V} + \rho \sum_{j \neq k} \mathbf{V}^\dagger \mathbf{R}_{\text{eff},j} \mathbf{V}) \mathbf{w}_k} \quad \text{s. t.} \quad \|\mathbf{V} \mathbf{w}_k\|^2 = 1 \quad (21)$$

Proposition 7. *The optimal digital precoder that maximizes (21) is derived as*

$$\mathbf{w}_k^* = \frac{\mathbf{u}_{\max} \left[\left(\mathbf{V}^\dagger \mathbf{V} + \rho \sum_{j \neq k} \mathbf{V}^\dagger \mathbf{R}_{\text{eff},j} \mathbf{V} \right)^{-1} \mathbf{V}^\dagger \mathbf{R}_{\text{eff},k} \mathbf{V} \right]}{\left\| \mathbf{V} \mathbf{u}_{\max} \left[\left(\mathbf{V}^\dagger \mathbf{V} + \rho \sum_{j \neq k} \mathbf{V}^\dagger \mathbf{R}_{\text{eff},j} \mathbf{V} \right)^{-1} \mathbf{V}^\dagger \mathbf{R}_{\text{eff},k} \mathbf{V} \right] \right\|} \quad (22)$$

Proof: The proof follows similar steps as the derivation of Proposition 1. ■

Given $\mathbf{w}_k^*$ in (22), the resultant lower-bound on $\mathbb{E}[\text{SLNR}_k]$ in (21) reduces to

$$\mathbb{E}[\text{SLNR}_k] \geq \rho \lambda_{\max} \left[\left(\mathbf{V}^\dagger \mathbf{V} + \rho \sum_{j \neq k} \mathbf{V}^\dagger \mathbf{R}_{\text{eff},j} \mathbf{V} \right)^{-1} \mathbf{V}^\dagger \mathbf{R}_{\text{eff},k} \mathbf{V} \right] \quad (23)$$

B. Analog Precoder Design

The optimization problem for the analog RF precoder $\mathbf{V}^*$ is formulated as

$$\mathbf{V}^* = \arg \max_{\mathbf{V}} \sum_{k=1}^{\mathcal{K}} \mathbb{E}[\text{SLNR}_k] \quad (24)$$

where the lower-bound on $\mathbb{E}[\text{SLNR}_k]$ is given in (23).

Proposition 8. *A sub-optimal solution to the analog precoder problem in (24) is derived as*

$$\mathbf{V}^* = \mathbf{U}_{[1:N_{RF}]} \left[\sum_{k=1}^{\mathcal{K}} \left(\mathbf{I}_{N_{BS}} + \rho \sum_{j=1}^{\mathcal{K}} \mathbf{R}_{\text{eff},j} \right)^{-1} \mathbf{R}_{\text{eff},k} \right] \quad (25)$$

Proof: The proof is given in Appendix E. ■

V. TWO-STAGE CSIT

As was previously mentioned, relying exclusively on statistical CSIT creates a performance bottleneck, due to the remaining multiuser interference; this, in turn, creates a saturation of the spectral efficiency at high SNRs. Therefore, to alleviate this issue, a two-stage precoding design technique is developed here, where the analog precoder stage is designed based on the knowledge of reduced rank statistical CSIT to steer the beams towards the dominant eigendirections of the desired users while eliminating the impact of multiuser interference in the statistical sense. Moreover, the baseband precoder stage eliminates the multiuser interference by utilizing only the instantaneous knowledge of a reduced

dimensionality effective channel. In particular, in the two-stage CSIT structure, $\mathbf{V}$ is designed exploiting only ρ and $\mathbf{R}_{\text{eff},k} \forall k \in \{1, 2, \dots, \mathcal{K}\}$, while the design of $\mathbf{W}$ utilizes the reduced dimensionality reconstructed effective channel $\widehat{\mathbf{H}}_{\text{eff}} = \mathbf{V}^\dagger \widehat{\mathbf{H}}$, where $\widehat{\mathbf{H}} = [\widehat{\mathbf{h}}_1, \widehat{\mathbf{h}}_2, \dots, \widehat{\mathbf{h}}_{\mathcal{K}}]$. Hence, rather than acquiring $N_{BS} \times \mathcal{K}$ instantaneous CSIT elements for $\widehat{\mathbf{H}}$, the BS obtains only $N_{RF} \times \mathcal{K}$ instantaneous CSIT elements for $\widehat{\mathbf{H}}_{\text{eff}}$ instead. This is a huge overhead reduction for the estimation and feedback of the instantaneous CSIT as the DL training and UL feedback overhead scales down to the fewer number of RF chains (i.e., the number of scheduled users) rather than the rank of the channel covariance matrix.

A. Baseband Precoder Design

The baseband precoder is designed based on the knowledge of the reduced dimensionality reconstructed effective channel $\widehat{\mathbf{H}}_{\text{eff}} = \mathbf{V}^\dagger \widehat{\mathbf{H}}$. As such, the two-stage baseband precoder problem formulation and derivations follow similar steps to that of the full CSI case in (12) and (13).

Proposition 9. *The optimal two-stage baseband precoder is given as*

$$\mathbf{w}_k^* = \frac{\left(\mathbf{V}^\dagger \mathbf{V} + \rho \sum_{j \neq k} \mathbf{V}^\dagger \widehat{\mathbf{h}}_j \widehat{\mathbf{h}}_j^\dagger \mathbf{V} \right)^{-1} \mathbf{V}^\dagger \widehat{\mathbf{h}}_k}{\left\| \mathbf{V} \left(\mathbf{V}^\dagger \mathbf{V} + \rho \sum_{j \neq k} \mathbf{V}^\dagger \widehat{\mathbf{h}}_j \widehat{\mathbf{h}}_j^\dagger \mathbf{V} \right)^{-1} \mathbf{V}^\dagger \widehat{\mathbf{h}}_k \right\|} \quad (26)$$

Proof: This is exactly similar to the derived optimal baseband precoder in (13). ■

Note that in contrast to (13), $\mathbf{V}$ in (26) relies on statistical CSIT. Therefore, the average of (14), $\mathbb{E}[\text{SLNR}_k]$ is lower-bounded as

$$\mathbb{E}[\text{SLNR}_k] = \mathbb{E} \left[\rho \widehat{\mathbf{h}}_k^\dagger \mathbf{V} \left(\mathbf{V}^\dagger \mathbf{V} + \rho \sum_{j \neq k} \mathbf{V}^\dagger \widehat{\mathbf{h}}_j \widehat{\mathbf{h}}_j^\dagger \mathbf{V} \right)^{-1} \mathbf{V}^\dagger \widehat{\mathbf{h}}_k \right] \quad (27)$$

$$= \rho \mathbb{E} \left[\text{Tr} \left(\left(\mathbf{V}^\dagger \mathbf{V} + \rho \sum_{j \neq k} \mathbf{V}^\dagger \widehat{\mathbf{h}}_j \widehat{\mathbf{h}}_j^\dagger \mathbf{V} \right)^{-1} \mathbf{V}^\dagger \widehat{\mathbf{h}}_k \widehat{\mathbf{h}}_k^\dagger \mathbf{V} \right) \right] \quad (28)$$

$$\geq \rho \text{Tr} \left(\left(\mathbf{V}^\dagger \mathbf{V} + \rho \sum_{j \neq k} \mathbf{V}^\dagger \mathbf{R}_{\text{eff},j} \mathbf{V} \right)^{-1} \mathbf{V}^\dagger \mathbf{R}_{\text{eff},k} \mathbf{V} \right), \quad (29)$$

where the lower-bound in (29) is obtained by utilizing the Jensen's inequality through the convexity of the function.

B. Analog Precoder Design

The design of $\mathbf{V}$ relies on the dominant eigenspace channel statistics information, hence, we utilize the sum of the average SLNR_k of all users introduced in (24) as our design metric.

Proposition 10. A sub-optimal solution to the analog precoder in (24) is derived as

$$\mathbf{V}^* = \mathbf{U}_{[1:N_{RF}]} \left[\sum_{k=1}^{\mathcal{K}} \left(\mathbf{I}_{N_{BS}} + \rho \sum_{j=1}^{\mathcal{K}} \mathbf{R}_{eff,j} \right)^{-1} \mathbf{R}_{eff,k} \right]. \quad (30)$$

Proof: The proof follows similar steps to the derivations of (25). ■

C. Extension to Multi-cell Systems

In this subsection, we introduce the extension of the proposed precoders to the multi-cell scenario, i.e., L cells. In [16] and [31], inter-cell interference (inter-group interference) is being mitigated in the outer precoder by leveraging the channel covariance information, presuming that users in the same cell (group) share the same covariance statistics. On the contrary, we propose a two-stage precoder, where we design the analog precoder without such strong assumption, based on the channel statistics information in order to reduce acquisition and feedback overhead required for the digital precoder design. The full and statistical CSIT scenarios are also considered.

The received signal at the k^{th} user in cell j is given as

$$y_{jk} = \mathbf{h}_{jjk}^\dagger \mathbf{x}_j + \sum_{l=1, l \neq j}^L \sqrt{\frac{1}{\beta_{lj}}} \mathbf{h}_{ljk}^\dagger \mathbf{x}_l + n_{jk}, \quad (31)$$

where $\mathbf{h}_{ljk}$ denotes the channel vector from the l^{th} BS to the k^{th} user in cell j , distributed similarly to (1), β_{lj} denotes the path loss between the j^{th} BS to the l^{th} BS, assuming that $\beta_{jj} = 1$, $\mathbf{x}_l$ is the transmit vector of BS l , given as $\mathbf{x}_l = \sum_{i=1}^{\mathcal{K}_l} \sqrt{P_{li}} \mathbf{V}_l \mathbf{w}_{li} s_{li}$, n_{jk} is the AWGN, distributed as $n_{jk} \sim \mathcal{CN}(0, \sigma^2)$, $\mathbf{V}_l$ is the unconstrained analog precoder for BS l , $\mathbf{w}_{li}$ is the digital precoder for the i^{th} user in BS l , s_{li} is the payload data symbol intended for the i^{th} user in cell l , with its equally allocated power $P_{li} = \frac{P}{\mathcal{K}_l}$, and $\mathcal{K}_l$ is the number of scheduled users in cell l . The received signal y_{jk} in (31) can be simplified as

$$y_{jk} = \underbrace{\sqrt{P_{jk}} \mathbf{h}_{jjk}^\dagger \mathbf{V}_j \mathbf{w}_{jk} s_{jk}}_{\text{desired signal}} + \underbrace{\sum_{i=1, i \neq k}^{\mathcal{K}_j} \sqrt{P_{ji}} \mathbf{h}_{jjk}^\dagger \mathbf{V}_j \mathbf{w}_{ji} s_{ji}}_{\text{intra-cell interference}} + \underbrace{\sum_{l=1, l \neq j}^L \sum_{i=1}^{\mathcal{K}_l} \sqrt{\frac{P_{li}}{\beta_{lj}}} \mathbf{h}_{ljk}^\dagger \mathbf{V}_l \mathbf{w}_{li} s_{li}}_{\text{inter-cell interference}} + \underbrace{n_{jk}}_{\text{AWGN noise}}. \quad (32)$$

Therefore, the SINR at the k^{th} user in cell j is written as

$$\text{SINR}_{jk} = \frac{\rho_j \left| \mathbf{h}_{jjk}^\dagger \mathbf{V}_j \mathbf{w}_{jk} \right|^2}{1 + \sum_{l=1}^L \sum_{i=1}^{\mathcal{K}_l} \frac{\rho_l}{\beta_{li}} \left| \mathbf{h}_{ljk}^\dagger \mathbf{V}_l \mathbf{w}_{li} \right|^2 - \rho_j \left| \mathbf{h}_{jjk}^\dagger \mathbf{V}_j \mathbf{w}_{jk} \right|^2}, \quad (33)$$

where $\rho_l = \frac{P}{\mathcal{K}_l \sigma^2}$. The per-cell sum spectral efficiency is given as $\mathcal{R}_{\text{sum},j} = \sum_{k=1}^{\mathcal{K}_j} \mathbb{E} [\log_2 (1 + \text{SINR}_{jk})]$.

Meanwhile, the SLNR at the k^{th} user in cell j is defined as

$$\text{SLNR}_{jk} = \frac{\rho_j \left| \mathbf{h}_{jjk}^\dagger \mathbf{V}_j \mathbf{w}_{jk} \right|^2}{1 + \sum_{l=1}^L \sum_{i=1}^{\mathcal{K}_l} \frac{\rho_l}{\beta_{li}} \left| \mathbf{h}_{ljk}^\dagger \mathbf{V}_l \mathbf{w}_{li} \right|^2 - \rho_j \left| \mathbf{h}_{jjk}^\dagger \mathbf{V}_j \mathbf{w}_{jk} \right|^2}. \quad (34)$$

1) *Full CSIT:* The problem formulation for the digital and analog precoders in multi-cell systems can be defined as $\mathbf{w}_{jk}^* = \arg \max_{\mathbf{w}_{jk}} \text{SLNR}_{jk}$ and $\mathbf{V}_j^* = \arg \max_{\mathbf{V}_j} \sum_{k=1}^{\mathcal{K}_j} \text{SLNR}_{jk}$, respectively, with the constraint that $\|\mathbf{V}_j \mathbf{w}_{jk}\| = 1$. Following similar steps to the derivations of (13) and (19), it is straightforward to show that the digital and analog precoders can be derived as in (35) and (36), shown at the top of next page.

2) *Statistical CSIT:* The problem formulation for the digital and analog precoders in multi-cell systems can be defined as $\mathbf{w}_{jk}^* = \arg \max_{\mathbf{w}_{jk}} \mathbb{E} [\text{SLNR}_{jk}]$ and $\mathbf{V}_j^* = \arg \max_{\mathbf{V}_j} \sum_{k=1}^{\mathcal{K}_j} \mathbb{E} [\text{SLNR}_{jk}]$, respectively, with the constraint that $\|\mathbf{V}_j \mathbf{w}_{jk}\| = 1$. Following similar steps to the derivations of (22) and (25), the digital precoder can be derived as in (37), shown at the top of next page, while the analog precoder is given as

$$\mathbf{V}_j^* = \mathbf{U}_{[1:N_{RF}]} \left[\sum_{k=1}^{\mathcal{K}_j} \left(\mathbf{I}_{N_{BS}} + \rho_j \sum_{l=1}^L \sum_{i=1}^{\mathcal{K}_l} \frac{1}{\beta_{jl}} \mathbf{R}_{jli} \right)^{-1} \mathbf{R}_{jjk} \right]. \quad (38)$$

3) *Two-Stage CSIT:* The problem formulation for $\mathbf{w}_{jk}$ and $\mathbf{V}_j$ in multi-cell systems are given by $\mathbf{w}_{jk}^* = \arg \max_{\mathbf{w}_{jk}} \text{SLNR}_{jk}$ and $\mathbf{V}_j^* = \arg \max_{\mathbf{V}_j} \sum_{k=1}^{\mathcal{K}_j} \mathbb{E} [\text{SLNR}_{jk}]$, respectively, with the constraint that $\|\mathbf{V}_j \mathbf{w}_{jk}\| = 1$. Following similar steps to the derivations of (26) and (30), the digital precoder can be derived as in (39), shown at the top of next page, while the analog precoder is written as

$$\mathbf{V}_j^* = \mathbf{U}_{[1:N_{RF}]} \left[\sum_{k=1}^{\mathcal{K}_j} \left(\mathbf{I}_{N_{BS}} + \rho_j \sum_{l=1}^L \sum_{i=1}^{\mathcal{K}_l} \frac{1}{\beta_{jl}} \mathbf{R}_{jli} \right)^{-1} \mathbf{R}_{jjk} \right]. \quad (40)$$

It is to be emphasized that the design of the $\mathbf{w}_{jk}^*$ and $\mathbf{V}_j^*$ in (39)-(40) are based on the knowledge of all multi-cell channel statistics and effective channels $\mathbf{R}_{jli}$, $\mathbf{V}_j^\dagger \mathbf{h}_{jli}$, respectively, along with ρ_j and $\beta_{jl} \forall k, l, i$. In the single cell scenario, i.e., $L = 1$, equations (39) and (40) reduce to (26) and (30), respectively.

VI. USER SCHEDULING

Multiuser diversity could be leveraged to further enhance the achievable sum spectral efficiency of the system. The BS can intelligently schedule a subset of users to be served

$$\mathbf{w}_{jk}^* = \frac{\left(\mathbf{V}_j^\dagger \mathbf{V}_j + \rho_j \mathbf{V}_j^\dagger \left(\sum_{l=1}^L \sum_{i=1}^{\mathcal{K}_l} \frac{1}{\beta_{jl}} \mathbf{h}_{jli} \mathbf{h}_{jli}^\dagger - \mathbf{h}_{jjk} \mathbf{h}_{jjk}^\dagger \right) \mathbf{V}_j \right)^{-1} \mathbf{V}_j^\dagger \mathbf{h}_{jjk}}{\left\| \mathbf{V}_j \left(\mathbf{V}_j^\dagger \mathbf{V}_j + \rho_j \mathbf{V}_j^\dagger \left(\sum_{l=1}^L \sum_{i=1}^{\mathcal{K}_l} \frac{1}{\beta_{jl}} \mathbf{h}_{jli} \mathbf{h}_{jli}^\dagger - \mathbf{h}_{jjk} \mathbf{h}_{jjk}^\dagger \right) \mathbf{V}_j \right)^{-1} \mathbf{V}_j^\dagger \mathbf{h}_{jjk} \right\|} \quad (35)$$

and

$$\mathbf{V}_j^* = \mathbf{U}_{[1:N_{RF}]} \left[\sum_{k=1}^{\mathcal{K}_j} \left(\mathbf{I}_{N_{BS}} + \rho_j \sum_{l=1}^L \sum_{i=1}^{\mathcal{K}_l} \frac{1}{\beta_{jl}} \mathbf{h}_{jli} \mathbf{h}_{jli}^\dagger \right)^{-1} \mathbf{h}_{jjk} \mathbf{h}_{jjk}^\dagger \right]. \quad (36)$$

$$\mathbf{w}_{jk}^* = \frac{\mathbf{u}_{\max} \left[\left(\mathbf{V}_j^\dagger \mathbf{V}_j + \rho_j \mathbf{V}_j^\dagger \left(\sum_{l=1}^L \sum_{i=1}^{\mathcal{K}_l} \frac{1}{\beta_{jl}} \mathbf{R}_{jli} - \mathbf{R}_{jjk} \right) \mathbf{V}_j \right)^{-1} \mathbf{V}_j^\dagger \mathbf{R}_{jjk} \mathbf{V}_j \right]}{\left\| \mathbf{V}_j \mathbf{u}_{\max} \left[\left(\mathbf{V}_j^\dagger \mathbf{V}_j + \rho_j \mathbf{V}_j^\dagger \left(\sum_{l=1}^L \sum_{i=1}^{\mathcal{K}_l} \frac{1}{\beta_{jl}} \mathbf{R}_{jli} - \mathbf{R}_{jjk} \right) \mathbf{V}_j \right)^{-1} \mathbf{V}_j^\dagger \mathbf{R}_{jjk} \mathbf{V}_j \right] \right\|}. \quad (37)$$

$$\mathbf{w}_{jk}^* = \frac{\left(\mathbf{V}_j^\dagger \mathbf{V}_j + \rho_j \mathbf{V}_j^\dagger \left(\sum_{l=1}^L \sum_{i=1}^{\mathcal{K}_l} \frac{1}{\beta_{jl}} \mathbf{h}_{jli} \mathbf{h}_{jli}^\dagger - \mathbf{h}_{jjk} \mathbf{h}_{jjk}^\dagger \right) \mathbf{V}_j \right)^{-1} \mathbf{V}_j^\dagger \mathbf{h}_{jjk}}{\left\| \mathbf{V}_j \left(\mathbf{V}_j^\dagger \mathbf{V}_j + \rho_j \mathbf{V}_j^\dagger \left(\sum_{l=1}^L \sum_{i=1}^{\mathcal{K}_l} \frac{1}{\beta_{jl}} \mathbf{h}_{jli} \mathbf{h}_{jli}^\dagger - \mathbf{h}_{jjk} \mathbf{h}_{jjk}^\dagger \right) \mathbf{V}_j \right)^{-1} \mathbf{V}_j^\dagger \mathbf{h}_{jjk} \right\|}. \quad (39)$$

when the available number of users $K \gg N_{RF}$, resulting in $\mathcal{S} \subset \{1, 2, \dots, K\}$, $|\mathcal{S}| \leq N_{RF}$ scheduled users, where $\mathcal{S}$ is the scheduler output and $|\mathcal{S}|$ is the cardinality of $\mathcal{S}$. In the presence of instantaneous CSIT, the BS utilizes both the channel gain and direction information in order to further improve the achievable sum spectral efficiency of the system, i.e., the BS can select users with high channel gains and at the same time spatially well separated in order to reduce multiuser interference [35]–[37]. However, user scheduling is a combinatorial optimization problem, where the optimal user set $\mathcal{S}^*$ requires a brute-force exhaustive search over all possible combinations of user sets $\mathcal{S}$, and the size of the search space grows exponentially with the number of users K and number of RF chains N_{RF} through $\sum_{i=1}^{N_{RF}} K! / (i! (K-i)!)$. Owing to the prohibitive complexity of exhaustive search over all sets of possible user combinations, low-complexity sub-optimal user scheduling techniques are more desirable. Specifically, in [36], the authors proposed a greedy user scheduler based on ZF precoding, by projecting each user's channel from the users' pool onto the orthogonal complement of the subspace spanned by the scheduled users, where the user with the highest projected channel is scheduled. The authors in [37] extended the work in [36] by introducing a semi-orthogonal user scheduler that enforces orthogonality amongst the scheduled users.

Most previous works on user scheduling require full CSIT, which is a fundamental bottleneck not only in FDD multiuser massive MIMO systems, but also TDD systems when the number of users K is large. More recently, a two-phase-feedback user scheduler and precoder for FDD multiuser massive MIMO systems was proposed in [35] for the two-stage precoder. In this work, the scheduling of semi-orthogonal users with large channel gain is introduced without full CSI feedback from all users. Scheduling is done via orthogonal reference beams utilizing only the channel covariance matrices

in the first phase. In the second phase, only scheduled users are required to feedback their CSIT to further reduce multiuser interference utilizing ZF precoding. However, the authors assumed that all users in each group have the same channel covariance matrix. Note that in order to have a group of users that share the same channel covariance matrix, this by itself is a user scheduling exercise that requires a massive number of users [25].

Motivated by the proposed analog and baseband precoders, we hereafter investigate three low-complexity leakage based user scheduling algorithms. Firstly, a joint scheduler and analog precoder followed by a baseband precoder based on full CSIT of the reconstructed channels is introduced for benchmarking. Second, in order to have the least possible CSIT acquisition and feedback overhead at the cost of extra multiuser interference, the statistical CSIT case is considered, where the joint scheduler and analog precoder followed by a baseband precoder are designed based on the reconstructed channel statistics information. Finally, to alleviate the burden of DL pilot training and UL feedback overhead required for the one-stage full CSIT, and at the same time reduce the negative impact of multiuser interference in the one-stage statistical CSIT, a two-stage approach is proposed, where the joint scheduler and beamformer stage leverages only the reduced rank channel statistics information of the K users, followed by a baseband precoder stage that relies on the reduced dimensionality reconstructed effective channel of the scheduled users.

More specifically, at any given time, the BS serves up to N_{RF} users concurrently. Let us denote the scheduler output as $\mathcal{S} = \{\pi(1), \pi(2), \dots, \pi(|\mathcal{S}|)\}$, where $\pi(i)$ is the index of the i^{th} scheduled user. Thus, here, y_k and SINR_k expressions are defined similarly to (7) and (8), respectively, by replacing $\sum_{j \neq k}$ with $\sum_{j \in \mathcal{S}, j \neq k}$.

Therefore, the achievable sum spectral efficiency of the

Algorithm 1 Full CSIT User Scheduler

- 1) Initialization: Select the first scheduled user:

$$\mathcal{T} = \{1, 2, \dots, K\} \quad (42)$$

$$\mathcal{S} = \{\} \quad (43)$$

$$\pi(1) = \arg \max_{k \in \mathcal{T}} \|\hat{\mathbf{h}}_k\| \quad (44)$$

$$\mathcal{S} \leftarrow \mathcal{S} \cup \{\pi(1)\} \quad (45)$$

$$\mathcal{T} \leftarrow \mathcal{T} \setminus \{\pi(1)\}, \quad (46)$$
 where $\mathcal{T}$ is the K users pool, $\{\}$ denotes the empty set and $\setminus$ denotes remove or subtract.
- 2) While $|\mathcal{S}| < N_{RF}$, schedule the other users:

$$\widehat{\mathbf{H}}(\mathcal{S}) = [\hat{\mathbf{h}}_{\pi(1)}, \hat{\mathbf{h}}_{\pi(2)}, \dots, \hat{\mathbf{h}}_{\pi(|\mathcal{S}|)}] \quad (47)$$

$$\pi(|\mathcal{S}| + 1) = \arg \max_{k \in \mathcal{T}} \hat{\mathbf{h}}_k^\dagger \left(\mathbf{I}_{N_{BS}} + \rho \widehat{\mathbf{H}}(\mathcal{S}) \widehat{\mathbf{H}}^\dagger(\mathcal{S}) \right)^{-1} \hat{\mathbf{h}}_k \quad (48)$$

$$\mathcal{S} \leftarrow \mathcal{S} \cup \{\pi(|\mathcal{S}| + 1)\} \quad (49)$$

$$\mathcal{T} \leftarrow \mathcal{T} \setminus \{\pi(|\mathcal{S}| + 1)\}. \quad (50)$$
- 3) RF Analog Precoder:

$$\mathbf{V}^* = \mathbf{U}_{[1:N_{RF}]} \left[\sum_{k \in \mathcal{S}} \left(\mathbf{I}_{N_{BS}} + \rho \sum_{j \in \mathcal{S}} \hat{\mathbf{h}}_j \hat{\mathbf{h}}_j^\dagger \right)^{-1} \hat{\mathbf{h}}_k \hat{\mathbf{h}}_k^\dagger \right]. \quad (51)$$
- 4) Digital Baseband Precoder:

$$\mathbf{w}_k^* = \frac{\left(\mathbf{V}^\dagger \mathbf{V} + \rho \sum_{j \in \mathcal{S}, j \neq k} \mathbf{V}^\dagger \hat{\mathbf{h}}_j \hat{\mathbf{h}}_j^\dagger \mathbf{V} \right)^{-1} \mathbf{V}^\dagger \hat{\mathbf{h}}_k}{\left\| \mathbf{V} \left(\mathbf{V}^\dagger \mathbf{V} + \rho \sum_{j \in \mathcal{S}, j \neq k} \mathbf{V}^\dagger \hat{\mathbf{h}}_j \hat{\mathbf{h}}_j^\dagger \mathbf{V} \right)^{-1} \mathbf{V}^\dagger \hat{\mathbf{h}}_k \right\|}. \quad (52)$$

system is written as

$$\mathbf{R}_{\text{sum}}(\mathcal{S}) = \sum_{k \in \mathcal{S}} \mathbb{E} [\log_2 (1 + \text{SINR}_k)]. \quad (41)$$

A. One-Stage Full CSIT User Scheduler

User scheduling in the full CSIT assumption for FDD multiuser massive MIMO systems is infeasible due to the overwhelming training symbols and feedback overhead required for the acquisition of instantaneous CSIT, which is limited by the channel coherence time. Note that in practice the channel is estimated at each user with the help of training sequences and fed back to the BS through a dedicated finite rate feedback channel. Therefore, the assumption of full CSIT is impractical (and included here only for benchmarking) as there will be inherently estimation and quantization errors that depend on the pilot symbols' length, power and codebook length, respectively. Long training sequences with high power and long codebook results in small error in the acquisition of CSIT. However, this consumes huge amount of the channel coherence time and system resources, leaving very small fraction for the transmission of useful information. Given that $\mathbf{V}$ and $\mathbf{W}$ in (19) and (13), respectively, are designed based on the reconstructed channels, a low-complexity one-stage instantaneous CSIT user scheduling approach is introduced in Algorithm 1.

In full CSIT, the BS leverages both channel gain and direction information for the proposed one-stage instantaneous CSIT user scheduling approach. Therefore, in order to further

Algorithm 2 Statistical and Two-Stage CSIT User Scheduler

- 1) Initialization: Select the first two scheduled users:

$$\mathcal{T} = \{1, 2, \dots, K\} \quad (53)$$

$$\mathcal{S} = \{\} \quad (54)$$

$$\{\pi(1), \pi(2)\} = \arg \min_{\substack{k \& l \in \mathcal{T} \\ k \neq l}} \left| \mathbf{u}_{\max}^\dagger [\mathbf{R}_{\text{eff}}, k] \mathbf{u}_{\max} [\mathbf{R}_{\text{eff}}, l] \right| \quad (55)$$

$$\mathcal{S} \leftarrow \mathcal{S} \cup \{\pi(1), \pi(2)\} \quad (56)$$

$$\mathcal{T} \leftarrow \mathcal{T} \setminus \{\pi(1), \pi(2)\}. \quad (57)$$
- 2) While $|\mathcal{S}| < N_{RF}$, schedule the other users:

$$\widehat{\mathbf{H}}(\mathcal{S}) = [\hat{\mathbf{h}}_{\pi(1)}, \hat{\mathbf{h}}_{\pi(2)}, \dots, \hat{\mathbf{h}}_{\pi(|\mathcal{S}|)}] \quad (58)$$

$$\pi(|\mathcal{S}| + 1) = \arg \max_{k \in \mathcal{T}} \lambda_{\max} \left[\left(\mathbf{I}_{N_{BS}} + \rho \sum_{j \in \mathcal{S}} \mathbf{R}_{\text{eff}, j} \right)^{-1} \mathbf{R}_{\text{eff}, k} \right] \quad (59)$$

$$\mathcal{S} \leftarrow \mathcal{S} \cup \{\pi(|\mathcal{S}| + 1)\} \quad (60)$$

$$\mathcal{T} \leftarrow \mathcal{T} \setminus \{\pi(|\mathcal{S}| + 1)\}. \quad (61)$$
- 3) RF Analog Precoder:

$$\mathbf{V}^* = \mathbf{U}_{[1:N_{RF}]} \left[\sum_{k \in \mathcal{S}} \left(\mathbf{I}_{N_{BS}} + \rho \sum_{j \in \mathcal{S}} \mathbf{R}_{\text{eff}, j} \right)^{-1} \mathbf{R}_{\text{eff}, k} \right]. \quad (62)$$
- 4) Digital Baseband Precoder (Statistical CSIT Case):

$$\mathbf{w}_k^* = \frac{\mathbf{u}_{\max}^\dagger \left[\left(\mathbf{V}^\dagger \mathbf{V} + \rho \sum_{j \in \mathcal{S}, j \neq k} \mathbf{V}^\dagger \mathbf{R}_{\text{eff}, j} \mathbf{V} \right)^{-1} \mathbf{V}^\dagger \mathbf{R}_{\text{eff}, k} \mathbf{V} \right]}{\left\| \mathbf{V} \mathbf{u}_{\max} \left[\left(\mathbf{V}^\dagger \mathbf{V} + \rho \sum_{j \in \mathcal{S}, j \neq k} \mathbf{V}^\dagger \mathbf{R}_{\text{eff}, j} \mathbf{V} \right)^{-1} \mathbf{V}^\dagger \mathbf{R}_{\text{eff}, k} \mathbf{V} \right] \right\|}. \quad (63)$$
- 5) Digital Baseband Precoder (Two-Stage CSIT):

$$\mathbf{w}_k^* = \frac{\left(\mathbf{V}^\dagger \mathbf{V} + \rho \sum_{j \in \mathcal{S}, j \neq k} \mathbf{V}^\dagger \hat{\mathbf{h}}_j \hat{\mathbf{h}}_j^\dagger \mathbf{V} \right)^{-1} \mathbf{V}^\dagger \hat{\mathbf{h}}_k}{\left\| \mathbf{V} \left(\mathbf{V}^\dagger \mathbf{V} + \rho \sum_{j \in \mathcal{S}, j \neq k} \mathbf{V}^\dagger \hat{\mathbf{h}}_j \hat{\mathbf{h}}_j^\dagger \mathbf{V} \right)^{-1} \mathbf{V}^\dagger \hat{\mathbf{h}}_k \right\|}. \quad (64)$$

improve the achievable sum spectral efficiency, the BS schedule users that have high channel gains and semi-orthogonal to each other in order to reduce multiuser interference. More specifically, step 1 selects the user with the strongest channel, while step 2 successively (in every iteration) schedules the user with the least leakage (highest SLNR_k according to (48)) to the scheduled users in (47), then $\mathbf{V}$ and $\mathbf{W}$ are designed in steps 3 and 4 respectively based on the scheduled users $\widehat{\mathbf{H}}(\mathcal{S})$.

B. One-Stage Statistical CSIT User Scheduler

Exploiting statistical CSIT for FDD multiuser massive MIMO systems is preferable as it incurs the least possible training symbols and feedback overhead, especially when its performance is close to the full CSIT case. Therefore, when the performance gap between the full CSIT and statistical CSIT is small, e.g., at low to medium SNRs and/or high multiuser diversity K and number of BS antennas N_{BS} , statistical CSIT is far more desirable. In contrast to full CSIT, only channels statistics directional information are leveraged here through the long-term first and reduced rank second order channel statistics. Thus, users are scheduled only based on

Figure 1. The sum spectral efficiency versus the SNR for the unconstrained $\mathbf{V}$ and constrained $\mathbf{F}$ precoders, with $N_{BS} = 64$, and $\mathcal{K} = N_{RF} = 10$ users.

their statistical degree of orthogonality, i.e., semi-orthogonal dominant eigendirections are selected. Hence, based on the derived $\mathbf{V}$ and $\mathbf{W}$ in (25) and (22), respectively, a low-complexity one-stage statistical CSIT user scheduler is envisioned in Algorithm 2.

In the case of one-stage statistical CSIT, due to the lack of channel gain information, the BS leverages only the channel statistics (reconstructed statistical eigendirections) information to schedule users. Firstly, two users which have the highest degree of statistical semi-orthogonality are scheduled in step 1 according to (55). Then, successively, the user which has the least leakage to the scheduled users according to (59) is scheduled in step 2. Hence, $\mathbf{V}$ and $\mathbf{W}$ are designed in steps 3 and 4 respectively based on the reconstructed channel statistics of the scheduled users $\hat{\mathbf{H}}(\mathcal{S})$. Therefore, in the presence of statistical multiuser diversity, the BS can schedule users that have the least leakage and have non overlapping angular support based only on statistical CSIT, thus resulting in $\mathbf{U}_i^\dagger \mathbf{U}_j \rightarrow \mathbf{0} \forall i \neq j$ asymptotically for a large number of users K and BS antennas.

C. Algorithm 3: Two-Stage CSIT User Scheduler

The rationale behind the two-stage CSIT approach is justified by the gap between the considered one-stage CSIT extremes, the full and statistical CSIT approaches. If the gap is huge, e.g., at high SNRs and low number of BS antennas and multiuser diversity K , the statistical CSIT case saturates and cannot provide any sum spectral efficiency improvement at high SNRs. Therefore, in this case, based on the derived $\mathbf{V}$ and $\mathbf{W}$ in (30) and (26), respectively, a two-stage CSIT user scheduling approach is proposed in Algorithm 2. In the first-stage, we jointly schedule users and design the analog precoder based on the dominant eigenspace reconstructed channel statistics information. Statistical multiuser diversity provides several statistical eigenspaces (channel statistics eigendirections) which are leveraged here to schedule statistically semi-orthogonal users, i.e., steps 1-3 in Algorithm 2, whereas in the second-stage, full CSIT of the reduced dimensionality reconstructed effective channels $\mathbf{V}^\dagger \hat{\mathbf{h}}_j$, $j = 1, \dots, |\mathcal{S}|$ of the

Figure 2. The sum spectral efficiency versus the Rician K-factor for the unconstrained precoder $\mathbf{V}$, with $\Delta_k = 10^\circ \forall k$, $\rho = 10$ dB, $N_{BS} = 32$, and $\mathcal{K} = N_{RF} = 10$ users.

scheduled users are acquired and fed back to the BS, through which the baseband precoder is designed in step 5.

VII. NUMERICAL RESULTS

In this section, a set of numerical results are provided to evaluate the performance of the proposed analog and baseband precoders. In our simulations, we set $\phi_k \sim \mathcal{U}(-\frac{\pi}{2}, \frac{\pi}{2})$, $\Delta_k = 15^\circ \forall k$, and $K_k = 10$ dB $\forall k$ except for Fig. 2.¹¹

In Fig. 1, the achievable sum spectral efficiency of the proposed precoders are shown for the unconstrained analog precoder $\mathbf{V}$ and constrained analog precoder $\mathbf{F}$. The results of Figs. 1-2 are based on the full knowledge of $\mathbf{R}_k$, $\mathbf{h}_k \forall k$, i.e., utilizing all r_k nonzero eigenvalues and the corresponding eigenvectors of the channel covariance matrices. It is clearly seen that the performance of the proposed unconstrained and constrained precoders are very close. The sum spectral efficiency is evaluated by substituting the resultant SINR_k from (8) into (9). The full CSIT scenario is evaluated by substituting (13) and (19) into (8) for the unconstrained case, while utilizing $\mathbf{F}$ derived from (19) according to (11) for the constrained case. In addition, the statistical CSIT scenario is estimated by substituting (22) and (25) into (8) for the unconstrained case, while exploiting $\mathbf{F}$ derived from (25) according to (11) for the constrained case. Furthermore, the two-stage CSIT scenario is analyzed by substituting (26) and (30) into (8) for the unconstrained case, while using $\mathbf{F}$ derived from (30) according to (11) for the constrained case.

The achievable sum spectral efficiency of the proposed two-stage CSIT precoder is very close to that of its full CSIT counterpart, whereas the performance of the statistical CSIT scenario saturates at high SNRs. This comes at a remarkably less number of DL training sequences and UL feedback overhead. Traditionally, the full CSIT case acquires $N_{BS} \times \mathcal{K}$ instantaneous CSIT elements for $\mathbf{H}$ at the BS. However, given that correlated massive MIMO channels are rank deficient, the full CSIT case acquires about $\sum_{k=1}^{\mathcal{K}} r_k \leq \max(r_1, \dots, r_{\mathcal{K}}) \times \mathcal{K}$ instantaneous CSIT elements for $\mathbf{H}$ at

¹¹In order to evaluate the average sum spectral efficiency over different user locations, the sum spectral efficiency is averaged over random AoDs herein.

Figure 3. The sum spectral efficiency versus the SNR for the dominant eigenspace reconstructed channel, with $N_{BS} = 64$, and $K = N_{RF} = 10$ users.

the BS, which requires a proportional number of DL training sequences and UL feedback overhead. On the other hand, the two-stage CSIT scenario acquires only $N_{RF} \times K$ (where $N_{RF} \ll \max(r_1, \dots, r_K)$) instantaneous CSIT elements for $\mathbf{H}_{\text{eff}} = \mathbf{V}^\dagger \mathbf{H}$ at the BS. Therefore, for every channel coherence block, with the settings of Fig. 1, it was found that the highest possible rank is $\max(r_1, \dots, r_K) = 50$, hence, in this case, the full CSIT precoder acquires at most 50×10 channel elements while the two-stage CSIT precoder acquires only 10×10 channel elements. This is a substantial reduction in the number of DL training sequences and UL feedback overhead. Nevertheless, statistical CSIT has the least overhead burden as the channel statistics remain fixed over multiple channel coherence intervals.

Fig. 2 presents the sum spectral efficiency versus the Rician K-factor for the unconstrained analog precoder $\mathbf{V}$. In contrast to the full CSIT case, the Rician K-factor has a higher impact on the statistical and two-stage CSIT scenarios, where a significant performance improvement is noticed as a result of higher Rician K-factor. This is expected as statistical CSIT relies solely on the first and second order channel statistics, whereas the analog precoding stage in the two-stage CSIT precoder depend on the channel statistics information. For comparison purposes, the MRT beamformer derived in [23, Theorem 1] is considered from the literature.

In Fig. 3, the achievable sum spectral efficiency of the proposed precoders for the dominant eigenspace reconstructed channel (6), is shown for the constrained analog precoder $\mathbf{F}$. The sum spectral efficiency is analyzed here by substituting the SINR_k from (8) into (9). The SINR_k in (8) quantifies the impact of the remaining residual channel error on the performance of the system. The reconstructed channel of each user leverages only the $r_{\text{eff}, k}$ (based on (5)) dominant eigenvalues and their corresponding eigenvectors, whereas the classical case utilizes all the r_k nonzero eigenvalues and their corresponding eigenvectors which consumes considerable amount of unnecessary downlink training and uplink feedback overhead. The inefficiency of exploiting r_k is clearly seen from Fig. 3 through its tightness with the proposed $r_{\text{eff}, k}$.

Figure 4. The sum spectral efficiency versus an arbitrarily chosen rank for the dominant eigenspace reconstructed channel with $\rho = 15$ dB, $N_{BS} = 64$, and $K = N_{RF} = 10$ users.

Figure 5. The sum spectral efficiency versus the SNR for the constrained precoder $\mathbf{F}$, with $N_{BS} = 64$, $K = 10$ & 100 users and $K = N_{RF} = 10$ scheduled users.

For finite size codebooks, in contrast to using r_k channel eigenmodes, utilizing only $r_{\text{eff}, k}$ dominant eigenmodes would result in extremely less quantization errors, thus resulting in higher system performance.

Fig. 4 depicts the achievable sum spectral efficiency versus an arbitrarily chosen rank ($r_k = r \forall k$) which is used for the dominant eigenspace reconstructed channel. The impact of arbitrarily choosing the rank (number of dominant eigendirections) to reconstruct the channel covariance matrices and their channels is clearly seen. A substantial performance improvement is noticed for the full and two-stage CSIT scenarios as the rank increases until saturation beyond a given rank. The specified rank beyond which the performance saturates is the optimal rank, which is about $r = 27$ in these settings. This matches very well with the proposed effective rank $r_{\text{eff}, k}$ in (5), that is given in Definition 2. The key point here is that the same spectral efficiency performance could be attained with the least possible DL training sequences and UL feedback overhead, that is only proportionate to $r_{\text{eff}, k}$. However, statistical CSIT shows a marginal performance improvement with the rank before ceiling. This is expected as statistical CSIT utilizes only the dominant eigenmode of the channel

Figure 6. The sum spectral efficiency versus the number of users for the constrained precoder $\mathbf{F}$ with $\rho = 15$ dB, $N_{BS} = 16$ & 64 , and $\mathcal{K} = N_{RF} = 10$ scheduled users.

correlation matrix.

In Fig. 5, the sum spectral efficiency versus the SNR for the dominant eigenspace reconstructed channels is presented, where $\mathcal{K}$ users are scheduled from the available number of users K . The positive impact of user scheduling on the achievable sum spectral efficiency is clearly seen, where a remarkable performance improvement is observed as a result of user scheduling. The quantification of the gap between the one-stage full and statistical CSIT user schedulers determines the importance of the two-stage CSIT user scheduler. We first point out that the one-stage full CSIT user scheduler acquires $\sum_{k=1}^K r_{\text{eff},k}$ instantaneous CSIT elements for $\widehat{\mathbf{H}}$ at the BS, while the two-stage CSIT user scheduler acquires only $N_{RF} \times \mathcal{K}$ instantaneous CSIT elements for $\widehat{\mathbf{H}}_{\text{eff}} = \mathbf{V}^\dagger \widehat{\mathbf{H}}$, with the emphasis that $N_{RF} \ll r_{\text{eff},k} \ll N_{BS}$ and $\mathcal{K} \leq N_{RF} \ll K$. On the other hand, there are no instantaneous CSIT requirements for the one-stage statistical CSIT user scheduler. Our purpose here is to show that when the performance of the one-stage statistical CSIT user scheduler approaches that of the two-stage CSIT user scheduler, i.e., if the gap between the one-stage statistical CSIT and two-stage CSIT user schedulers is small, then the one-stage statistical CSIT user scheduler is far more desirable, whereas the proposed two-stage CSIT user scheduler is justifiable if the gap is huge. It is to be emphasized that the performance loss due to the DL training sequences and UL feedback overhead has not been taken into account in (9) and (41).

Fig. 5 shows that in the presence of massive MIMO systems, very little spectral efficiency improvement is seen by the one-stage instantaneous and the two-stage CSIT user schedulers, while the one-stage statistical CSIT user scheduler shows remarkable performance enhancement, that works very close to the two-stage CSIT user scheduler at higher SNRs. For example, with statistical multiuser diversity of $K = 100$ users where only $\mathcal{K} = 10$ users are scheduled, the performance of the one-stage statistical CSIT user scheduler approaches that of the two-stage CSIT user scheduler when $\rho \leq 10$ dB. However, without multiuser diversity, the performance gap is very small only when $\rho \leq 5$ dB with increasingly higher gap

Figure 7. The sum spectral efficiency versus the SNR for the constrained precoder $\mathbf{F}$, with $K = 100$ users, and $\mathcal{K} = N_{RF} = 4$ scheduled users.

Figure 8. The per-cell sum spectral efficiency versus the SNR for multi-cell systems with the unconstrained precoder $\mathbf{V}$, for $N_{BS} = 32$, $\beta_{jl} = 20 \forall l \neq j$, $L = 3$, and $\mathcal{K}_l = N_{RF} = 4 \forall l$.

at higher SNRs. This is thanks to the multiuser eigenspace diversity which at asymptotically high K and N_{BS} , results in scheduling users that have nearly orthogonal eigenspaces $\mathbf{U}_i^\dagger \mathbf{U}_j \rightarrow \mathbf{0} \forall i \neq j$ without any instantaneous CSIT.

In Fig. 6, the effect of multiuser diversity K on the sum spectral efficiency performance is shown. It is clearly seen that multiuser diversity has huge impact on the sum spectral efficiency at lower number of BS antennas, while very little impact on the performance of one-stage full CSIT and two-stage CSIT user schedulers at massive number of BS antennas. However, the one-stage statistical CSIT user scheduler still shows remarkable performance improvement, reducing the gap with its two-stage CSIT user scheduler counterpart.

Fig. 7 shows the sum spectral efficiency performance comparison between the proposed user scheduling in Algorithm 1 and greedy ZF user selection algorithm proposed in [36]. Due to the two-stage nature of our system model, the proposed ZF precoder in (16) is utilized here instead. It is clearly seen that for low to medium number of BS antennas, the proposed leakage based algorithm outperforms the greedy ZF algorithm, whereas at high number of BS antennas, the performance of these two schedulers coincide. However, this comes with

the merit of outstanding reduction in CSIT acquisition and feedback overhead.

In Fig. 8, the per-cell sum spectral efficiency of the proposed analog and digital precoders in multi-cell systems are shown for the unconstrained analog precoder $\mathbf{V}$. The negative impact of inter-cell interference is clearly seen when compared to the single cell scenario ($L = 1$). For comparison with the literature, the multi-cell minimum mean squared error (M-MMSE) precoder, proposed for the traditional fully digital structure in [38, Eq. (37)] is included. The performance of the proposed hybrid precoding with full CSIT matches that of the M-MMSE precoder from the literature. This not only comes with substantially less cost and power consumption, but also reduced training pilots and feedback overhead. However, due to the effect of inter-cell and intra-cell interferences, the performances of the proposed two-stage CSIT and statistical CSIT saturate at high SNRs.

VIII. CONCLUSIONS

A new hybrid precoding technique that leverages the reconstructed reduced rank long-term channel statistics to jointly design a scheduler and analog precoder was proposed for FDD multiuser massive MIMO systems. Moreover, two CSIT scenarios were considered for the digital precoder design, the statistical CSIT and full CSIT of the reduced dimensionality reconstructed effective channel. Exploiting the rank deficiency of the per-user channel covariance matrix, a reconstructed channel with its covariance statistics based on a reduced effective rank was developed. The analog and digital precoders were jointly derived by utilizing the SLNR metric. The motivation for developing the two-stage CSIT beamformer was based on the performance gap between the full CSIT and statistical CSIT beamformers. It was clearly shown that the sum spectral efficiency of the unconstrained analog precoder performs very close to that of the constrained analog precoder. Furthermore, the performance of the proposed two-stage CSIT beamformer is very close to that of the full CSIT beamformer with the merit of outstanding reduction in CSIT acquisition and feedback overhead. Finally, with the aid of the proposed analog and digital precoders, a low-complexity one-stage statistical CSIT user scheduler and two-stage CSIT user scheduler were proposed to further enhance the sum spectral efficiency by utilizing the concept of statistical multiuser diversity.

APPENDIX A PROOF FOR LEMMA 3

The generalized Rayleigh quotient problem in (12) is efficiently solved by the generalized eigenvalue decomposition method, as follows: An equivalent desired formulation for the unconstrained problem in (12) can be re-written as

$$\begin{aligned} \mathbf{w}_k^* &= \arg \max_{\mathbf{w}_k} \mathbf{w}_k^\dagger \mathbf{V}^\dagger \hat{\mathbf{h}}_k \hat{\mathbf{h}}_k^\dagger \mathbf{V} \mathbf{w}_k \\ \text{s. t.} \quad & \mathbf{w}_k^\dagger \left(\mathbf{V}^\dagger \mathbf{V} + \rho \sum_{j \neq k} \mathbf{V}^\dagger \hat{\mathbf{h}}_j \hat{\mathbf{h}}_j^\dagger \mathbf{V} \right) \mathbf{w}_k = c \end{aligned} \quad (65)$$

where c is constant.

This problem can be easily solved using the Lagrange multiplier, as follows

$$L(\mathbf{w}_k) = \mathbf{w}_k^\dagger \mathbf{V}^\dagger \hat{\mathbf{h}}_k \hat{\mathbf{h}}_k^\dagger \mathbf{V} \mathbf{w}_k - \kappa \left(\mathbf{w}_k^\dagger \left(\mathbf{V}^\dagger \mathbf{V} + \rho \sum_{j \neq k} \mathbf{V}^\dagger \hat{\mathbf{h}}_j \hat{\mathbf{h}}_j^\dagger \mathbf{V} \right) \mathbf{w}_k - c \right) \quad (66)$$

where κ is the Lagrange multiplier.

Therefore, once we maximize with respect to $\mathbf{w}_k$, we get

$$\mathbf{V}^\dagger \hat{\mathbf{h}}_k \hat{\mathbf{h}}_k^\dagger \mathbf{V} \mathbf{w}_k = \kappa \left(\mathbf{V}^\dagger \mathbf{V} + \rho \sum_{j \neq k} \mathbf{V}^\dagger \hat{\mathbf{h}}_j \hat{\mathbf{h}}_j^\dagger \mathbf{V} \right) \mathbf{w}_k. \quad (67)$$

This is a generalized eigenvalue problem, i.e., $\mathbf{w}_k^*$ is the generalized eigenvector corresponding to the largest generalized eigenvalue $\lambda_{\max}$ of $\left\{ \mathbf{V}^\dagger \hat{\mathbf{h}}_k \hat{\mathbf{h}}_k^\dagger \mathbf{V}, \mathbf{V}^\dagger \mathbf{V} + \rho \sum_{j \neq k} \mathbf{V}^\dagger \hat{\mathbf{h}}_j \hat{\mathbf{h}}_j^\dagger \mathbf{V} \right\}$. Given that $\left(\mathbf{V}^\dagger \mathbf{V} + \rho \sum_{j \neq k} \mathbf{V}^\dagger \hat{\mathbf{h}}_j \hat{\mathbf{h}}_j^\dagger \mathbf{V} \right)$ is invertible, the generalized eigenvalue problem reduces to the standard eigenvalue problem. Hence, (67) reduces to

$$\left(\mathbf{V}^\dagger \mathbf{V} + \rho \sum_{j \neq k} \mathbf{V}^\dagger \hat{\mathbf{h}}_j \hat{\mathbf{h}}_j^\dagger \mathbf{V} \right)^{-1} \mathbf{V}^\dagger \hat{\mathbf{h}}_k \hat{\mathbf{h}}_k^\dagger \mathbf{V} \mathbf{w}_k = \kappa \mathbf{w}_k. \quad (68)$$

Therefore, the optimal baseband precoder is derived as

$$\begin{aligned} \mathbf{w}_k^* &= \beta \mathbf{u}_{\max} \left[\left(\mathbf{V}^\dagger \mathbf{V} + \rho \sum_{j \neq k} \mathbf{V}^\dagger \hat{\mathbf{h}}_j \hat{\mathbf{h}}_j^\dagger \mathbf{V} \right)^{-1} \mathbf{V}^\dagger \hat{\mathbf{h}}_k \hat{\mathbf{h}}_k^\dagger \mathbf{V} \right] \\ &= \beta \left(\mathbf{V}^\dagger \mathbf{V} + \rho \sum_{j \neq k} \mathbf{V}^\dagger \hat{\mathbf{h}}_j \hat{\mathbf{h}}_j^\dagger \mathbf{V} \right)^{-1} \mathbf{V}^\dagger \hat{\mathbf{h}}_k, \end{aligned} \quad (69)$$

where β is a normalization constant that satisfies the power constraint $\|\mathbf{V} \mathbf{w}_k^*\|^2 = 1$. Therefore, once taking this normalization into account, (69) reduces to (13), and this concludes the proof.

APPENDIX B PROOF FOR COROLLARY 4

The optimal solution for (15) is achieved by utilizing the maximal-ratio combining (MRC) principle, in the following form $\mathbf{w}_k = \frac{\mathbf{P}_k^\perp \mathbf{V}^\dagger \hat{\mathbf{h}}_k}{\|\mathbf{V} \mathbf{P}_k^\perp \mathbf{V}^\dagger \hat{\mathbf{h}}_k\|}$, where both constraints are exploited in this form, more specifically, 1) the null-space projection matrix $\mathbf{P}_k^\perp$ guarantees that $\mathbf{w}_k \perp \mathbf{V}^\dagger \hat{\mathbf{H}}_k$ (due to the fact that $\hat{\mathbf{h}}_k^\dagger \mathbf{V} \mathbf{w}_j = 0 \ \forall \ k \neq j$), 2) the division by $\|\mathbf{V} \mathbf{P}_k^\perp \mathbf{V}^\dagger \hat{\mathbf{h}}_k\|$ satisfies the norm constraint on $\mathbf{w}_k$, such that $\|\mathbf{V} \mathbf{w}_k\|^2 = 1$. This concludes the proof.

In light of comparison between (13) and the ZF scheme, we use the Woodbury matrix identity, as follows

$$(\mathbf{I} + \mathbf{A} \mathbf{B})^{-1} = \mathbf{I} - \mathbf{A} (\mathbf{I} + \mathbf{B} \mathbf{A})^{-1} \mathbf{B} \quad (70)$$

Exploiting the fact that the analog precoder satisfies $\mathbf{V}^\dagger \mathbf{V} \approx \mathbf{I}_{N_{BS}}$ when N_{BS} is large [17], (69) can be simplified as in (71) and (72), shown at the top of next page.

$$\beta \left(\mathbf{I}_{N_{RF}} + \rho \sum_{j \neq k} \mathbf{V}^\dagger \hat{\mathbf{h}}_j \hat{\mathbf{h}}_j^\dagger \mathbf{V} \right)^{-1} \mathbf{V}^\dagger \hat{\mathbf{h}}_k = \beta \left(\mathbf{I}_{N_{RF}} + \rho \mathbf{V}^\dagger \hat{\mathbf{H}}_k \hat{\mathbf{H}}_k^\dagger \mathbf{V} \right)^{-1} \mathbf{V}^\dagger \hat{\mathbf{h}}_k \quad (71)$$

$$= \beta \left(\mathbf{I}_{N_{RF}} - \mathbf{V}^\dagger \hat{\mathbf{H}}_k \left(\frac{1}{\rho} \mathbf{I}_{N_{RF}} + \hat{\mathbf{H}}_k^\dagger \mathbf{V} \mathbf{V}^\dagger \hat{\mathbf{H}}_k \right)^{-1} \hat{\mathbf{H}}_k^\dagger \mathbf{V} \right) \mathbf{V}^\dagger \hat{\mathbf{h}}_k. \quad (72)$$

APPENDIX C PROOF FOR PROPOSITION 5

Utilizing the identity $\mathbf{x}^\dagger \mathbf{A} \mathbf{x} = \text{Tr}(\mathbf{A} \mathbf{x} \mathbf{x}^\dagger)$, equation (18) is re-written as

$$\begin{aligned} & \sum_{k=1}^{\mathcal{K}} \text{SLNR}_k \\ &= \sum_{k=1}^{\mathcal{K}} \rho \mathbf{h}_k^\dagger \mathbf{V} \left(\mathbf{V}^\dagger \mathbf{V} + \rho \sum_{j \neq k} \mathbf{V}^\dagger \hat{\mathbf{h}}_j \hat{\mathbf{h}}_j^\dagger \mathbf{V} \right)^{-1} \mathbf{V}^\dagger \hat{\mathbf{h}}_k \\ &= \rho \text{Tr} \left(\sum_{k=1}^{\mathcal{K}} \left(\mathbf{V}^\dagger \mathbf{V} + \rho \sum_{j \neq k} \mathbf{V}^\dagger \hat{\mathbf{h}}_j \hat{\mathbf{h}}_j^\dagger \mathbf{V} \right)^{-1} \mathbf{V}^\dagger \hat{\mathbf{h}}_k \hat{\mathbf{h}}_k^\dagger \mathbf{V} \right) \\ &\geq \rho \text{Tr} \left(\sum_{k=1}^{\mathcal{K}} \left(\mathbf{V}^\dagger \mathbf{V} + \rho \sum_{j=1}^U \mathbf{V}^\dagger \hat{\mathbf{h}}_j \hat{\mathbf{h}}_j^\dagger \mathbf{V} \right)^{-1} \mathbf{V}^\dagger \hat{\mathbf{h}}_k \hat{\mathbf{h}}_k^\dagger \mathbf{V} \right) \end{aligned} \quad (73)$$

where the inequality in (73) comes from the rank-1 perturbation lemma. We emphasize that the introduced rank-1 perturbation lemma entails almost sure convergence as the number of BS antennas approaches infinity [3].

Hence, a sub-optimal solution to the analog precoder $\mathbf{V}$ is given as

$$\begin{aligned} & \mathbf{V}^* = \\ & \arg \max_{\mathbf{V}} \text{Tr} \left(\sum_{k=1}^{\mathcal{K}} \left(\mathbf{V}^\dagger \mathbf{V} + \rho \sum_{j=1}^{\mathcal{K}} \mathbf{V}^\dagger \hat{\mathbf{h}}_j \hat{\mathbf{h}}_j^\dagger \mathbf{V} \right)^{-1} \mathbf{V}^\dagger \hat{\mathbf{h}}_k \hat{\mathbf{h}}_k^\dagger \mathbf{V} \right). \end{aligned} \quad (74)$$

With the aid of (67)-(68), we can re-write $(\mathbf{V}^\dagger \mathbf{B} \mathbf{V})^{-1} \mathbf{V}^\dagger \mathbf{A} \mathbf{V} = \mathbf{V}^\dagger \mathbf{B}^{-1} \mathbf{A} \mathbf{V}$, hence, (74) reduces to [39, Eq. (7)]

$$\begin{aligned} & \mathbf{V}^* = \\ & \arg \max_{\mathbf{V}^\dagger \mathbf{V} = \mathbf{I}} \text{Tr} \left(\mathbf{V}^\dagger \sum_{k=1}^{\mathcal{K}} \left(\mathbf{I}_{N_{BS}} + \rho \sum_{j=1}^{\mathcal{K}} \hat{\mathbf{h}}_j \hat{\mathbf{h}}_j^\dagger \right)^{-1} \hat{\mathbf{h}}_k \hat{\mathbf{h}}_k^\dagger \mathbf{V} \right). \end{aligned} \quad (75)$$

The optimal solution to (75) is the $N_{BS} \times N_{RF}$ matrix containing the N_{RF} leading eigenvectors associated with the N_{RF} strongest eigenvalues of the matrix $\sum_{k=1}^{\mathcal{K}} (\mathbf{I}_{N_{BS}} + \rho \sum_{j=1}^{\mathcal{K}} \hat{\mathbf{h}}_j \hat{\mathbf{h}}_j^\dagger)^{-1} \hat{\mathbf{h}}_k \hat{\mathbf{h}}_k^\dagger$ [39], [40].

APPENDIX D PROOF FOR PROPOSITION 6

From (10), in statistical CSIT, $\mathbf{V}$ and $\mathbf{w}_k$ are deterministic and do not depend on the instantaneous variations of $\mathbf{h}_k$, hence, the average SLNR_k is expressed as

$$\begin{aligned} & \mathbb{E}[\text{SLNR}_k] \\ &= \mathbb{E} \left[\frac{\rho \mathbf{w}_k^\dagger \mathbf{V}^\dagger \hat{\mathbf{h}}_k \hat{\mathbf{h}}_k^\dagger \mathbf{V} \mathbf{w}_k}{1 + \rho \mathbf{w}_k^\dagger \mathbf{V}^\dagger \tilde{\mathbf{h}}_k \tilde{\mathbf{h}}_k^\dagger \mathbf{V} \mathbf{w}_k + \rho \sum_{j \neq k} \mathbf{w}_k^\dagger \mathbf{V}^\dagger \mathbf{h}_j \mathbf{h}_j^\dagger \mathbf{V} \mathbf{w}_k} \right] \\ &= \mathbb{E} \left[\rho \mathbf{w}_k^\dagger \mathbf{V}^\dagger \hat{\mathbf{h}}_k \hat{\mathbf{h}}_k^\dagger \mathbf{V} \mathbf{w}_k \right] \times \\ & \quad \mathbb{E} \left[\frac{1}{\mathbf{w}_k^\dagger \mathbf{V}^\dagger (\mathbf{I}_{N_{BS}} + \rho \tilde{\mathbf{h}}_k \tilde{\mathbf{h}}_k^\dagger + \rho \sum_{j \neq k} \mathbf{h}_j \mathbf{h}_j^\dagger) \mathbf{V} \mathbf{w}_k} \right] \\ &\geq \frac{\rho \mathbf{w}_k^\dagger \mathbf{V}^\dagger \mathbb{E}[\hat{\mathbf{h}}_k \hat{\mathbf{h}}_k^\dagger] \mathbf{V} \mathbf{w}_k}{\mathbf{w}_k^\dagger \mathbf{V}^\dagger (\mathbf{I}_{N_{BS}} + \rho \mathbb{E}[\tilde{\mathbf{h}}_k \tilde{\mathbf{h}}_k^\dagger] + \rho \sum_{j \neq k} \mathbb{E}[\mathbf{h}_j \mathbf{h}_j^\dagger]) \mathbf{V} \mathbf{w}_k}. \end{aligned} \quad (76)$$

The equality in (77) utilizes the independence between the numerator and denominator of SLNR_k in (76), while the inequality in (78) results from Jensen's inequality through the convexity of the function $f(x) = \frac{1}{x}$, $\forall x \geq 0$. Once evaluating the expectations in (78), we arrive at the desired result in (20).

APPENDIX E PROOF FOR PROPOSITION 8

We first recall the identity

$$\sum_{k=1}^{\mathcal{K}} \lambda_{\max}(\mathbf{R}_{\text{eff}, k}) \leq \text{Tr} \left(\sum_{k=1}^{\mathcal{K}} \mathbf{R}_{\text{eff}, k} \right). \quad (79)$$

From (23)-(24), a sub-optimal formulation for the analog precoder $\mathbf{V}^*$ is defined as

$$\begin{aligned} & \mathbf{V}^* = \\ & \arg \max_{\mathbf{V}} \sum_{k=1}^{\mathcal{K}} \text{Tr} \left(\left(\mathbf{V}^\dagger \mathbf{V} + \rho \sum_{j \neq k} \mathbf{V}^\dagger \mathbf{R}_{\text{eff}, j} \mathbf{V} \right)^{-1} \mathbf{V}^\dagger \mathbf{R}_{\text{eff}, k} \mathbf{V} \right). \end{aligned} \quad (80)$$

Consequently, by following similar steps with the derivation of (19), we arrive at (25), and this concludes the proof.

REFERENCES

- [1] A. Almradi, M. Matthaiou, P. Xiao, and V. Fusco, "Two-stage hybrid precoding in FDD massive MIMO systems with low rank correlated Rician fading channels," in *Proc. IEEE ICC Workshops*, To Appear.

- [2] M. Dai and B. Clerckx, "Multiuser millimeter wave beamforming strategies with quantized and statistical CSIT," *IEEE Trans. Wireless Commun.*, vol. 16, no. 11, pp. 7025–7038, Nov. 2017.
- [3] S. Park, J. Park, A. Yazdan, and R. W. Heath, Jr., "Exploiting spatial channel covariance for hybrid precoding in massive MIMO systems," *IEEE Trans. Signal Process.*, vol. 65, no. 14, pp. 3818–3832, Jul. 2017.
- [4] I. Boukhedimi, A. Kammoun, and M. S. Alouini, "LMMSE receivers in uplink massive MIMO systems with correlated Rician fading," *IEEE Trans. Commun.*, vol. 67, no. 1, pp. 230–243, Jan. 2019.
- [5] Z. Jiang, A. F. Molisch, G. Caire, and Z. Niu, "Achievable rates of FDD massive MIMO systems with spatial channel correlation," *IEEE Trans. Wireless Commun.*, vol. 14, no. 5, pp. 2868–2882, May 2015.
- [6] V. Raghavan *et al.*, "Single-user versus multi-user precoding for millimeter wave MIMO systems," *IEEE J. Sel. Areas Commun.*, vol. 35, no. 6, pp. 1387–1401, Jun. 2017.
- [7] A. Alkhateeb, G. Leus, and R. W. Heath, Jr., "Limited feedback hybrid precoding for multi-user millimeter wave systems," *IEEE Trans. Wireless Commun.*, vol. 14, no. 11, pp. 6481–6494, Nov. 2015.
- [8] A. Alkhateeb, J. Mo, N. Gonzalez-Prelcic, and R. W. Heath, Jr., "MIMO precoding and combining solutions for millimeter-wave systems," *IEEE Commun. Mag.*, vol. 52, no. 12, pp. 122–131, Dec. 2014.
- [9] A. F. Molisch *et al.*, "Hybrid beamforming for massive MIMO: A survey," *IEEE Commun. Mag.*, vol. 55, no. 9, pp. 134–141, Sep. 2017.
- [10] M. Sadek, A. Tarighat, and A. H. Sayed, "A leakage-based precoding scheme for downlink multi-user MIMO channels," *IEEE Trans. Wireless Commun.*, vol. 6, no. 5, pp. 1711–1721, May 2007.
- [11] —, "Active antenna selection in multiuser MIMO communications," *IEEE Trans. Signal Process.*, vol. 55, no. 4, pp. 1498–1510, Apr. 2007.
- [12] Q. Zhang, S. Jin, K. K. Wong, H. Zhu, and M. Matthaiou, "Power scaling of uplink massive MIMO systems with arbitrary-rank channel means," *IEEE J. Sel. Topics Signal Process.*, vol. 8, no. 5, pp. 966–981, Oct. 2014.
- [13] A. Liu and V. K. N. Lau, "Phase only RF precoding for massive MIMO systems with limited RF chains," *IEEE Trans. Signal Process.*, vol. 62, no. 17, pp. 4505–4515, Sep. 2014.
- [14] A. Adhikary, J. Nam, J. 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, Oct. 2013.
- [15] 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, Jun. 2014.
- [16] J. Chen and V. K. N. Lau, "Two-tier precoding for FDD multi-cell massive MIMO time-varying interference networks," *IEEE J. Sel. Areas Commun.*, vol. 32, no. 6, pp. 1230–1238, June 2014.
- [17] F. Sotrobiani and W. Yu, "Hybrid digital and analog beamforming design for large-scale antenna arrays," *IEEE J. Sel. Topics Signal Process.*, vol. 10, no. 3, pp. 501–513, Apr. 2016.
- [18] W. Ni, X. Dong, and W. Lu, "Near-optimal hybrid processing for massive MIMO systems via matrix decomposition," *IEEE Trans. Signal Process.*, vol. 65, no. 15, pp. 3922–3933, Aug. 2017.
- [19] W. Ni and X. Dong, "Hybrid block diagonalization for massive multiuser MIMO systems," *IEEE Trans. Commun.*, vol. 64, no. 1, pp. 201–211, Jan. 2016.
- [20] W. Utschick, C. Stockle, M. Joham, and J. Luo, "Hybrid LISA precoding for multiuser millimeter-wave communications," *IEEE Trans. Wireless Commun.*, vol. 17, no. 2, pp. 752–765, Feb. 2018.
- [21] C. Zhang, Y. Huang, Y. Jing, S. Jin, and L. Yang, "Sum-rate analysis for massive MIMO downlink with joint statistical beamforming and user scheduling," *IEEE Trans. Wireless Commun.*, vol. 16, no. 4, pp. 2181–2194, Apr. 2017.
- [22] V. Raghavan, J. J. Choi, and D. J. Love, "Statistical beamforming for the large antenna broadcast channel," in *Proc. IEEE ISIT*, June 2017, pp. 2508–2512.
- [23] S. Jin, W. Tan, M. Matthaiou, J. Wang, and K. K. Wong, "Statistical eigenmode transmission for the MU-MIMO downlink in Rician fading," *IEEE Trans. Wireless Commun.*, vol. 14, no. 12, pp. 6650–6663, Dec. 2015.
- [24] M. Wang, F. Gao, S. Jin, and H. Lin, "An overview of enhanced massive MIMO with array signal processing techniques," *IEEE J. Sel. Topics Signal Process.*, vol. 13, no. 5, pp. 886–901, Sep. 2019.
- [25] J. Nam, A. Adhikary, J. Ahn, and G. Caire, "Joint spatial division and multiplexing: Opportunistic beamforming, user grouping and simplified downlink scheduling," *IEEE J. Sel. Topics Signal Process.*, vol. 8, no. 5, pp. 876–890, Oct. 2014.
- [26] D. W. H. Cai, T. Q. S. Quek, and C. W. Tan, "A unified analysis of max-min weighted SINR for MIMO downlink system," *IEEE Trans. Signal Process.*, vol. 59, no. 8, pp. 3850–3862, Aug. 2011.
- [27] A. Adhikary *et al.*, "Joint spatial division and multiplexing for mm-wave channels," *IEEE J. Sel. Areas Commun.*, vol. 32, no. 6, pp. 1239–1255, June 2014.
- [28] O. Roy and M. Vetterli, "The effective rank: A measure of effective dimensionality," in *Proc. IEEE EUSIPCO*, Sep. 2007, pp. 606–610.
- [29] M. R. Castellanos *et al.*, "Channel-reconstruction-based hybrid precoding for millimeter-wave multi-user MIMO systems," *IEEE J. Sel. Topics Signal Process.*, vol. 12, no. 2, pp. 383–398, May 2018.
- [30] Y. Han, S. Jin, J. Zhang, J. Zhang, and K. K. Wong, "DFT-based hybrid beamforming multiuser systems: Rate analysis and beam selection," *IEEE J. Sel. Topics Signal Process.*, vol. 12, no. 3, pp. 514–528, Jun. 2018.
- [31] D. Kim, G. Lee, and Y. Sung, "Two-stage beamformer design for massive MIMO downlink by trace quotient formulation," *IEEE Trans. Commun.*, vol. 63, no. 6, pp. 2200–2211, Jun. 2015.
- [32] H. Xie, F. Gao, S. Jin, J. Fang, and Y. Liang, "Channel estimation for TDD/FDD massive MIMO systems with channel covariance computing," *IEEE Trans. Wireless Commun.*, vol. 17, no. 6, pp. 4206–4218, June 2018.
- [33] M. Barzegar Khalilsarai, S. Haghighatshoar, X. Yi, and G. Caire, "FDD massive MIMO via UL/DL channel covariance extrapolation and active channel sparsification," *IEEE Trans. Wireless Commun.*, vol. 18, no. 1, pp. 121–135, Jan. 2019.
- [34] K. Huang, H. Wang, J. Hou, and S. Jin, "Joint spatial division and diversity for massive MIMO systems," *IEEE Trans. Commun.*, vol. 67, no. 1, pp. 258–272, Jan. 2019.
- [35] G. Lee and Y. Sung, "A new approach to user scheduling in massive multi-user MIMO broadcast channels," *IEEE Trans. Commun.*, vol. 66, no. 4, pp. 1481–1495, Apr. 2018.
- [36] G. Dimic and N. D. Sidiropoulos, "On downlink beamforming with greedy user selection: Performance analysis and a simple new algorithm," *IEEE Trans. Signal Process.*, vol. 53, no. 10, pp. 3857–3868, Oct. 2005.
- [37] T. Yoo and A. Goldsmith, "On the optimality of multi-antenna broadcast scheduling using Zero-Forcing beamforming," *IEEE J. Sel. Areas Commun.*, vol. 24, no. 3, pp. 528–541, Mar. 2006.
- [38] E. Bjornson, J. Hoyerdis, and L. Sanguinetti, "Massive MIMO has unlimited capacity," *IEEE Transactions on Wireless Communications*, vol. 17, no. 1, pp. 574–590, Jan. 2018.
- [39] H. Wang, S. Yan, D. Xu, X. Tang, and T. Huang, "Trace ratio vs. ratio trace for dimensionality reduction," in *Proc. IEEE CVPR*, June 2007, pp. 1–8.
- [40] T. T. Ngo, M. Bellalij, and Y. Saad, "The trace ratio optimization problem," *SIAM Rev.*, vol. 54, no. 3, pp. 545–569, 2012.

Ahmed Almradi (S'14) received the B.Sc. degree from the University of Tripoli, Tripoli, Libya, in 2004, the M.Sc. degree from the Rochester Institute of Technology, Rochester, NY, USA in 2012, and the Ph.D. degree from The University of Manchester, Manchester, U.K, in 2017, all in electrical and electronic engineering. From November 2017 through March 2019, he was a Postdoctoral Research Fellow with the 5G Innovation Centre, Institute for Communication Systems, University of Surrey. He is currently a Postdoctoral Research Fellow at the Centre for Wireless Innovation, Queen's University Belfast. His research interests are in the modelling, design, and performance analysis of wireless communication systems with special emphasis on MIMO half-duplex and full-duplex relaying systems, wireless information and power transfer (energy harvesting) systems, multiuser massive MIMO in hybrid analogue and digital precoding systems.

Michail Matthaiou (S'05–M'08–SM'13) was born in Thessaloniki, Greece in 1981. He obtained the Diploma degree (5 years) in Electrical and Computer Engineering from the Aristotle University of Thessaloniki, Greece in 2004. He then received the M.Sc. (with distinction) in Communication Systems and Signal Processing from the University of Bristol, U.K. and Ph.D. degrees from the University of Edinburgh, U.K. in 2005 and 2008, respectively. From September 2008 through May 2010, he was with the Institute for Circuit Theory and Signal Processing,

Munich University of Technology (TUM), Germany working as a Postdoctoral Research Associate. He is currently a Reader (equivalent to Associate Professor) in Multiple-Antenna Systems at Queen's University Belfast, U.K. after holding an Assistant Professor position at Chalmers University of Technology, Sweden. His research interests span signal processing for wireless communications, massive MIMO systems, hardware-constrained communications, mm-wave systems and deep learning for communications. Dr. Matthaiou and his coauthors received the IEEE Communications Society (ComSoc) Leonard G. Abraham Prize in 2017. He was awarded the prestigious 2018/2019 Royal Academy of Engineering/The Leverhulme Trust Senior Research Fellowship and recently received the 2019 EURASIP Early Career Award. His team was also the Grand Winner of the 2019 Mobile World Congress Challenge. He was the recipient of the 2011 IEEE ComSoc Best Young Researcher Award for the Europe, Middle East and Africa Region and a co-recipient of the 2006 IEEE Communications Chapter Project Prize for the best M.Sc. dissertation in the area of communications. He has co-authored papers that received best paper awards at the 2018 IEEE WCSP and 2014 IEEE ICC and was an Exemplary Reviewer for IEEE COMMUNICATIONS LETTERS for 2010. In 2014, he received the Research Fund for International Young Scientists from the National Natural Science Foundation of China. He is currently the Editor-in-Chief of Elsevier Physical Communication and a Senior Editor for IEEE WIRELESS COMMUNICATIONS LETTERS. In the past, he was an Associate Editor for the IEEE TRANSACTIONS ON COMMUNICATIONS and Associate Editor/Senior Editor for IEEE COMMUNICATIONS LETTERS.

Pei Xiao (SM'11) is currently a professor of Wireless Communications at the Institute for Communication Systems, home of 5G Innovation Center (5GIC) at the University of Surrey. He is the technical manager of 5GIC, leading the research team in the new physical layer work area, and coordinating/supervising research activities across all the work areas within 5GIC (WWW.surrey.ac.uk/5gic/research). Prior to this, he worked at Newcastle University and Queen's University Belfast. He also held positions at Nokia Networks in Finland. His technical expertise and research interests are in the fields of communication theory and signal processing for wireless communications.

Vincent F. Fusco (S'82–M'82–SM'96–F'04) received the Bachelor's degree (Hons.) in electrical and electronic engineering, the Ph.D. degree in microwave electronics, and the D.Sc. degree from Queen's University Belfast (QUB), Belfast, U.K. in 1979, 1982, and 2000, respectively. He is currently the Director of Research with the ECIT Research Institute, QUB. His fundamental work on active antenna front-end techniques has provided generic advances in low-cost phased and self-tracking antenna array architectures. He has authored or co-authored more than 500 papers and 2 books. He holds a number of antenna-related patents. Dr. Fusco is a Fellow of the Institution of Engineering and Technology (IET) and the Royal Academy of Engineering. He was a recipient of the IET Senior Achievement Award and the Mountbatten Medal for seminal contributions in the field of microwave electronics and its impact on U.K. industry, in 2012.