AoD-Adaptive Channel Feedback in FDD Massive MIMO Systems with Multiple-Antenna Users

Mahmoud A. AlaaEldin*, Karim G. Seddik*, and Wessam Mesbah†

*Department of Electronics and Communications Engineering, American University in Cairo, Cairo, Egypt 11835
†Electrical Engineering Department, King Fahd University of Petroleum and Minerals, Dhahran, Saudi Arabia 31261
mahmoud_alaa_eldin@aucegypt.edu, kseddik@aucegypt.edu, mesbahw@gmail.com

***Abstract*—In this paper, we consider the problem of Angle of Departure (AoD) based channel feedback in Frequency Division Duplex (FDD) massive Multiple-Input Multiple-Output (MIMO) systems with multiple antennas at the users. We consider the use of Zero-Forcing Block Diagonalization (BD) as the downlink precoding scheme. We consider two different cases; one in which the number of streams intended for a user equals the number of antennas at that user and the other case in which the number of streams is less than the number of user antennas. BD requires the feedback of the subspace spanned by the channel matrix at the user or a subspace of it in the case of having less number of streams than the number of antennas at a specific user. Based on our channel model, we propose a channel feedback scheme that requires less feedback overhead compared to feeding back the whole channel matrix. Then, we quantify the rate gap between the rate of the system with perfect Channel State Information (CSI) at the massive MIMO Basestation (BS) and our proposed channel feedback scheme for a given number of feedback bits. Finally, we design feedback codebooks based on optimal subspace packing in the Grassmannian manifold. We show that our proposed codes achieve performance that is very close to the performance of the system with perfect CSI at the BS.**

I. Introduction

Massive MIMO wireless communication systems have been shown to introduce dramatic improvements in both spectral and energy efficiency [1]. Channel feedback is a crucial part in FDD massive MIMO systems to perform precoding and digital beam-forming on the transmitted signals. In FDD systems, channel reciprocity cannot be used to obtain the downlink CSI at the BS. Therefore, channel feedback is necessary. However, the challenge that massive MIMO systems face is that it has a very large number of antennas and hence, the codebook size is very large, and hence, the feedback overhead is overwhelming.

Many channel feedback schemes were proposed in order to reduce the amount of feedback overhead as well as the size of the codebook for massive MIMO systems. In [2], a spatially common sparsity-based adaptive channel estimation and feedback scheme for FDD massive MIMO systems was proposed. In [3], a compressed channel feedback scheme for correlated massive MIMO channels was proposed. The channels were quantized based on compressive sensing technique in order to be fed back to the base station with low overhead. A limited feedback scheme for massive MIMO systems based on principal component analysis (PCA) was discussed in [4]. In [5], an AoD-adaptive subspace codebook for channel feedback was proposed. The paper utilized the idea that the angles of departure vary much slower than the channel gains, which results in a massive reduction in the required feedback overhead. This is because that the channel vector is constrained to be in a lower dimensional subspace of the full M-dimensional space (where M is the number of transmitting antennas at the BS) during the angle coherence time. Exploiting this fact can result in a significant reduction in the required feedback overhead . However, the work in [5] did not consider the case of equipping the users with multiple receive antennas. In addition, the work in [5] has assumed random feedback codebooks and no structured feedback codebooks design was considered.

In this paper, we extend the massive MIMO model in [5], by equipping each user with multiple receive antennas rather than only a single antenna. We use the concept of AoD-adaptive subspace codebook to reduce the amount of required feedback overhead. We jointly feed back the CSI of the multiple receiving antennas at each user. This joint feedback results in a massive reduction of feedback bits compared to feeding back the CSI of each receive antenna separately. In order to achieve this, we use BD [6] as ourprecoding scheme, which is a generalization of the zero-forcing channel inversion technique. BD is a linear precoding scheme that involves simultaneous transmissions of multiple data streams to each user while cancelling the interference from other users. Hence, BD only needs the channel subspace of each user's channel matrix at the BS, which requires fewer feedback bits if compared to reporting the actual channel matrix. In addition, we present a BD-based AoD-adaptive codebook design. A structured quantization codebook design is proposed based on subspace packing in Grassmannian manifolds. Finally, we quantify the rate loss resulting from using the proposed codebook. We prove that the required number of feedback bits to achieve a constant rate gap, from the system with perfect CSI at the BS, only increases linearly with the Signal to Noise Ratio (SNR).

II. System Model

A. Downlink Massive MIMO Channel Model

In this paper, we assume a Millimeter Wave (mmWave) massive MIMO broadcast (downlink) system with a single BS communicating with K multi-antenna users. The BS has M transmitting antennas while the $k^{\text{th}}(\forall k \in \{1, 2, \cdots, K\})$ user has N_k receiving antennas. We assume that the typical model

in massive MIMO systems is used. This model assumes that the number of transmitting antennas is much higher than the number of users (i.e., $M >> K$). We consider a narrowband ray-based downlink channel model for the downlink channel vectors $\mathbf{H}_k \in \mathbb{C}^{N_k \times M}$ at the k^{th} user [5]

$$\mathbf{H}_k = \mathbf{G}_k \mathbf{A}_k(\theta_{k,1}, \theta_{k,2}, \cdots, \theta_{k,P_k}). \tag{1}$$

The matrix $\mathbf{A}_k(\theta_{k,1}, \theta_{k,2}, \cdots, \theta_{k,P_k}) \in \mathbb{C}^{P_k \times M}$ is defined as:

$$\mathbf{A}_k(\theta_{k,1}, \theta_{k,2}, \cdots, \theta_{k,P_k}) = \begin{bmatrix} \mathbf{a}(\theta_{k,1}) \\ \mathbf{a}(\theta_{k,2}) \\ \vdots \\ \mathbf{a}(\theta_{k,P_k}) \end{bmatrix} \tag{2}$$

where P_k is the number of resolvable paths from the BS to the k^{th} user. The parameter $\theta_{k,i} (1 \leq i \leq P_k)$ represents the AoDs of the i^{th} path of the k^{th} user. We assume that the transmitting antennas at the BS form a Uniform Linear Array (ULA) as in [7]. Hence, $\mathbf{a}(\theta_{k,i}) \in \mathbb{C}^{M \times 1}$ is a steering vector that represents the antenna response of the i^{th} propagation path of the k^{th} user, and it can be written as

$$\mathbf{a}(\theta_{k,i}) = [1, e^{-j2\pi \frac{d}{\lambda} \sin(\theta_{k,i})}, \cdots, e^{-j2\pi \frac{d}{\lambda}(M-1) \sin(\theta_{k,i})}], \tag{3}$$

where λ is the signal wavelength and d is the spacing between every two successive antennas at the BS. From Eq. (1), we can notice that the k^{th} user's channel vector for each antenna is a linear combination of its P_k steering vectors scaled by the complex paths' gains of that antenna. $\mathbf{G}_k \in \mathbb{C}^{N_k \times P_k}$ is a matrix whose rows contain the complex path gains of each antenna at the k^{th} user (i.e., the entry $\mathbf{G}_k(i,j)$ represents the complex gain of the j^{th} path of the i^{th} antenna at user k). The complex path gains in $\mathbf{G}_k$ are assumed to be Independently and Identically Distributed (i.i.d.) circularly-symmetric complex Gaussian random variables with zero mean and unit variance.

During the angle coherence time of $\theta_{k,i}$, the channel vector of each antenna of user k is only distributed in a P_k-dimensional subspace, called as the channel subspace in this paper, of the full M-dimensional space. We assume throughout the paper that the channel subspace $\mathbf{A}_k$, which is a function of the AoDs, is known at both user k and the BS. The AoDs can be estimated at the k^{th} user using the standard Multiple Signal Classification (MUSIC) algorithm [8], then they are fed back to the BS once after every angle coherence time. Consequently, the BS only needs to know the low dimensional path gains matrix $\mathbf{G}_k \in \mathbb{C}^{N_k \times P_k}$ in order to generate the actual channel matrix $\mathbf{H}_k$. In this paper, we neglect the overhead coming from reporting the AoDs to the BS because it is very low compared to the overhead coming from reporting the path gains in $\mathbf{G}_k$.

The BS sends m_k streams to user k, where $m_k \leq N_k$. Let $\mathbf{u}_k \in \mathbb{C}^{m_k \times 1}$ contains the m_k data symbols to be transmitted simultaneously to the k^{th} user such that

$$\mathbf{u}_k = [u_{k,1} u_{k,2} \cdots u_{k,m_k}]^T. \tag{4}$$

Before transmitting the users' data symbols over the channel, the k^{th} user symbol vector is multiplied by the precoding matrix $\mathbf{F}_k \in \mathbb{C}^{M \times m_k}$. Thus, the overall transmitted vector $\mathbf{x} \in \mathbb{C}^{M \times 1}$, which contains all the data symbols intended for all users, is given by:

$$\mathbf{x} = \sum_{j=1}^{K} \mathbf{F}_j \mathbf{u}_j \tag{5}$$

and the received signal at the k^{th} user can be written as:

$$\mathbf{y}_k = \mathbf{H}_k \mathbf{x} + \mathbf{n}_k = \mathbf{H}_k \mathbf{F}_k \mathbf{u}_k + \mathbf{H}_k \sum_{\substack{j=1 \\ j \neq k}}^{K} \mathbf{F}_j \mathbf{u}_j + \mathbf{n}_k, \tag{6}$$

where $\mathbf{n}_k \in \mathbb{C}^{N_k \times 1}$ is the circularly symmetric complex Gaussian noise vector at the k^{th} user with a zero vector mean and identity covariance matrix.

The second term in Eq. (6) represents the summation of the interference, from the signals intended to all other users in the cell, at user k. The users' precoding matrices, $\mathbf{F}_k$'s, are unitary matrices (i.e., $\mathbf{F}_k^{\text{H}} \mathbf{F}_k = \mathbf{I}_{m_k}$), and in order to adhere to the power constraint, we have $E\left[\|\mathbf{u}_k\|^2\right] = \frac{\gamma}{K}, \forall k \in \{1, 2, \cdots, K\}$, where γ is the total transmit power at the BS.

B. Partial CSI Feedback

The training overhead to perform channel estimation at the receiver side increases in massive MIMO systems as the number of transmit antennas at the BS increases [1]. However, there are many effective downlink channel estimation schemes that address this problem with a highly reduced amount of training overhead [2], [9], [10]. Consequently, we assume in this paper that each user knows its downlink channel matrix.

The channel matrix $\mathbf{H}_k$ of each user is required at the BS in order to perform precoding and power allocation. However, we assume in this paper that the total power of each user is uniformly allocated across its multiple data streams. Hence, in order to perform BD, which will be discussed thoroughly in Sec. III-A, it is only required to feed back the spatial direction of each user's effective channel. The spatial direction of the k^{th} user is defined as the subspace spanned by the rows of $\widetilde{\mathbf{H}}_k \in \mathbb{C}^{m_k \times M}$, where $\widetilde{\mathbf{H}}_k$ represents the subspace of the effective channel of user k. In case of $m_k = N_k$, the spatial direction of the k^{th} user is the subspace spanned by the rows of its channel matrix itself $\mathbf{H}_k \in \mathbb{C}^{N_k \times M}$. The quantization of the spatial direction $\widetilde{\mathbf{H}}_k$, say $\widehat{\mathbf{H}}_k \in \mathbb{C}^{m_k \times M}$, is chosen from the codebook $\mathcal{C}_k = \{\mathbf{C}_{k,1}, \mathbf{C}_{k,2}, \cdots, \mathbf{C}_{k,2^B}\}$, that consists of 2^B matrices in $\mathbb{C}^{m_k \times M}$, where B is the number of feedback bits for each user and the rows of $\mathbf{C}_{k,i}$ are orthonormal. The details of the beamforming matrix design as well as the codebook design are discussed in Sec. III and Sec. IV, respectively. The k^{th} user quantizes its spatial direction $\widetilde{\mathbf{H}}_k$ to a quantization subspace $\widehat{\mathbf{H}}_k = \mathbf{C}_{k,Z_k}$, where the index Z_k is calculated such that:

$$Z_k = \underset{i \in [1, 2^B]}{\arg\min}\, d^2(\widetilde{\mathbf{H}}_k, \mathbf{C}_{k,i}), \tag{7}$$

where $d(\mathbf{H}_k, \mathbf{C}_{k,i})$ is the distance metric between the two matrices $\mathbf{H}_k$ and $\mathbf{C}_{k,i}$. In this paper, we adopt the chordal distance as our distance metric [11], which is given by:

$$d(\widetilde{\mathbf{H}}_k, \mathbf{C}_{k,i}) = \sqrt{\sin^2\theta_1 + \sin^2\theta_2 + \cdots + \sin^2\theta_{m_k}}, \tag{8}$$

where the θ_j's are the principal angles between the two subspaces spanned by the rows of the matrices $\mathbf{H}_k$ and $\mathbf{C}_{k,i}$ [11]. The principal angles only depend on the subspaces spanned by the rows of the matrices. Hence, the rows of each matrix $\mathbf{C}_{k,i} \in \mathcal{C}_k$ are orthonormal (i.e., $\mathbf{C}_{k,i}\mathbf{C}_{k,i}^{\mathrm{H}} = \mathbf{I}_{m_k} \forall\ \mathbf{C}_{k,i} \in \mathcal{C}_k$), and each $\mathbf{C}_{k,i}$ represents a quantization subspace in the codebook. The chordal distance can be calculated using an alternate form of Eq. (8) as follows:

$$d(\widetilde{\mathbf{H}}_k, \mathbf{C}_{k,i}) = \left[N_k - \left\| \widetilde{\mathbf{H}}_k \mathbf{C}_{k,i}^{\mathrm{H}} \right\|_{\mathrm{F}}^2 \right]^{1/2}, \tag{9}$$

where the values of this distance metric range between 0 and $\sqrt{m_k}$. Note that we do not feed back any channel magnitude information to the BS.

III. Design of BD based beamforming matrices

In this section, we present the details of the block diagonalization (BD) precoding scheme. Then, we analyze the per-user data rates of the BD scheme.

A. Design of Users' Beamforming Matrices

In this paper, we consider BD as our linear BS precoding technique. BD is a zero-forcing technique which completely nulls the interference at each user due to the signals transmitted to all other users. Thus, BD can be thought of as a generalization of channel inversion in cases of multiple antennas per user. Following the BD algorithm, each $\mathbf{F}_k$ is chosen under the constraint of having $\mathbf{H}_j\mathbf{F}_k = \mathbf{0},\ \forall j \neq k$. This leads to obtaining an orthonormal basis for the null space of the matrix formed by stacking all $\{\mathbf{H}_j\}_{j\neq k}$ matrices. This procedure nulls the interference terms in Eq. (6) at each user. BD is different from the conventional Zero-Forcing (ZF) precoding, where every complex data symbol to be transmitted to the n^{th} antenna (among the N_k antennas) of the k^{th} user is precoded by a vector which is orthogonal to all the rows of $\mathbf{H}_j, j \neq k$, and is orthogonal to all rows of $\mathbf{H}_k$ except the n^{th} one. In other words, conventional ZF forces every transmitted data symbol to be received by only one antenna at the intended user. This results in more restrictions in designing the BS precoders and results in a degraded performance if compared to BD based precoders design.

However, in practice, we cannot achieve zero interference as the BS does not have perfect knowledge of $\{\mathbf{H}_k\}_{k=1}^K$. In the case of limited feedback, BS has access to a quantized version of the subspace spanned by the rows of each $\mathbf{H}_k$, namely $\widehat{\mathbf{H}}_k$. We follow the strategy in [12], where the BS treats the quantized subspaces $\widehat{\mathbf{H}}_1, \widehat{\mathbf{H}}_2, \cdots, \widehat{\mathbf{H}}_K$ as the true channel subspaces while performing the BD procedure. In that case, we denote the generated precoding matrices as $\widehat{\mathbf{F}}_1, \widehat{\mathbf{F}}_2, \cdots, \widehat{\mathbf{F}}_K$ in order to distinguish them from those selected with perfect channel knowledge at the BS.

We assume in this paper that the number of antennas of the k^{th} user, N_k, is smaller than the number of resolvable paths P_k, (i.e., $N_k < P_k$). Thus, all antennas of user k are independent from each other since they experience P_k independent paths with independent path gains (i.e., entries of $\mathbf{G}_k$ are independent). We consider two different cases when designing the precoding matrices $\widehat{\mathbf{F}}_k,\ \forall k \in \{1, 2, \cdots, K\}$ as follows.

1) **Case I: $N_k = m_k$:** In this case, it is assumed that the number of antennas of the k^{th} user, N_k, is equal to the number of complex data symbols m_k to be simultaneously transmitted to it. Define $\widetilde{\mathbf{W}}_k$ as

$$\widetilde{\mathbf{W}}_k = \left[\widehat{\mathbf{H}}_1^T \cdots \widehat{\mathbf{H}}_{k-1}^T \widehat{\mathbf{H}}_{k+1}^T \cdots \widehat{\mathbf{H}}_K^T \right]^T, \tag{10}$$

where $\widehat{\mathbf{H}}_k,\ k \in \{1, 2, \cdots, K\}$, is the quantized feedback version of the original spatial direction $\widetilde{\mathbf{H}}_k$ of the k^{th} user. The zero-interference constraint forces the precoding matrix $\widehat{\mathbf{F}}_k$ of the k^{th} user to lie in the null space of $\widetilde{\mathbf{W}}_k$. The channel subspace of the k^{th} user $\mathbf{A}_k$ only depends on the AoDs of the user which are assumed to be independent from one user to another. Thus, we can conclude that the spatial directions of different users $\widehat{\mathbf{H}}_k$ are linearly independent from each other. Consequently, the rank of $\widetilde{\mathbf{W}}_k$ of the k^{th} user is $\widetilde{L}_k = \text{rank}(\widetilde{\mathbf{W}}_k) = N_R - N_k$, where N_R is the aggregate number of receive antennas (i.e., $N_R = \sum_{k=1}^K N_k$). Define the Singular Value Decomposition (SVD) of $\widetilde{\mathbf{W}}_k$ as

$$\widetilde{\mathbf{W}}_k = \mathbf{U}_k \mathbf{\Sigma_k} \left[\mathbf{V_k^{(1)}} \ \ \mathbf{V_k^{(0)}} \right]^{\mathrm{H}}, \tag{11}$$

where $\mathbf{V}_k^{(1)}$ holds the first $\widetilde{L}_k$ right singular vectors, while $\mathbf{V}_k^{(0)}$ have the remaining $(M - \widetilde{L}_k)$ right singular vectors. Hence, $\mathbf{V}_k^{(0)}$ forms an orthonormal basis for the null space of $\widetilde{\mathbf{W}}_k$, and therefore, its columns are candidates for the columns of the k^{th} user precoding matrix, $\widehat{\mathbf{F}}_k$.

The effective channel of the k^{th} user is the product $\widehat{\mathbf{H}}_k\mathbf{V}_k^{(0)}$. Due to nulling the interference of other users, this is now equivalent to the single-user MIMO capacity maximization problem and the best precoder is, thus, the right singular vectors of that effective channel [13]. Let $\bar{L}_k$ be the rank of the product $\widehat{\mathbf{H}}_k\mathbf{V}_k^{(0)}$ and it is upper bounded by $\bar{L}_k \leq min\{L_k, \widetilde{L}_k\}$, where L_k is the rank of $\widehat{\mathbf{H}}_k$. Thus, the SVD of the effective channel of the k^{th} user is given by:

$$\widehat{\mathbf{H}}_k\mathbf{V}_k^{(0)} = \mathbf{Q}_k \begin{bmatrix} \mathbf{\Lambda_k} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{bmatrix} \left[\mathbf{R}_k^{(1)} \ \ \mathbf{R}_k^{(0)} \right]^{\mathrm{H}}, \tag{12}$$

where $\mathbf{\Lambda_k}$ is $\bar{L}_k \times \bar{L}_k$ and the columns of $\mathbf{R}_k^{(1)}$ are the first $\bar{L}_k$ singular vectors. Finally, the product $\mathbf{V}_k^{(0)}\mathbf{R}_k^{(1)}$ forms an orthonormal basis of dimension $\bar{L}_k$, and it represents the precoding matrix that maximizes the capacity of the k^{th} user while achieving zero interference.

$$\widehat{\mathbf{F}}_k = \mathbf{V}_k^{(0)}\mathbf{R}_k^{(1)}. \tag{13}$$

2) **Case II: $N_k > m_k$:** In this case, it is assumed that the number of antennas of the k^{th} user, N_k, is larger than the number of complex data symbols, m_k, to be simultaneously transmitted to that user. Adding more antennas at each receiver enhances the diversity gain at each user. In addition, having more antennas at the users than the number of data streams means that we only feed back a smaller subspace of the right singular vectors of the channel matrix $\mathbf{H}_k \in \mathbb{C}^{N_k \times M}$ of user

k. This, in turn, enhances the capacity of the system. Let the SVD of the channel matrix $\mathbf{H}_k$ of the k^{th} user be:

$$\mathbf{H}_k = \mathbf{U}_k \Sigma_k \mathbf{V}_k^{\text{H}}, \tag{14}$$

where $\mathbf{U}_k \in \mathbb{C}^{N_k \times N_k}$ and $\mathbf{V}_k \in \mathbb{C}^{M \times M}$ are unitary matrices, and $\Sigma_k \in \mathbb{C}^{N_k \times M}$ is a rectangular matrix that has the singular values on its diagonal. Let $\mathbf{V}_{k,m_k}$ be a matrix that contains the first m_k columns of $\mathbf{V}_k$. From Eq. (14), we can notice that each row of $\mathbf{H}_k$ is a linear combination of the complex conjugate of the first N_k columns of $\mathbf{V}_k$. Thus, the subspace spanned by the first N_k columns of $\mathbf{V}_k$ is equivalent to the subspace spanned by the complex conjugate of the rows of the channel matrix $\mathbf{H}_k \in \mathbb{C}^{N_k \times M}$. Consequently, we can conclude that the subspace spanned by the first N_k columns of $\mathbf{V}_k$ always lies in the subspace spanned by the rows of $\mathbf{A}_k^* \in \mathbb{C}^{P_k \times M}$. This is important since $\mathbf{A}_k$ is assumed to be already known at the BS. Then, we can use a low dimensional codebook, to be designed in Sec. IV, in order to quantize $\mathbf{V}_{k,m_k}$. It was proved in [14] that the columns of $\mathbf{V}_{k,m_k}$ are isotropically distributed on the subspace they lie in. Hence, a Grassmannian packing based codebook, to be presented in Sec. IV-B, can be used to quantize $\mathbf{V}_{k,m_k}$. Let the quantized version of $\mathbf{V}_{k,m_k}$ be $\widehat{\mathbf{V}}_{k,m_k} \in \mathbb{C}^{M \times m_k}$, and it is chosen from the codebook $\mathcal{C}$ according to Eq. (7).

Now, following the conventional BD procedure, let $\mathbf{S}_k \in \mathbb{C}^{M \times (M - \sum_{i=1, i\neq k}^{K} m_i)}$ represent the orthonormal basis of the null space of $\mathbf{W}_k$, where

$$\mathbf{W}_k = \left[\widehat{\mathbf{V}}_{1,m_1} \cdots \widehat{\mathbf{V}}_{k-1,m_{k-1}} \widehat{\mathbf{V}}_{k+1,m_{k+1}} \cdots \widehat{\mathbf{V}}_{K,m_K}\right]^{\text{H}}. \tag{15}$$

The effective channel of the k^{th} user will be the product $\widehat{\mathbf{V}}_{k,m_k}^{\text{H}} \mathbf{S}_k$. The SVD of this product is given by:

$$\widehat{\mathbf{V}}_{k,m_k}^{\text{H}} \mathbf{S}_k = \mathbf{Q}_k \mathbf{\Lambda_k} \left[\mathbf{R_k^{(1)}} \quad \mathbf{R_k^{(0)}}\right]^{\text{H}}, \tag{16}$$

where $\mathbf{R}_k^{(1)}$ represents the first m_k right singular vectors. Finally, the product $\mathbf{S}_k \mathbf{R}_k^{(1)}$ form an orthonormal basis of dimension m_k, and it represents the precoding matrix $\widehat{\mathbf{F}}_k \in \mathbb{C}^{M \times m_k}$ that maximizes the capacity of the k^{th} user while achieving zero interference. The precoding matrix $\widehat{\mathbf{F}}_k$ is given by

$$\widehat{\mathbf{F}}_k = \mathbf{S}_k \mathbf{R}_k^{(1)}. \tag{17}$$

Hence, the received vector $\mathbf{y}_k \in \mathbb{C}^{N_k \times 1}$ at user k becomes

$$\mathbf{y}_k = \mathbf{H}_k \mathbf{F}_k \mathbf{u}_k + \sum_{j=1, j\neq k}^{K} \mathbf{H}_k \widehat{\mathbf{F}}_j \mathbf{u}_j + \mathbf{n}_k. \tag{18}$$

The received vector $\mathbf{y}_k$, in Eq. (18), is finally left multiplied by $\mathbf{U}_{k,m_k}^{\text{H}}$, where $\mathbf{U}_{k,m_k} \in \mathbb{C}^{N_k \times m_k}$ is the matrix that contains the first m_k columns of the matrix $\mathbf{U}_k$ given in Eq. (14).

B. The Per-User Rate

The BS can perform downlink precoding on the data vectors $\mathbf{u}_k \in \mathbb{C}^{m_k \times 1}$ intended for each user based on the fed back quantized spatial directions represented by $\widehat{\mathbf{H}}_k$. As described above, we consider the BD based linear precoding at the BS to obtain the beamforming matrices for each user $\mathbf{F}_k$. The BD strategy involves linear precoding that eliminates the interference at each user due to all other users as discussed in Sec. III-A. Hence, the second term in Eq. (6), which represents the interference at the k^{th} user due to all other users, is canceled in the case of perfect CSI at the BS (i.e., $\widehat{\mathbf{H}}_k \equiv \mathbf{H}_k$). Then, the per-user ergodic rate for case I is given by [6]:

$$R_{\text{CSIT,I}}(\gamma) = \mathbb{E} \log_2 \left| \mathbf{I}_{m_k} + \frac{\gamma}{K m_k} \mathbf{H}_k \mathbf{F}_k \mathbf{F}_k^{\text{H}} \mathbf{H}_k^{\text{H}} \right|. \tag{19}$$

For case II, the total effective channel after left multiplying Eq. (18) by $\mathbf{U}_{k,m_k}^{\text{H}}$ is $\mathbf{U}_{k,m_k}^{\text{H}} \mathbf{H}_k \mathbf{F}_k$. Hence the per-user ergodic rate for case II is given by:

$$R_{\text{CSIT,II}}(\gamma) = \mathbb{E} \log_2 \left| \mathbf{I}_{m_k} + \frac{\gamma}{K m_k} \mathbf{U}_{k,m_k}^{\text{H}} \mathbf{H}_k \mathbf{F}_k \mathbf{F}_k^{\text{H}} \mathbf{H}_k^{\text{H}} \mathbf{U}_{k,m_k} \right|, \tag{20}$$

where k is the user index, and a uniform power allocation policy is adopted. The expectation is evaluated over the distribution of the channel matrix, $\mathbf{H}_k$.

In the case of limited feedback of B bits for each user, the interference at the k^{th} user due to all other users cannot be completely eliminated because the quantized spatial direction spanned by the rows of $\widehat{\mathbf{H}}_k$ is not exactly the same as the original spatial direction spanned by the rows of $\widetilde{\mathbf{H}}_k$. As a result, this quantization leads to residual interference power, and the per-user rate for case I is given by [12]:

$$\begin{aligned} R_{\text{QUANT,I}}(\gamma) \; = \; & \mathbb{E} \log_2 \left| \mathbf{I}_{m_k} + \frac{\gamma}{K m_k} \sum_{j=1}^{K} \mathbf{H}_k \widehat{\mathbf{F}}_j \widehat{\mathbf{F}}_j^{\text{H}} \mathbf{H}_k^{\text{H}} \right| - \\ & \mathbb{E} \log_2 \left| \mathbf{I}_{m_k} + \frac{\gamma}{K m_k} \sum_{\substack{j=1 \\ j\neq k}}^{K} \mathbf{H}_k \widehat{\mathbf{F}}_j \widehat{\mathbf{F}}_j^{\text{H}} \mathbf{H}_k^{\text{H}} \right|. \end{aligned} \tag{21}$$

Similarly, the per-user rate for case II due to quantization is given by:

$$\begin{aligned} R_{\text{QUANT,II}}(\gamma) \; = \; & \mathbb{E} \log_2 \left| \mathbf{I}_{m_k} + \frac{\gamma}{K m_k} \sum_{j=1}^{K} \mathbf{U}_{k,m_k}^{\text{H}} \mathbf{H}_k \widehat{\mathbf{F}}_j \widehat{\mathbf{F}}_j^{\text{H}} \mathbf{H}_k^{\text{H}} \mathbf{U}_{k,m_k} \right| - \\ & \mathbb{E} \log_2 \left| \mathbf{I}_{m_k} + \frac{\gamma}{K m_k} \sum_{\substack{j=1 \\ j\neq k}}^{K} \mathbf{U}_{k,m_k}^{\text{H}} \mathbf{H}_k \widehat{\mathbf{F}}_j \widehat{\mathbf{F}}_j^{\text{H}} \mathbf{H}_k^{\text{H}} \mathbf{U}_{k,m_k} \right|, \end{aligned} \tag{22}$$

where k is the user index, and the expectation is evaluated over the distribution of the channel matrices, $\mathbf{H}_k \; \forall k \in \{1, 2, \cdots, K\}$, and the corresponding quantized precoding matrices, $\widehat{\mathbf{F}}_j$. The term $\mathbf{H}_k \widehat{\mathbf{F}}_k \widehat{\mathbf{F}}_k^{\text{H}} \mathbf{H}_k^{\text{H}}$ represents the useful signal intended for user k and, $\sum_{j=1, j\neq k}^{K} \mathbf{H}_k \widehat{\mathbf{F}}_j \widehat{\mathbf{F}}_j^{\text{H}} \mathbf{H}_k^{\text{H}}$ represents the muti-user interference at user k.

IV. AoD-adaptive Subspace Codebook

The path angles of departure of the k^{th} user, $\theta_{k,i}$s, defined in Eq. (1) depend on the obstacles that surround the BS. These obstacles are expected to change their physical positions in a much longer time than the channel coherence time. On the

other hand, for the path gains represented by $\mathbf{G}_k$ in Eq. (1), one resolvable path is formed by a set of scatters around user k, which consists of a number of unresolvable paths. Hence, path gains, $\mathbf{G}_k$'s, are expected to change much faster than path AoDs, $\theta_{k,i}$'s [5], [15]. However, the size of $\mathbf{G}_k$ is very low compared to the original channel matrix $\mathbf{H}_k$ and here comes the reuction in feedback overhead. During the angle coherence time, the spatial direction of the k^{th} user $\widetilde{\mathbf{H}}_k$ is isotropically distributed in the channel subspace, which is spanned by the rows of $\mathbf{A}_k(\theta_{k,1}, \theta_{k,2}, \cdots, \theta_{k,P_k})$. As shown in Eq. (1), each row of the channel matrix $\mathbf{H}_k$ is composed of P_k paths, where $\mathbf{A}_k(\theta_{k,1}, \theta_{k,2}, \cdots, \theta_{k,P_k})$ is completely determined by the path AoDs. The reason for the uniform distribution of $\widetilde{\mathbf{H}}_k$ in the row space of $\mathbf{A}_k$ is that the rows of $\mathbf{A}_k$ (steering vectors) are asymptotically orthogonal to each other (i.e., $\mathbf{A}_k\mathbf{A}_k^{\text{H}} \approx M\mathbf{I}_{P_k}$) [5]. Additionally, the path gains in $\mathbf{G}_k$ are modeled as i.i.d. circularly symmetric complex Gaussian random variables with zero mean and unit variance, which causes the user's spatial direction to be uniformly distributed in its channel subspace during the angle coherence time.

Due to limited scattering of mmWave, the number of paths P_k is much smaller than the number of transmit antennas M at the BS [16]. Therefore, the row space of $\mathbf{A}_k$ is only a subspace of the full M-dimensional space. Thus, assuming that the BS knows the AoDs, we can only quantize and feed back the path gains matrix $\mathbf{G}_k \in \mathbb{C}^{N_k \times P_k}$. Then, the quantization subspace $\mathbf{C}_{k,i}$ of the proposed AoD-adaptive subspace codebook $\mathcal{C} = \{\mathbf{C}_{k,1}, \mathbf{C}_{k,2}, \cdots, \mathbf{C}_{k,2^B}\}$ is formed by:

$$\mathbf{C}_{k,i} = \frac{1}{\sqrt{M}}\mathbf{X}_i\mathbf{A}_k, \tag{23}$$

where $\mathbf{X}_i \in \mathbb{C}^{m_k \times P_k}$ is a matrix whose rows are orthonormal, and its row space is isotropically distributed over the complex P_k-dimensional space.

A. Random Subspace Quantization codebooks

In general, the design of optimal quantization codebooks is a very hard problem, especially when the number of subspaces to be separated is large. Hence, the performance in such cases can be studied by averaging over random codebooks [17]. It is easier to analyze the performance of random codes in this case, and this would provide us with some performance bounds for structured codes. In our problem, a number of 2^B subspaces, each having a dimension of m_k, are picked at random in a P_k-dimensional Euclidean space. The set of all m_k-dimensional subspaces in a P_k-dimensional space represent a Grassmannian manifold, which is denoted by $\mathcal{G}_{P_k,m_k}$. The 2^B random subspaces, that form the random quantization codebook, are uniformly distributed over $\mathcal{G}_{P_k,m_k}$. A random subspace chosen uniformly over $\mathcal{G}_{P_k,m_k}$ can be generated by generating an $m_k \times P_k$ matrix whose elements are i.i.d. complex Gaussian. Then, an orthonormal basis for the row space of this matrix is calculated using QR decomposition.

B. Grassmannian Subspace Packing

The design of the quantized path gain matrices $\mathbf{X}_i$ is done using subspace packing in Grassmannian manifold. The packing problem tends to find 2^B subspaces in a higher dimensional space such that the minimum distance between two subspaces is maximized. There are many distance metrics that have been used for packing subspaces in the Grassmannian manifold. In this paper, we adopt the chordal distance, defined in Eq. (9), as our distance metric. The codebook design is done by solving the packing problem of 2^B m_k-dimensional subspaces in a complex Euclidean space of dimensionality P_k. We follow the iterative algorithm stated in [11] in order to solve the subspace packing problem. The solution of this problem is usually simpler when the number of subspaces in the codebook 2^B is lower than P_k^2. In that case, the minimum distance between two subspaces in the codebook can reach the Rankin bound [11], which is the maximum attainable theoretical distance.

V. Throughput analysis

In this section, we calculate the rate gap between the ideal rate and the rate using a random subspace quantization scheme. Due to space limitations, we only study the rate gap for case I assuming that all users have the same number of receive antennas (i.e., $N_k = m_k = N$) and same number of resolvable paths (i.e., $P_k = P$). We derive an expression for the required number of feedback bits to achieve some constant rate gap, where we prove that the number of bits scales linearly with the transmit power γ_{dB} in dB.

A. Rate Gap

The per-user rate of the ideal case of case I is given by Eq. (19), and the per-user rate of the practical case of case I is given by Eq. (21). Following Theorem 1 of [12], which gives an upper bound for the rate gap in Multi-User MIMO systems, we derive an expression for the per-user rate gap due to limited feedback in our massive MIMO system model. The per-user rate gap $\Delta R(\gamma) = R_{CSIT}(\gamma) - R_{QUANT}(\gamma)$ can be upper bounded as (details of the proof are omitted due to space limitations):

$$\Delta R(\gamma) \leq \log_2\left(1 + \frac{\gamma(K-1)M}{KN(KP-N)}D\right), \tag{24}$$

where D is the average subspace quantization error which is given by:

$$D = \mathbb{E}\left[d^2(\widetilde{H}_k, \widehat{H}_k)\right], \tag{25}$$

and $d(\widetilde{H}_k, \widehat{H}_k)$ is the chordal distance defined in Eq. (9).

B. Quantization Error

In this subsection, we calculate the quantization error, D, of the spatial direction of user k when the AoD-adaptive subspace codebook is used. We have $\mathbf{C}_{k,Z_k} = \frac{1}{\sqrt{M}}\mathbf{X}_{Z_k}\mathbf{A}_k$ and $\widetilde{\mathbf{H}}_k = \frac{1}{\sqrt{M}}\widetilde{\mathbf{G}}_k\mathbf{A}_k$; then, the quantization error is given by

$$D = \mathbb{E}\left[N - \left\|\widetilde{\mathbf{H}}_k\mathbf{C}_{k,Z_k}^{\text{H}}\right\|_{\text{F}}^2\right] = \mathbb{E}\left[N - \left\|\frac{1}{M}\widetilde{\mathbf{G}}_k\mathbf{A}_k\mathbf{A}_k^{\text{H}}\mathbf{X}_{Z_k}^{\text{H}}\right\|_{\text{F}}^2\right] \tag{26}$$

$$\overset{(a)}{\approx} \mathbb{E}\left[N - \left\|\widetilde{\mathbf{G}}_k\mathbf{X}_{Z_k}^{\text{H}}\right\|_{\text{F}}^2\right] \tag{27}$$

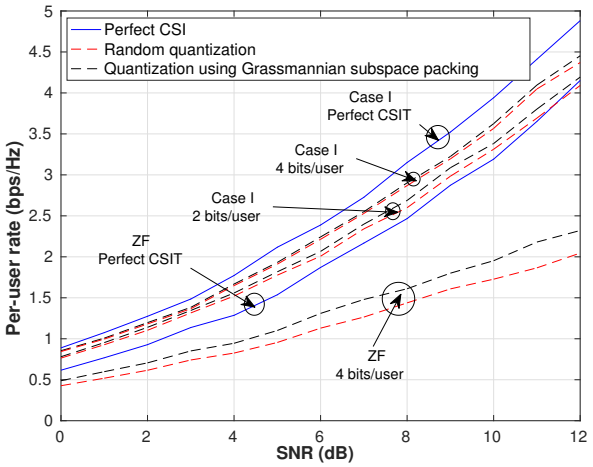

Fig. 1: BD vs conventional ZF: case I with $N_k = m_k = 2$

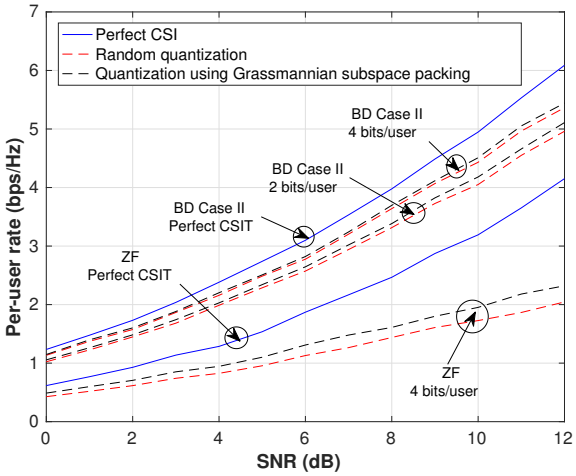

Fig. 2: BD vs conventional ZF: case II with $N_k = 3$, $m_k = 2$

where $\widetilde{\mathbf{G}}_k \in \mathbb{C}^{N\times P}$ is a matrix whose rows are orthonormal and its row space represents the subspace spanned by the rows of $\widetilde{\mathbf{G}}_k$. Step (a) is true due to $\mathbf{A}_k\mathbf{A}_k^{\mathrm{H}} \approx M\mathbf{I}_P$. Both $\widetilde{\mathbf{G}}_k$ and $\mathbf{X}_{Z_k}$ are isotropically distributed subspaces on the P-dimensional space. Then, we can bound the quantization error as [12]:

$$D \leq \bar{D} = \frac{\Gamma(\frac{1}{T})}{T}(C_{PN})^{-\frac{1}{T}}2^{-\frac{B}{T}}, \tag{28}$$

where $T = N(P-N)$ and $C_{PN} = \frac{1}{T!}\prod_{i=1}^{N}\frac{(P-i)!}{(N-i)!}$.

C. Feedback Bits

Now, we discuss the required number of feedback bits B that results in a constant rate gap. After bounding the quantization error by $\bar{D}$, the rate loss can be bounded as:

$$\Delta R(\gamma) \leq \log_2\left(1 + \frac{\gamma(K-1)M}{KN(KP-N)}\bar{D}\right). \tag{29}$$

Let the rate gap be such that $\Delta R(\gamma) \leq \log_2(b)$ bps/Hz, and substituting for $\bar{D}$ from Eq. (28), then the number of feedback bits that guarantees this rate loss is given by:

$$B = 3.3\,T\log_{10}(\gamma) - T\log_2\left[(b^{\frac{1}{N}}-1)\frac{KN(KP-N)}{(K-1)M}\right] + T\log_2\left(\frac{\Gamma(\frac{1}{T})}{T}\right) - \log_2(C_{PN}), \tag{30}$$

where B scales linearly with the transmit power γ_{dB} in dB.

VI. Simulation Results

In this section, the performance of the proposed feedback system and codebook design is examined and verified. The system parameters are set as follows. The number of antennas at the BS is $M = 128$, the number of users in the system is $K = 8$, the number of antennas at each user is $N_k = N = 2$, the number of data streams transmitted simultaneously to each user is $m_k = 2$ and the number of resolvable paths is $P = 3$. The path AoDs of the users are independent and uniformly distributed in $\left[-\frac{1}{2}\pi, \frac{1}{2}\pi\right]$.

Fig. 1 compares the performance of BD and the conventional ZF scheme for case I with $N_k = m_k = 2$. Fig. 1 also compares the performance of the ideal case, where perfect CSI is assumed available at the BS, and the limited feedback case where quantized CSI is fed back to the BS with $B = 2$ and 4 per user. Note that in the case of conventional ZF scheme, the channel vector of each antenna at the k^{th} user is separately quantized and fed back to the BS; therefore, the feedback bits for each user are divided among its receive antennas in this case. This is because in the case of conventional ZF, any user antenna is used to receive a single stream and all other streams must be nulled (even other streams intended for the same user), which is not the case for BD. In Fig. Fig. 1, we plot the per-user rate using the AoD-adaptive codebook with both random subspace quantization and using Grassmannian subspace packing based coodbook. From this figure, we can easily see the performance gains of the BD approach as compared to the conventional ZF approach. In addition, it can be noticed that Grassmannian codes are always better (or slightly better) than random codes. Note that Grassmannian codes are more structured, which deem them suitable for practical implementation, while random codes are impractical. Finally, it is clear that increasing the number of feedback bits enhances the system performance, and we can get arbitrary close to the performance of the ideal system with perfect CSI at the BS.

Fig. 2 compares the performance of BD with ideal and quantized CSI against the ideal and quantized CSI of the conventional ZF scheme for case II with $N_k = 3, m_k = 2$ with $B = 2$ and 4 per user. The same observations mentioned above while commenting on the results of Fig. Fig. 1 apply in this case as well. Moreover, it is noticeable that case II has higher per-user rate than case I for the same number of user streams. This is due to the fact that in case II we assume more receiving antennas at each user than the number of streams, which introduces diversity gain at the users.

In Fig. 3, we present numerical results for the practical per-user rate when using random quantization codebook. The

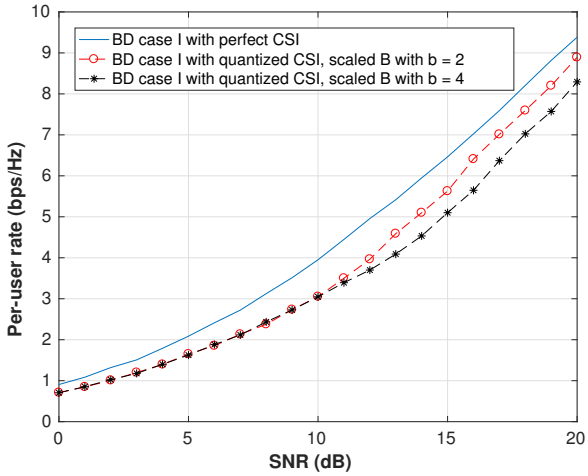

Fig. 3: Ideal vs quantized CSI for case I with B as in (30)

required number of feedback bits is scaled as per Eq. (30) in order to guarantee a maximum rate gap of $\log_2(b)$, where we show the results for $b = 2$ and 4. We notice in Fig. 3 that the rate gap between the ideal (perfect CSI at the BS) and practical cases does not increase as the SNR increases; this is due to scaling the number of feedback bits B with the transmitted power γ_{dB} as explained above. It is clear that the rate gap at any SNR does not exceed the maximum value of $\log_2(b)$, which validates the expression in (30).

VII. Conclusions

In this paper, we have considered the problem of channel feedback in FDD massive MIMO systems with multiple antennas at the users. We have considered the use of BD at the massive basestation. Based on the nature of our channel model, we have devised a channel feedback scheme to reduce the required feedback bits. We have quantified the rate loss due to the use of channel feedback (compared to the case with perfect CSI at the BS). Finally, we have proposed a systematic approach to design the channel feedback codebooks in which the codebook design is formulated as a subspace packing problem over the Grassmannian manifold.

References

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

[2] Z. Gao, L. Dai, Z. Wang, and S. Chen, "Spatially common sparsity based adaptive channel estimation and feedback for fdd massive mimo," *IEEE Transactions on Signal Processing*, vol. 63, no. 23, pp. 6169–6183, Dec 2015.

[3] Y. G. Lim and C. B. Chae, "Compressed channel feedback for correlated massive mimo systems," in *IEEE International Conference on Communications Workshops (ICC)*, June 2014.

[4] A. Ge, T. Zhang, Z. Hu, and Z. Zeng, "Principal component analysis based limited feedback scheme for massive mimo systems," in *IEEE International Symposium on Personal, Indoor, and Mobile Radio Communications (PIMRC)*, Aug 2015.

[5] W. Shen, L. Dai, G. Gui, Z. Wang, R. W. Heath, and F. Adachi, "Aod-adaptive subspace codebook for channel feedback in fdd massive mimo systems," in *IEEE International Conference on Communications (ICC)*, May 2017.

[6] Q. H. Spencer, A. L. Swindlehurst, and M. Haardt, "Zero-forcing methods for downlink spatial multiplexing in multiuser mimo channels," *IEEE Transactions on Signal Processing*, vol. 52, no. 2, pp. 461–471, Feb 2004.

[7] F. Rusek, D. Persson, B. K. Lau, E. G. Larsson, T. L. Marzetta, O. Edfors, and F. Tufvesson, "Scaling up mimo: Opportunities and challenges with very large arrays," *IEEE Signal Processing Magazine*, vol. 30, no. 1, pp. 40–60, Jan 2013.

[8] R. Schmidt, "Multiple emitter location and signal parameter estimation," *IEEE Transactions on Antennas and Propagation*, vol. 34, no. 3, pp. 276–280, Mar 1986.

[9] H. Xie, F. Gao, and S. Jin, "An overview of low-rank channel estimation for massive mimo systems," *IEEE Access*, vol. 4, pp. 7313–7321, 2016.

[10] P. Zhao, Z. Wang, and C. Sun, "Angular domain pilot design and channel estimation for fdd massive mimo networks," in *IEEE International Conference on Communications (ICC)*, May 2017.

[11] I. S. Dhillon, J. R. W. Heath, T. Strohmer, and J. A. Tropp, "Constructing packings in grassmannian manifolds via alternating projection," *Experimental Mathematics*, vol. 17, no. 1, pp. 9–35, 2008.

[12] N. Ravindran and N. Jindal, "Limited feedback-based block diagonalization for the mimo broadcast channel," *IEEE Journal on Selected Areas in Communications*, vol. 26, no. 8, pp. 1473–1482, October 2008.

[13] G. G. Raleigh and J. M. Cioffi, "Spatio-temporal coding for wireless communication," *IEEE Transactions on Communications*, vol. 46, no. 3, pp. 357–366, Mar 1998.

[14] D. J. Love and R. W. Heath, "Limited feedback unitary precoding for spatial multiplexing systems," *IEEE Transactions on Information Theory*, vol. 51, no. 8, pp. 2967–2976, Aug 2005.

[15] V. Va, J. Choi, and R. W. Heath, "The impact of beamwidth on temporal channel variation in vehicular channels and its implications," *IEEE Transactions on Vehicular Technology*, vol. 66, no. 6, pp. 5014–5029, June 2017.

[16] T. S. Rappaport, S. Sun, R. Mayzus, H. Zhao, Y. Azar, K. Wang, G. N. Wong, J. K. Schulz, M. Samimi, and F. Gutierrez, "Millimeter wave mobile communications for 5g cellular: It will work!" *IEEE Access*, vol. 1, pp. 335–349, 2013.

[17] W. Santipach and M. L. Honig, "Asymptotic capacity of beamforming with limited feedback," in *International Symposium onInformation Theory (ISIT)*, June 2004.