Fundamental Limits in Correlated Fading MIMO Broadcast Channels: Benefits of Transmit Correlation Diversity

Junyoung Nam

Abstract

We investigate the impact of transmit correlation on the capacity of correlated fading MIMO broadcast channels (BCs) and establish capacity characterizations in various regimes of system parameters, with a particular interest in the large-scale array (or massive MIMO) regime. It is advocated in this paper that transmit correlation can be of use to increase both multiplexing gain and power gain in multiuser MIMO systems. We show the potential gains and succinctly characterize fundamental limits in correlated fading MIMO BCs, assuming an ideal condition for which transmit correlation diversity is well defined. In particular, transmit correlation is shown to improve the system multiplexing gain up to by a factor of the degrees of transmit correlation diversity. Not relying on the ideal condition on transmit correlations, we further propose a joint spatial division and multiplexing (JSDM) scheme based on opportunistic beamforming that, for a large number of users, can achieve near optimal sum-rate scaling with very limited channel state feedback in realistic channels.

Index Terms

Multiuser MIMO, transmit antenna correlation, large-scale (massive) MIMO.

I. INTRODUCTION

Channel fading effect was once considered as a harmful source to combat with transmit or receive diversity, but since [1], [2] independent fading in multiple-antenna channels has been the useful fountain for increasing the degrees of freedom available for wireless communications. The independent channel randomness across transmit and receive antenna pairs may be significantly impaired by antenna correlation typical in realistic MIMO wireless communications. Transmit (or receive) antenna correlation occurs in the presence of a correlation between transmit (or receive) antennas, turning independent fading into correlated fading.

The fundamental limits and optimal precoding of spatially correlated fading MIMO channels have been well characterized in a variety of transmit correlation models (e.g., see [3]–[6]). In this point-to-point case, transmit correlation is well known to play a detrimental role to the capacity (e.g., as a power offset at high signal-to-noise ratio (SNR) [7]). On the contrary, the fundamental characteristics of correlated fading MIMO BC have not been fully understood in general models on transmit correlation matrices $R_i$ of users $i = 1, \ldots, K$. In particular, the impact of transmit correlation on the capacity of MIMO BC has not received much attention and it has not been well characterized yet. For example, the work of [8] extended the sum-rate scaling result of [9] to the special case where all users have the same transmit correlation such that $R_i = R, \forall i$, which concludes that transmit correlation has a fairly detrimental impact on the sum capacity of MIMO BC, compared with the independent fading case. This coincides with the traditional view on the impact of transmit correlation on capacity. Like this, transmit correlation has been considered as a harmful source for MIMO communications.

A different line of thought is that transmit correlation can be in fact advantageous for multiuser MIMO (MU-MIMO) communications, since there exist diverse transmit correlations across multiple users to leverage. Basically, different transmit correlations imply different large-scale (or long-term)
channel directions of users so that the diversity of transmit correlations can be favorably exploited in the multiuser context. For such effect, we coin the term transmit correlation diversity. In the more general transmit correlation model where users have different $R_{ii}$, there is a number of literature on the results that take advantage of transmit correlation and address its potential benefits in several aspects. The benefits of transmit correlation have been observed with respect to channel state information (CSI) feedback overhead, scheduling, and codebook design (e.g., see [10]–[12]) in practice. A beneficial impact of transmit correlation on the sum capacity was also addressed in a simple and very special channel model [13], [14]. Even if these results might shed light on potential gains of transmit correlation, there is no comprehensive understanding on the gains and fundamental characterization for correlated fading MIMO BCs. Consequently, it is indeed necessary to revisit the impact of transmit correlation and establish the capacity characterization in a more general channel model and in various regimes of system parameters of interest. This is a main goal of the present work.

A key fundamental limit on the sum rate of i.i.d. Rayleigh block-fading MIMO BC consisting of a transmitter with $M$ antennas and $K$ receivers (users) with a single antenna each immediately follows from the work of Zheng and Tse [15, Sec. V] (see also [16]) by taking downlink training into account and by allowing cooperation among users. Namely, the high-SNR capacity of the resulting pilot-aided systems is at best

$$M_0^* (1 - M_0^*/T_c) \log \text{SNR} + O(1)$$

where $T_c$ is the coherence time interval[$1$] and $M_0^* = \max \{ M, K, \lfloor T_c/2 \rfloor \}$. For typical cellular downlink systems with $M$ small, where $\min \{ M, K \} \ll T_c$, the factor $T_c/2$ does not significantly affect the system performance. However, in the large scale array regime with $M > T_c$, which has been paid great attention in practice since [17] and is also of particular interest in this work, this factor is shown to have a critical impact on the system performance. To be specific, no matter how large $M$ and $K$ are, multiplexing gain is fundamentally limited by $T_c/2$, which was also observed in [18].

Assuming the perfect CSI at the transmitter (CSIT) and hence not taking the training overhead into account, another well-known characteristic is the asymptotic behavior of the sum capacity of i.i.d. Rayleigh fading MIMO BC in the large number of users (large $K$) regime. Sharif and Hassibi [9] showed that the sum rate scales at best as $M \log \log K + M \log P_M + o(1)$, where $M$ is finite, $o(1)$ goes to zero as $K \to \infty$, and $M \log \log K$ is the multiuser diversity gain. Furthermore, the same scaling law was shown to be achievable by random beamforming (RBF), which requires only partial CSIT. On the other hand, the impact of transmit correlation on the high-SNR behavior of the capacity of correlated fading MIMO channels was well refined by adding a power offset term in [6], [7]. By letting all users fully cooperate again, the same behavior would be the upper bound on the sum rate of correlated fading MIMO BCs.

Probably, some common lessons based on the previous capacity results are as follows.

- For both $M$ and $K$ large, the coherence time $T_c$ plays as a serious limiting factor in the performance of MIMO wireless communications.
- The fundamental limits and asymptotic behaviors of MIMO BCs are upper-bounded by those of MIMO point-to-point channels, where transmit and receive antennas are co-located around a single point, respectively.
- Transmit correlation is detrimental in MIMO wireless networks.

Albeit the first two statements are indeed true in independent fading MIMO wireless communications and the third one is the same in a MIMO BC with common transmit correlation, we argue in the present work that this is not necessarily the case with correlated fading channels or with different transmit correlations among users.

In order to validate the above argument, we make use of the notion of JSDM, proposed by the authors in [19]. The idea is that diverse transmit correlations across users can be exploited to create minimally-interfering virtual sectors, not depending on instantaneous channel realizations. To this end, we partition

\footnote{The unit of coherence time interval can be represented as the number of transmit symbols because each training signal is transmitted over a symbol interval, which in turn corresponds to a channel use.}
users with similar transmit correlations into a group. Then, separate certain groups, whose eigenspaces are quasi-orthogonal each other, by beamforming across such eigenspaces and simultaneously serve users within each group by spatial multiplexing. The maximum number of virtual sectors is called the degrees of transmit correlation diversity available in a correlated fading MIMO BC.

In the first half of this work, we restrict our attention to an optimistic condition defined in [19] to intuitively expose a potential gain of transmit correlation diversity and to see what fundamental limits are in correlated fading channels. The ideal condition is called the tall unitary structure of transmit correlations among users for which users in a group have the same transmit correlation matrix of rank \( r \leq M \) and the eigenspaces of all groups are orthogonal to each other. In this case, the degrees of transmit correlation diversity is given by the constant \( G = \lfloor M/r \rfloor \). Assuming the ideal condition, we have the following main results.

- If \( r \leq \lceil T_c/2 \rceil \), the high-SNR capacity of pilot-aided MU-MIMO systems is upper-bounded by
  \[
  M^* \left( 1 - \frac{M^*}{T_c G} \right) \log \text{SNR} + O(1)
  \]
  where \( M^* = \min\{M, K\} \). This important result advocates that multiplexing gain can continue growing as \( M^* \) increases, provided that transmit correlations among users are sufficiently high and well structured.
- The term \( O(1) \) in (2) is further refined as a power gain term, which accounts for a potential gain due to eigenbeamforming across the long-term eigenmodes.
- For the perfect CSIT and large \( K \)
  \[
  M \log \log K + M \log \frac{P}{M} + \sum_{g=1}^{G} \log |\Lambda_g| + o(1)
  \]
  where \( \Lambda_g \) is a diagonal matrix consisting of the eigenvalues of transmit correlation matrix of group \( g \).

These results show that transmit correlation diversity can be leveraged to increase both multiplexing gain and power gain in correlated fading MIMO BCs. Also, fundamental limits for these channels are not upper-bounded any longer by those of the point-to-point case in practical pilot-aided systems, where CSIT is provided at the cost of downlink training. As a reference on how large the potential gain of transmit correlation diversity could be, we show the gap between the multiplexing gains of (1) and (2) in Fig. 1, where the maximum possible gain is shown to increase the multiplexing gain of pilot-aided MU-MIMO systems by a factor of \( G \). For \( M \) or \( K \) small, the gap is insignificant but it quickly becomes much larger as both \( M \) and \( K \) increase.

Although the tall unitary structure is essential to fully exploit transmit correlation diversity, it seems very unusual in realistic systems. On one hand, in order to address this issue in the case of uniform linear arrays (ULAs) with \( M \) large, the authors in [20] showed using Szego’s asymptotic results on Toeplitz matrices [21] that the ideal condition can be closely approximated, as long as the user groups have nonoverlapping supports of their angle of departure (AoD) distributions. Therefore, the ideal structure can be easily obtained in the large \( M \) regime. On the other hand, this ceases to hold for small \( M \).

The second part of this paper is thus devoted to show that we can get a fairly large portion of the potential gain for not-so-large \( M \), not relying on the ideal structure, with very limited CSIT. For any transmit correlations among users, we can achieve the asymptotic sum-rate scaling in (3) at the cost of some penalty factors. This result can be used to characterize the capacity of correlated fading MIMO BCs in general settings. To this end, we propose a new JSDM scheme using the idea of opportunistic beamforming [9], [22]. In sharp contrast to RBF that requires a large number of users, \( K \gg M \), the proposed scheme is shown to get a good system throughput even for not-so-large \( K \), compared to the zero-forcing beamforming with semi-orthogonal user selection (ZFBF-SUS) [23]. Finally, we provide
Fig. 1. Impact of transmit correlation diversity \((G)\) on multiplexing gain for different numbers of \(\min(M, K)\) in pilot-aided systems, where the solid lines indicate \(T_c = 32\) and the dash-dotted lines indicate \(T_c = 100\).

A more complete view on our fundamental limits and scheme by taking into account some practical considerations to realize a large fraction of the potential gain of transmit correlation diversity in practice.

**Notation:** \(A^H\), \(\|A\|_F\), and \(\lambda_i(A)\) denote the Hermitian transpose, the Frobenius norm, the \(i\)th eigenvalue (in descending order) of matrix \(A\). \(\text{tr}(A)\) and \(|A|\) denote the trace and the determinant of a square matrix \(A\). \(I_n\) denotes the \(n \times n\) identity matrix. \(\|a\|\) denotes the \(\ell^2\) norm of vector \(a\). \(|\mathcal{X}|\) is the cardinality of a discrete set \(\mathcal{X}\). We also use \(x \sim \mathcal{CN}(0, \Sigma)\) to indicates that \(x\) is a zero-mean complex circularly-symmetric Gaussian vector with covariance \(\Sigma\). Finally, \(C^{\text{sum}}(a, b)\) denotes the asymptotic sum capacity when the system parameters \(a\) and \(b\) are sufficiently large.

II. SYSTEM DESCRIPTION

A. Channel Model

In this paper, we consider a family of spatially correlated Rayleigh fading channels obeying the well-known Kronecker model [3], [4] (or separable correlation model)

\[
H = R_T^{1/2}W R_R^{1/2}
\]  

(4)

where the elements of \(W\) are i.i.d. \(\sim \mathcal{CN}(0, 1)\), and \(R_T\) and \(R_R\) denote the deterministic transmit and receive correlation matrices, respectively. The random matrix \(H\) follows the frequency-flat block-fading model for which it remains constant during the coherence time interval of \(T_c\) but changes independently every interval. Most of our results in this paper remain valid in the more general unitary-independent-unitary (UIU) model (for which see [6]), since the elements of \(W\) are allowed to be independent nonidentically distributed (IND) to apply well-known results of random matrix theory. Throughout this paper, correlated fading channels imply not only the i.i.d. Rayleigh fading model but also the UIU model. We also assume wide-sense stationarity of the channel. Within the observation time for channel estimation and data detection, the long-term channel properties do not change.
Consider a MIMO broadcast channel (downlink) with $M$ transmit antennas and $K$ users equipped with a single antenna each. Since users have no receive correlation in this case, we let $\mathbf{R} \triangleq \mathbf{R}_T$ for notational simplicity. In general, $\mathbf{R}$ is of full algebraic rank but with eigenvalues except dominant ones decaying quickly. We normalize $\text{tr}(\mathbf{R}) = M$ without loss of generality for all users and let $r = \text{rank}(\mathbf{R})$ and $\tilde{r} \leq r$ denote the number of dominant eigenvalues (in descending order) of $\mathbf{R}$, which are defined as follows: for any integer $\tilde{r} > 0$, there exists a positive constant $m$ such that
\[
\frac{\lambda_1(\mathbf{R})}{\lambda_{\tilde{r}}(\mathbf{R})} \leq m < \infty
\]
i.e., dominant eigenvalues are uniformly bounded.

By using the Karhunen-Loeve transform, the channel vector of a user can be expressed as
\[
\mathbf{h} = \mathbf{U} \Lambda_{\frac{1}{2}} \mathbf{w}
\]
where $\mathbf{w} \in \mathbb{C}^{r \times 1} \sim \mathcal{CN}(\mathbf{0}, \mathbf{I})$, $\Lambda$ is an $r \times r$ diagonal matrix whose elements are the non-zero eigenvalues of $\mathbf{R}$, and $\mathbf{U} \in \mathbb{C}^{M \times r}$ is a matrix whose columns are the eigenvectors of $\mathbf{R}$ corresponding to the non-zero eigenvalues, i.e., $\mathbf{R} = \mathbf{U} \Lambda \mathbf{U}^H$. We call $\mathbf{R}$ the large-scale fading component and $\mathbf{w}$ the small-scale fading component.

Let $\mathbf{H}$ denote the $M \times K$ system channel matrix given by stacking the $K$ users channel vectors by columns. The signal vector received by the users is given by
\[
\mathbf{y} = \mathbf{H}^H \mathbf{d} + \mathbf{z} = \mathbf{H}^H \mathbf{x} + \mathbf{z}
\]
where $\mathbf{V}$ is the $M \times s$ precoding matrix with $s$ the rank of the input covariance $\mathbf{\Sigma} = \mathbb{E}[\mathbf{x} \mathbf{x}^H]$ (i.e., the total number of independent data streams), $\mathbf{d}$ is the $s$-dimensional transmitted data symbol vector such that the transmit signal vector is given by $\mathbf{x} = \mathbf{V} \mathbf{d}$, and $\mathbf{z} \sim \mathcal{CN}(\mathbf{0}, \mathbf{I})$ is the Gaussian noise at the receivers. The system has the total power constraint such that $\text{tr}(\mathbf{\Sigma}) \leq P$, where $P$ implies the total transmit SNR.

### B. Review of Joint Spatial Division and Multiplexing (JSDM)

This MU-MIMO strategy was originally designed to provide a feasible solution for frequency division duplexing (FDD) large-scale MIMO systems, in which keeping the CSI feedback overhead realistic is a critical issue to handle. The goal of JSDM is to reduce the downlink training and the CSI feedback overheads by exploiting the fact that some users have similar transmit correlation matrices and further by leveraging a useful structure of users transmit correlations.

1) **User Partitioning:** In order to create the useful structure on transmit correlations, for a given user and scatterer geometry, we have to classify users based on their transmit correlation matrices. To this end, put together users with similar transmit correlations into a group. Then, separate multiple groups by spatial division, whose “long-term” subspaces are quasi-orthogonal. Spatial division here is done by multiple beamforming along the quasi-orthogonal subspaces. Simultaneously, we can serve users in each group by spatial multiplexing which depends on instantaneous channels. In general, we have multiple sets of quasi-orthogonal groups, which we call classes in this work. Different classes are served with orthogonal time/frequency resources. Therefore, we partition the entire user set, $\mathcal{K} = \{1, 2, \cdots, K\}$, into $T$ non-overlapping subsets (classes), $\mathcal{T}^{(1)}, \mathcal{T}^{(2)}, \cdots, \mathcal{T}^{(T)}$.

Let the integer $K_g^{(t)}$ denote the number of users for group $g$ in class $t$ (denoted by $(t, g)$) such that $\sum_t \sum_g K_g^{(t)} = K$ and also let $G_t$ denote the number of groups in class $t$. Then, the $t$th class can be written as the subset $\mathcal{T}^{(t)} = \{ \mathcal{T}_1^{(t)}, \mathcal{T}_2^{(t)}, \cdots, \mathcal{T}_{G_t}^{(t)} \}$, where $\mathcal{T}_g^{(t)} = \{ \mathbf{g}_1^{(t)}, \mathbf{g}_2^{(t)}, \cdots, \mathbf{g}_{K_g^{(t)}}^{(t)} \}$ with $\mathbf{g}_k^{(t)} = \sum_{i=1}^{t-1} \sum_{j=1}^{g-1} K_j^{(i)} + k$, for $t = 1, \cdots, T, \ g = 1, \cdots, G_t, \ k = 1, \cdots, K_g^{(t)}$. By the above set partitioning, user $i \in \mathcal{K}$ is reordered with the index $\mathbf{g}_k^{(t)}$. For notational simplicity, the superscript $(t)$ for classes

\[2\]Although we treat only single-antenna users for convenience, multiple-antenna case can easily apply.
will be omitted hereafter unless it is essential for clarity. In this case, the above user index reduces to \( g_k = \sum_{h=1}^{g-1} K_h + k \), where \( k = 1, \ldots, K_g \).

Let \( R_{g_k} \) denote the transmit correlation matrix of user \( g_k \) and be eigen-decomposed as \( R_{g_k} = U_{g_k} \Lambda_{g_k} U_{g_k}^H \), with rank of \( r_{g_k} \) and \( \tilde{r}_{g_k} \leq r_{g_k} \) dominant eigenvalues. Denote by \( \tilde{U}_{g_k} \) the \( M \times \tilde{r}_{g_k} \) matrix collecting the eigenvectors corresponding to dominant eigenvalues. Also denote by \( R_g \) the transmit correlation matrix of group \( g \) decomposed as \( R_g = U_g \Lambda_g U_g^H \) of rank \( r_g \) with \( \lambda_g \)th eigenvalue of \( R_g \). The subspace spanned by the eigenvectors of \( R_g \) is called the eigenspace of group \( g \). The channel vector of user \( g_k \) is

\[
h_{g_k} = U_{g_k} \Lambda_{g_k}^{1/2} w_{g_k}.
\]  

(8)

Also, let \( H_g = [h_{g_1}, \ldots, h_{g_{K_g}}] \) and \( H = [H_1, \ldots, H_G] \) denote the group channel and the system channel matrix, respectively, reordered according to the index \( g_k \).

2) Transformed Channel: The precoding of JSDM has a two-stage structure given by \( V = BP \), where \( B \in \mathbb{C}^{M \times b} \) with \( b \leq M \) is a pre-beamforming matrix that depends only on the channel second-order statistics, \( \tilde{U}_{g_k} \), which is supposed to be similar to \( U_g \) by user partitioning. Then \( P \in \mathbb{C}^{b \times s} \) is a precoding matrix that depends on the instantaneous realization of the aggregate transformed channel (or projected channel) \( H \), defined by \( H^T \triangleq H^T B \). Now, we have the following transmit signal vector \( x = BPd \). We divide \( b \) such that \( b = \sum_g b_g \), where \( b_g \) is an integer not smaller than \( s_g \) (the number of independent data streams for group \( g \)), and let \( B_g \) be the \( M \times b_g \) pre-beamforming matrix of group \( g \). The received signal \( (7) \) can be rewritten as

\[
y = H^T Pd + z.
\]

Although the transformed channel \( H \) has a reduced dimension, its estimation and feedback may still be unaffordable. By virtue of user partitioning, we can consider feeding back only the \( G \) diagonal blocks

\[
H_g \triangleq B_g^H H_g, \ g = 1, \ldots, G
\]

(9)

and each group is independently processed by treating signals of other groups as interference. Therefore, the transformed channel \( H_g \) can be viewed as an effective channel for group \( g \). In this case, the precoding matrix takes on the block-diagonal form \( P = \text{diag}(P_1, \ldots, P_G) \), where \( P_g \in \mathbb{C}^{b_g \times s_g} \), yielding the vector broadcast plus interference Gaussian channel

\[
y_g = H_g^T B_g P_g d_g + \sum_{h \neq g} H_g^T B_h P_h d_h + z_g
\]

(10)

for \( g = 1, \ldots, G \). Assuming that the inter-group interference term is made sufficiently small by appropriate user partitioning and pre-beamforming, each group can be viewed as a virtual sector.

The reader has to distinguish between JSDM and space-division multiple access (SDMA). The latter refers to a wide class of beamforming schemes depending on instantaneous channel orthogonality (spatial separation) between users (e.g., [9], [22]–[25]). In contrast, the first-stage spatial separation of JSDM depends on the orthogonality between long-term eigenspaces of groups, and the second-stage separation is done by instantaneous channel orthogonality between users within a group.

III. Transmit Correlation Diversity

In this section, we introduce the notion of transmit correlation diversity. Once partitioning the user set \( K \), we have a useful structure of transmit correlations to be exploited by the two-stage precoding of JSDM. The structure can be characterized by the two properties defined as follows.

- **Similarity:** For users in group \( g \), each user subspace spanned by columns of \( U_{g_k}^* \) is similar to the eigenspace of group \( g \). In other words, the chordal distance between \( U_{g_k}^* \) and \( U_g \) should be minimized. Strictly, the perfect similarity is defined by \( U_{g_k} = U_g \) for all \((g, k)\).
- **Orthogonality:** For groups in a class, their eigenspaces are quasi-orthogonal such that \( U_g^H U_h \approx 0 \), for all \( h \neq g \). The exact orthogonality implies \( U_g^H U_h = 0 \).
Clearly, the above long-term (large-scale) orthogonality ensures the instantaneous (small-scale) orthogonality given by $h_i^H h_j = 0$, $i \neq j$, but not vice versa.

Fig. 2 depicts a simple example which explains illustratively the virtual sectorization given by making use of diverse transmit correlations in a three-sector base station (BS).

1) In the beginning, angular regions (pie slices drawn by AoD and angular spread (AS) at the BS) representing eigenspaces (given by group transmit correlations) are overlapped, i.e., user groups are interfering with each other and there is no noticeable structure.
2) Put together the red angular regions into class $t = 1$ and separate them by multiple pre-beamforming along eigenspaces. By doing so, each group can be viewed as a virtual sector.
3) Do the same thing on the blue regions for class $t = 2$.
4) Multiple users within each group (i.e., virtual sector) can be simultaneously served by spatial multiplexing.

Although the notion of transmit correlation diversity is subtle to clearly define unless an ideal condition is given, the formal definition can be stated as follows.

**Definition 1** (Transmit Correlation Diversity). A multi-user MIMO channel is said to have the $G$ degrees of transmit correlation diversity, if its classes after user partitioning have, on the average, $G$ groups whose eigenspaces are quasi-orthogonal.

With $G = 1$ representing rich scattering channels, $G$ can grow up to $M$ as AS of all users goes to zero. Roughly speaking, the number of degrees of transmit correlation diversity depends on the root-mean-square AS of users. The quasi-orthogonality here will be more or less clarified by introducing a “semi-unitary structure” in Sec. VI.

In order to elucidate a potential gain of transmit correlation diversity, we now define an ideal structure of transmit correlations, already introduced in [19], that strictly satisfies the two properties, similarity and

---

3 AoD and angle of arrival (AoA) are generally different in FDD. As AoD is more precise at the transmitter side, we prefer the terminology AoD to AoA.

4 More precisely, $G$ here corresponds to the largest integer smaller than or equal to the average number of groups over classes.
orthogonality. In this special case, we assume for notational simplicity that $G$ groups are formed in a symmetric manner such that $r_g = r$ for all $g$.

**Definition 2 (Tall Unitary Structure).** For $G \leq \lfloor \frac{M}{r} \rfloor$, a tall unitary structure of transmit correlations is obtained if $U_{g_k} = U_g$ for all $(g,k)$ and if the $M \times rG$ matrix $U = [U_1, \cdots, U_G]$ is tall unitary such that

$$U^H U = I.$$  \hspace{1cm} (11)

For $M = rG$, we call this the unitary structure since $U$ is now unitary. Under the tall unitary structure, we just let $b_g = r_g$ and $B = U$. Then, the tall unitary structure is translated into the orthogonality of $H_g^H B_h = 0$, for all $h \neq g$. These choices let the MU-MIMO channels in (10) take on the form

$$y_g = H_g^H P_g d_g + z_g = W_g^H \Lambda_g^{1/2} P_g d_g + z_g$$ \hspace{1cm} (12)

where $W_g$ is an $r_g \times K_g$ i.i.d. matrix with elements $\sim \mathcal{C}\mathcal{N}(0,1)$, for all $g$. Using this fact, we arrive at the following simple yet important result.

**Theorem 1 ([19]).** Under the tall unitary structure in (11), the sum capacity of the original MIMO BC with full CSI is equal to the sum capacity of the set of parallel channels (12), given by

$$\max_{\sum_{g=1}^G \text{tr}(S_g) \leq P} \sum_{g=1}^G \log \left| I + \Lambda_g^{1/2} W_g S_g W_g^H \Lambda_g^{1/2} \right|.$$ \hspace{1cm} (13)

where $S_g$ denotes the diagonal $K_g \times K_g$ input covariance matrix for group $g$ in the dual MAC channel.

This can be intuitively verified by noticing that the transformed channel $H_g$ with reduced dimensionality is unitarily equivalent to the original channel $H_g$ under the tall unitary condition so that we can get channel dimension reduction without loss of optimality. This reduction effect enables significant savings in both downlink training and CSI feedback. Throughout this paper, transmit correlation diversity is advocated as a hidden useful resource in MIMO wireless networks, as exemplified by Fig. 2.

**IV. IS TRANSMIT CORRELATION BENEFICIAL IN MU-MIMO COMMUNICATIONS?**

A fundamental question as to whether transmit correlation is harmful or not to the capacity of MU-MIMO wireless systems motivates our study in this section. Traditionally, transmit correlation has been considered to be a detrimental source to the capacity of MIMO wireless channels. This is the case in point-to-point MIMO systems [3], [4], [6]. One exceptional case where transmit correlation is shown to be beneficial is when the capacity-achieving input covariance is non-isotropic and when SNR is sufficiently low [6]. In the multiuser context, the traditional view on transmit correlation is inherited by the result of [8], which showed that transmit correlation is always harmful to the sum-rate scaling law in the special case when all users have the same transmit correlation.

On the contrary, as mentioned in the introduction, a beneficial impact of transmit correlation on the sum capacity was partially addressed in a simple channel model [13]. Throughout this paper, we advocate that transmit correlation is indeed advantageous to the capacity of a more general MIMO BC as well. In this section, under the unitary structure, we provide some high-SNR analyses for correlated fading MIMO BCs in a compact form. We also characterize the high-SNR capacity in the large $M$ regime. Furthermore, in the large $K$ regime, where the number of users is much larger than the number of transmit antennas, we address what is an optimal sum-rate scaling law of correlated fading MIMO BCs.
A. High-SNR Analysis

For $M$ fixed, we first investigate ergodic sum rates of MIMO BC at high SNR to clearly capture a rate gap between the i.i.d. Rayleigh fading channel and the correlated Rayleigh fading channel satisfying the unitary structure. We assume that every realization of $\mathbf{H}_g$ is perfectly known at the BS. Using Theorem 1 and some well-known results of random matrix theory in Appendix A, at high SNR, we get the following result.

**Theorem 2.** Suppose the perfect CSIT on $\mathbf{H}_g$ in (9) and the unitary structure. For $K_g > r_g, \forall g$, the high-SNR capacity of the corresponding correlated Rayleigh fading MIMO BC scales like

$$C_{\text{sum}}(P) = M \log \frac{P}{M} + \log e \sum_{g=1}^{G} r_g \left( -\gamma + \sum_{\ell=2}^{K_g} \frac{1}{\ell} + \frac{K_g - r_g}{r_g} \sum_{\ell=K_g-r_g+1}^{K_g} \frac{1}{\ell} \right) + \sum_{g=1}^{G} \log |\Lambda_g| + O(1)$$

(14)

where $\gamma$ is the Euler-Mascheroni constant. For $K_g \leq r_g$

$$C_{\text{sum}}(P) = K \log \frac{P}{K} + \log e \sum_{g=1}^{G} K_g \left( -\gamma + \sum_{\ell=2}^{r_g} \frac{1}{\ell} + \frac{r_g - K_g}{K_g} \sum_{\ell=K_g-r_g+1}^{r_g} \frac{1}{\ell} \right) + \sum_{g=1}^{G} \sum_{i=1}^{K_g} \log \lambda_{g,i} + O(1).$$

(15)

**Proof:** See Appendix B.

We can generalize the above result to the tall unitary structure for which $\sum_{g=1}^{G} r_g \leq M$ and $M$ in (14) and (15) is replaced by $\sum_{g=1}^{G} r_g$. Also, we can allow each group to have any of the two conditions on $K_g$ and $r_g$. In particular, (15) becomes asymptotically tight with $O(1)$ replaced by $o(1)$, when $K_g = r_g$.

An alternative expression of (15) can be found in (107) of Appendix B, based on the approach in [7], [27]. Comparing with the alternative characterization, we can see that (15) in Theorem 2 is more intuitive and insightful, even though the remaining term is $O(1)$ instead of $o(1)$.

B. Large System Analysis

We turn our attention to the large $M$ regime, i.e., the large system analysis. In this case, we need the asymptotic behavior of large-dimensional Wishart matrices. To this end, the best known approach is using the Marčenko-Pastur law [28]. In the present paper, we shall instead use (78) in Appendix A to avoid a more involved definite integral calculation than [29].

To concisely characterize the resulting asymptotic capacity, let $\Lambda$ be a diagonal matrix consisting of dominant eigenvalues and assume $\Lambda_g = \Lambda, r_g = r, K_g = K = K' = \frac{K}{G}, \forall g$, and $G$ divides $M$ and $K$. Let

$$G = \frac{M}{r} \quad \text{and} \quad \mu = \frac{M}{K}$$

be fixed. In other words, both $r$ and $K$ can increase to infinity along with $M$.

**Theorem 3.** Suppose the perfect CSIT on $\mathbf{H}_g$ and the unitary structure. As $M \rightarrow \infty$, for $\mu < 1$, the high-SNR capacity of correlated fading MIMO BCs scales linearly in $M$ with the ratio

$$C_{\text{sum}}(P, M, r) = \log \frac{\text{SNR}}{e\mu} + \left( \frac{1-\mu}{\mu} \right) \log \frac{1}{1-\mu} + O(1).$$

(16)

For $\mu \geq 1$, the high-SNR capacity scales linearly in $K$ with the ratio

$$C_{\text{sum}}(P, M, r) = \log \frac{\mu\text{SNR}}{e} + (\mu - 1) \log \frac{\mu}{\mu - 1} + O(1).$$

(17)

**Proof:** See Appendix C.

These asymptotic behaviors will be compared with i.i.d. fading channels in Sec. IV-D.

$^3$ Although these results assume that only the distribution of the channel is accessible at the transmitter, the difference with the full CSIT case that we are assuming vanishes at high SNR when the number of transmit antennas is greater than or equal to the total number of receive antennas (i.e., $K_g \leq r_g$).
C. Large $K$ Analysis

In this subsection, we consider the case of $K \gg M$. It is well known from Sharif and Hassibi \cite{9} that the sum capacity of i.i.d. Rayleigh fading MIMO Gaussian BCs scales like

$$C_{\text{sum}}^\text{i.i.d.}(K) = M \log \log K + M \log \frac{P}{M} + o(1)$$

(18)

where $o(1)$ goes to zero as $K \to \infty$. Furthermore, the same scaling law was shown to be achievable by the random beamforming (RBF), which requires only partial CSIT, and later on by the zero-forcing beamforming with semi-orthogonal user selection (ZFBF-SUS) \cite{23} that requires full CSI but exhibits a better system throughput.

If channels have transmit correlation, we can find some results on scaling laws for correlated MIMO BCs. In particular, when all users have both the same SNR and the common transmit correlation matrix $R$ of full rank, the authors in \cite{8} proved that the sum capacity scales like

$$M \log \log K + M \log \frac{P}{M} + \log |R| + o(1).$$

(19)

This implies that transmit correlation is purely detrimental to the asymptotic capacity since $|R| \leq 1$.

As an immediate consequence of this result, we have the following general form.

**Corollary 1.** Assuming that all users have the same $R$ with rank $r \leq M$, we have the sum-rate scaling law

$$r \log \log K + r \log \frac{P}{r} + \log |\Lambda| + o(1).$$

(20)

A simple proof can be given by the identity

$$\log \left| I_M + U \Lambda^{1/2} W S W^H \Lambda^{1/2} U^H \right| = \log \left| I_r + \Lambda W S W^H \right|$$

$$= \log |\Lambda^{-1} + W S W^H| + \log |\Lambda|$$

(21)

as before. In this rank-deficient case, we replaced $|R|$ in (19) with the product of non-zero eigenvalues of $R$ and also $M$ with $r$. The SNR $\frac{P}{M}$ must be replaced by $\frac{P}{r}$ since we transmit only $r$ independent data streams instead of $M$ streams for the full-rank case. Rewriting (20) as

$$M \log \log K + M \log \frac{P}{M} + c + o(1)$$

(22)

where $c = (r - M) \log \log K + r \log \frac{P}{r} - M \log \frac{P}{M} + \log \prod_{i=1}^{r} \lambda_i$, we can easily see that the full-rank case in (19) is larger than (22) regardless of $K$ unless $P$ is small.

The assumption that all users have the same transmit correlation is generally far from the distribution of transmit correlation in a realistic MU-MIMO system. In what follows, we present a scaling law for a more general MIMO BC, in which users have different transmit correlation with the unitary structure in Definition 2 as an intermediate step to a realistic case in Sec. VI not requiring the ideal structure.

**Theorem 4.** Suppose the perfect CSIT on $H_g$ and the unitary structure. For fixed $M$ and large $K$, the capacity of correlated Rayleigh fading MIMO BC is

$$C_{\text{sum}}(K) = M \log \log K + M \log \frac{P}{M} + \sum_{g=1}^{G} \log |\Lambda_g| + o(1).$$

(23)

**Proof:** See Appendix D.

An important implication of the above theorem is that the sum capacity of “spatially well-colored” Gaussian MIMO BC can be much larger than that of “spatially white” Gaussian MIMO BC, by fully exploiting transmit correlation diversity, as will be shown in the next subsection.
D. Comparison with Independent Fading Channels

For the purpose of comparison between independent fading and correlated fading channels, assume that \( r_g = r, K_g = K', \forall g \), and \( M \) and \( K \) are divided by \( G \), and also assume \( \lambda_{g,i} = \frac{M}{r} = G \) for all \((g,i)\). In this case, the eigen-beamforming gain term in (14) and (15) is given by

\[
\sum_{g=1}^{G} \min(r,K') \sum_{i=1}^{\min(M,K')} \log \lambda_{g,i} = \min(M,K) \log G
\]

(24)

which is in fact the upper bound in (111).

With the optimistic choice of (24), we can see that when \( r = K' \) (i.e., \( M = K \)), the high-SNR capacity (15) reduces to

\[
C_{\text{sum}}(P) = M \log \frac{P}{M} + M \left( -\gamma + \sum_{\ell=2}^{r} \frac{1}{\ell} \right) \log e + M \log G + o(1)
\]

(25)

which approaches close to the high-SNR capacity of the i.i.d. Rayleigh fading MIMO channel [7].

\[
C(P) = M \log \frac{P}{M} + M \left( -\gamma + \sum_{\ell=2}^{M} \frac{1}{\ell} \right) \log e + o(1).
\]

(26)

It is clear that (25) coincides with (26) as \( M \) (equivalently, \( r \)) increases. Compared to \( C(P) \), transmit correlation diversity incurs power loss in \( C_{\text{sum}}(P) \) due to channel dimension reduction, represented by the offset between the second terms in (25) and (26), i.e., \( -M \left( \sum_{\ell=2}^{M} \frac{1}{\ell} \right) \log e \), and simultaneously provides eigen-beamforming gain of \( M \log G \), which compensates the power loss. It can be further seen that for \( r \geq K' \), the high-SNR capacity (15) of correlated fading BC is the same as that of independent fading BC.

For \( r < K' \), suppose that the receivers are allowed to cooperate inside each group, which we call the partial cooperation in this work. Then, we have the following result.

Corollary 2. Assuming the partial cooperation between the receivers within each group, we have

\[
C_{\text{sum}}(P,K) = M \log K' + M \log \frac{P}{M} + \sum_{g=1}^{G} \log |\Lambda_g| + o(1).
\]

(27)

This can be easily verified with the high-SNR upper bound in (95) and the fact that

\[
\mathbb{E} \left[ \log |W_g W_g^H| \right] \simeq r \log K', \quad \text{for large } K'
\]

which follows from (97) and (80). The partial cooperation is shown to provide the additional power gain of \( M ( \log \frac{K'}{G} - \log \log K) \) at high SNR with large \( K' \), compared to (23) with no cooperation. Assuming (24), we easily see that (27) becomes \( M \log K + M \log \frac{P}{M} + o(1) \), which is the asymptotic capacity of the i.i.d. Rayleigh fading MIMO channel for large \( P \) and \( K \) (here, the number of receive antennas) [7, Corollary 1].

In the large \( M \) regime, we can easily see that when \( M = K \), (17) reduces to

\[
\frac{C_{\text{sum}}(P,M)}{M} = \log \frac{\text{SNR}}{e} + o(1)
\]

(28)

which equals the well-known ratio of the i.i.d. Rayleigh fading MIMO channel [1]. If \( \mu \to 0 \), (16) reduces to

\[
\frac{C_{\text{sum}}(P,M)}{M} = \log \frac{\text{SNR}}{e\mu} + O(1)
\]

Since users in a particular group are often closely located, the partial cooperation within such a group seems more feasible than the full cooperation across all users over the entire cell area.
and if \( \mu \to \infty \), (17) reduces to

\[
\frac{C_{\text{sum}}(P, M)}{K} = \log \frac{\mu \text{SNR}}{e} + O(1).
\]

The two ratios are also the same as those of the i.i.d. Rayleigh fading MIMO channel.

For \( r \ll K' \), we present the following corollary of Theorem 4.

**Corollary 3.** If \( \Delta_g \) is sufficiently small but there is no dominant line-of-sight component and hence Rayleigh fading is still valid, then

\[
\limsup_{\Delta_g \to 0, \forall g} C_{\text{sum}}^*(K) - C_{\text{i.i.d}}^*(K) = M \log M. \tag{29}
\]

**Proof:** Since the limit \( \Delta_g \to 0 \) implies \( r_g \to 1 \) for all \( g \), we have \( G = M \) and \( \lambda_{g,1} \to M \) due to the fact that \( \text{tr}(\Lambda_g) = M \). Using Theorem 4, the sum rate of this ideal MIMO BC scales in the limit of \( \Delta_g \to 0 \) like

\[
M \log \log K + M \log \frac{P}{M} + M \log M + o(1).
\]

Comparing this with (18), we conclude (29).

This corollary shows the interesting result that the potential sum-rate gap between the ideal correlated fading case and the independent fading case scales like \( M \log M \).

Under the unitary structure, the above full-CSIT capacity results can be summarized as follows.

- For \( r \geq K' \), the high-SNR capacity of correlated fading BCs can be equal to that of independent fading BCs.
- For \( r < K' \), correlated fading BCs can have the same high-SNR capacity as independent fading MIMO channels if the partial cooperation is allowed, since the loss due to channel dimension reduction can be fully compensated by the eigen-beamforming gain. In contrast, an independent fading BC needs the full cooperation to do so. But, this is not a BC any more.
- For \( r \ll K' \), correlated Rayleigh fading BCs can have even larger sum-rate scaling than the i.i.d Rayleigh fading BC. This is because the power gain due to eigen-beamforming can fully compensate the multiuser diversity gain reduction and even surpass the full multiuser diversity gain available in independent fading.

The above summary shifts the traditional view on transmit correlation from detrimental to beneficial effects. Assuming the prefect CSIT with no cost, the sum rate of even the most optimistic and ideal correlated fading MIMO BC is at best equal to the capacity of independent fading MIMO channel. It will be shown in the following section that this is not necessarily the case with practical pilot-aided systems, where CSIT is provided at the cost of downlink training.

Finally, the reader is encouraged to refer to our companion paper [30] for more detail in the large \( M \) regime with \( M > K \).

**V. FUNDAMENTAL LIMITS OF PILOT-AIDED MU-MIMO SYSTEMS**

In this section, we investigate the fundamental limits of pilot-aided MU-MIMO downlink systems, in which the resources for downlink training are taken into consideration. While FDD systems make use of downlink pilot and CSI feedback for downlink training, time division duplexing (TDD) can employ only uplink pilot thanks to channel reciprocity. In FDD, a pilot-aided downlink system devotes the training phase of length \( M_0^* \) to allow users to estimate the \( M_0^* \)-dimensional channel vectors, where

\[
M_0^* = \min\{M, K, \lfloor T_c/2 \rfloor\}.
\]

Assuming that CSIT is acquired by delay-free and error-free feedback with no channel estimation error, the high-SNR sum rate of MU-MIMO downlink systems is upper-bounded by \( M_0^*(1 - M_0^*/T_c) \log \text{SNR} + O(1) \),
as mentioned in [1] and also [20]. This upper bound is also valid in TDD. For example, the number of co-scheduled users (i.e., \(s\)) is limited by uplink pilot overhead (also affected by \(T_c/2\)) in TDD massive MIMO systems [17]. If instantaneous feedback within coherence time \(T_c\) is not possible, the impact of the resulting channel prediction error on the system multiplexing gain can be found in [31].

As already pointed out, in the large \(M\) regime, the factor \(T_c/2\) may significantly reduce the system performance. Noticing that this result holds true in the i.i.d. fading MIMO BC, however, we will characterize the fundamental limits in correlated fading channels with the notion of transmit correlation diversity. As before, we suppose that the unitary structure is satisfied and that \(T_c\) is finite.

A. Downlink Training: Pre-beamformed Common Pilot

The common pilot is in general isotropically transmitted, since it has to be seen by all users. We consider a training scheme where the common pilot signal is given by the pre-beamforming matrix \(B_g\) as follows.

\[
X^\text{tr}_g = B_g U^\text{tr}_g
\]

where \(U^\text{tr}_g\) is a scaled unitary matrix of size \(r_g \times r_g\). Based on the noisy observation of the pilot signal, user \(g_k\) can estimate the transformed channel \(h^\text{tr}_{g_k}\), which is unitarily equivalent to \(h_{g_k}\) under the unitary structure, as shown in Sec. III. Therefore, the proposed common pilot incurs no loss due to pre-beamforming, as if it were an isotropic pilot to users in each group.

At this point, we compare JSDM with SDMA in terms of downlink training. SDMA requires orthogonal common pilots since, unlike user data streams, the common pilot cannot be multiplexed by SDMA. In contrast, the virtual sectorization of JSDM allows a common pilot resource to be reused over all orthogonal groups (virtual sectors), yielding a remarkable saving in the training phase. Under the unitary structure, the \(G\) degrees of transmit correlation diversity allows the use of \(G\) multiple orthogonal training signals in a single channel use and hence it can reduce the training overhead by a factor of \(G\). This can provides an insight for designing a training scheme in large-scale MIMO. The training overhead reduction makes a viral impact on fundamental limits in pilot-aided systems, as will be seen in the following subsection.

B. Pilot-Aided System I

Following the pilot-aided system proposed in [15], suppose that we use \(M'\) of \(M\) BS antennas in the communication phase, where \(M' \leq M\). The total number of degrees of freedom for communication is upper-bounded by

\[
\min\{M', K\} \left( T_c - \frac{M'}{G} \right)
\]

where \(M'/G\) is due to the above training overhead reduction. So, we need to devote only \(M'/G\) channel uses to the training phase. It is easy to show that the optimal number of transmit antennas to use is

\[
M^* = \min \left\{ M, K, \left\lfloor \frac{T_c G}{2} \right\rfloor \right\}
\]

yielding the pre-log factor

\[
M^* \left( 1 - \frac{M^*}{T_c G} \right)
\]

and the fundamental limit (2), which is one of our main results. From (31), we can see that, for large \(M\) and \(K\), exploiting the \(G\) degrees of transmit correlation diversity can increase the system multiplexing gain up to by a factor of \(G\). Also, we can easily derive the following important observation.
Proposition 1. If \( G \geq \left\lceil \frac{2 \min\{M,K\}}{T_c} \right\rceil \), then
\[
M^* = \min\{M, K\}
\]

The above result implies that as long as the degrees of transmit correlation diversity is sufficiently large such that \( G \geq \left\lceil \frac{2 \min\{M,K\}}{T_c} \right\rceil \), the optimal number of transmit antennas \( M^* \) is not affected any longer by the coherence time \( T_c \). As a consequence, the multiplexing gain can continue growing as \( \min\{M, K\} \) increases.

Example: Let \( T_c \) be 32 LTE OFDM symbol time \([32]\) (approximately 60 km/h) and 100 symbol time (19 km/h). Also, suppose that the unitary condition is attained such that \( G = 4 \) and \( G = 8 \). Fig. [1] shows the Zheng-Tse upper bound, \( M^*_o(1 - M^*_o/T_c) \), and the new bound in \([32]\) on the system multiplexing gain as \( \min\{M, K\} \) increases. It can be seen that exploiting transmit correlation diversity can increase the multiplexing gain up to by a factor of 4 and 8 for \( G = 4 \) and \( G = 8 \), respectively.

In what follows, we refine the \( O(1) \) term in \((2)\) in the large \( G \) regime and then in the large \( r \) regime. So far, we have fixed the degrees of transmit correlation diversity, \( G \). We now turn our attention to the case that as \( M \to \infty \), \( G \) also grows. Therefore, suppose that the ratio \( G/M \) is not vanishing (i.e., bounded), as \( M \) increases. In practical systems, we generally consider a large-scale array in the high carrier frequency \( f_c \) due to the space limitation of large-scale array and the fact that the wavelength is inversely proportional to \( f_c \). So, it is fairly reasonable to increase \( M \) proportionally as \( f_c \) grows and hence to let \( M \) depend on \( f_c \). The higher \( f_c \), the smaller number of strong multipaths the receivers experience due to higher directionality (e.g., in millimeter wave (mm-Wave) channels \([33]\)). This sparsity of dominant multipath components is also verified by mm-Wave propagation measurement campaigns \([34]\). Thus, the high transmit correlation diversity is attainable in the high \( f_c \) case. Then, as both \( M \) and \( f_c \) go to infinity, \( r_g \) may remain unchanged such that \( G/M \) is fixed.

Since the purpose of this work is to concisely characterize the asymptotic capacity in a familiar and closed-from expression, assume \( \Lambda_g = \Lambda, r_g = r, K_g = K', \forall g \), as before, and also let \( r^* = \frac{M^*}{G} \) and \( \mu^* = \frac{M^*}{K} \) be fixed, where \( M^* \) is divided by \( G \). Given a finite \( T_c \), let the ratio
\[
\frac{M^*}{T_c G} = \nu
\]
be fixed with \( M^* \to \infty \). In this scenario, \( G \) grows to infinity along with \( M \) but both \( r \) and \( K' \) are finite, unlike Theorem 3. Therefore, we shall make use of \((76)\) instead of \((78)\) in Lemma 1 and then we can reuse Theorem 2 in the sequel.

Denote by \( C_{\text{pilot1}}(P, M^*, \nu) \) the high-SNR capacity of MIMO BC in pilot-aided system 1 for \( M^* \) and \( \nu \) large, where \( \nu = r \) or \( G \). Assuming that the perfect CSIT is provided by an ideal (i.e., delay-free and error-free feedback) uplink with no channel estimation error nor pilot contamination in TDD, \( C_{\text{pilot1}}(P, M^*, \nu) \) is simply given by
\[
C_{\text{pilot1}}(P, M^*, \nu) = (1 - \nu)C_{\text{sum}}(P, M^*, \nu).
\]

For \( \nu \leq 1/2 \) and \( \mu < 1 \), we have \( M^* = M, r^* = M/G = r, \mu^* = \mu \). By using \((14)\), the ratio at which the high-SNR capacity increases in the large \( G \) regime as \( M \to \infty \) is
\[
\frac{C_{\text{sum}}(P, M, G, \nu)}{M} = \log \frac{P}{M} + \log e \sum_{g=1}^{G} \frac{r}{M} \left( -\gamma + \sum_{\ell=2}^{K'} \frac{1}{\ell} + \frac{K' - r}{r} \sum_{\ell=K' - r + 1}^{K'} \frac{1}{\ell} \right) + \frac{1}{M} \sum_{g=1}^{G} \log |\Lambda_g| + O(1)
\]
\[
= \log \frac{\text{SNR}}{r} + \log e \left( -\gamma + \sum_{\ell=2}^{K' \mu} \frac{1}{\ell} + \left( \frac{1 - \mu}{\mu} \right) \sum_{\ell=(1-\mu)K'+1}^{K'} \frac{1}{\ell} \right) + O(1)
\]
(33)

\(^7\)Not assuming the perfect channel estimation and the unitary structure, the impact of the resulting noisy CSIT on the performance of JSDM is evaluated in \([20]\).
where we used (111) and replaced $P$ with SNR for a more familiar form. For $\nu \leq 1/2$ and $\mu \geq 1$, noticing that $M^* = K, r^* = K/G = K', \mu^* = 1$ and using (15), we have

$$\frac{C_{\text{sum}}(P, K, G)}{K} = \log \frac{P}{K} + \log e \sum_{g=1}^{G} \frac{K'}{K} \left( -\gamma + \sum_{\ell=2}^{K'} \frac{1}{\ell} \right) + \frac{1}{K} \sum_{g=1}^{G} \sum_{i=1}^{K'} \log \lambda_{g,i} + O(1)$$

$$= \log \frac{\text{SNR}}{K'} + \log e \left( -\gamma + \sum_{\ell=2}^{K'} \frac{1}{\ell} \right) + O(1)$$

(34)

where we used (112). When $\nu > 1/2$, the ratios for the two cases of $\mu < 1$ and $\mu \geq 1$ can be similarly obtained by noticing $M^* = \frac{T_c G}{g}$ and $r^* = \frac{r}{g}$. As a result, for all four cases in the large $G$ regime with $r^*$ fixed, we can get the unified formula

$$\frac{C_{\text{sum}}(P, M^*, G)}{M^*} = (1 - \nu) \left\{ \log \frac{\text{SNR}}{r^*} + \log e \left( -\gamma + \sum_{\ell=2}^{K'} \frac{1}{\ell} + \frac{1 - \mu^*}{\mu^*} \sum_{\ell=(1 - \mu^*)K'}^{K'} \frac{1}{\ell} \right) \right\} + O(1).$$

(35)

The following result shows the optimistic potential gain of transmit correlation diversity in the limit of $\Delta_g \to 0$ for all $g$, which we provide as a reference, even though this situation is unrealistic.

**Proposition 2.** Suppose the unitary structure. For $\mu = 1$ and $T_c \geq 2$, as $\Delta_g \to 0$ and $M \to \infty$ with $\mu$ and $\nu$ fixed, the high-SNR capacity of the pilot-aided system I scales linearly in $M$ with the maximum possible ratio

$$\limsup_{\Delta_g \to 0} \frac{C_{\text{sum}}(P, M)}{M} = (1 - T_c^{-1}) \log \frac{\text{SNR}}{e} + o(1).$$

(36)

We first notice that the condition of Proposition 1 can be restated as $T_c \geq 2 \min(r, K')$ and hence the sufficient condition is guaranteed just for $T_c \geq 2$, since $r \to 1$ as $\Delta_g \to 0$. Using this fact and (35), (36) immediately follows with $\nu = T_c^{-1}$. It is remarkable that if the unitary structure is attained with $T_c$ sufficiently large and $\Delta_g$ sufficiently small, the optimistic high-SNR capacity of MU-MIMO systems becomes equivalent to that of the full-CSI capacity in (28).

In the large $r$ regime where $r$ (instead of $G$) goes to infinity, using Theorem 3 we obtain the following result. For $\mu^* \leq 1$

$$\frac{C_{\text{sum}}(P, M^*, r^*)}{M^*} = (1 - \nu) \left\{ \log \frac{\text{SNR}}{e\mu^*} + \left( \frac{1 - \mu^*}{\mu^*} \right) \log \frac{1}{1 - \mu^*} \right\} + O(1).$$

(37)

Note that according to (31) and (32), the case of $\mu^* > 1$ does not happen in pilot-aided system I, since we make use of only $K$ transmit antennas regardless of how large $M$ is, i.e., $M^* = K$. This phenomenon in system I may cause a nontrivial rate loss for $M > K$, as will be discussed in the next subsection.

Fig. 3 shows that the asymptotic sum rate curves in pilot-aided systems I. For large $G$ and fixed $r^*$, the multiplexing gain grows linearly with $\min\{M, K\}$, whereas this is not the case with large $r^*$ and fixed $G$. For large $G$ and fixed $r^*$, the upper bound on the asymptotic capacity of $\mu < 1$ is shown to be larger than that of $\mu \geq 1$, which stems from the foregoing observation that the latter case uses only $K$ transmit antennas. To understand the large rate gap between $\mu = 1$ and $\mu = 10$, recall that at high SNR, a dual MAC is equivalent to the corresponding MIMO point-to-point channel with $K$ transmit antennas and $M$ receive antenna. The equivalent MIMO channel is well understood to have a power gain scales logarithmically with $M$ due to receive beamforming. For $\mu = 0.1$ case, the large rate gap from $\mu = 1$ is because the upper bound was given by allowing the partial cooperation within each group. In addition, we can see that there is a crossover between large $G$ and large $r^*$ when $\min\{M, K\}$ is not so large. Finally, for large $r^*$, the two cases of $\mu = 1$ and $\mu = 10$ collapse into the red solid line. This absurd situation will be resolved in the following subsection.
C. Pilot-Aided System II

In the large $M$ regime, the $\mu > 1$ case would be more frequently encountered in practice. We introduce a new pilot-aided system to address the foregoing issue for this case with $r^*$ large ($G$ fixed). In sharp contrast to pilot-aided system I [15], in which only $K$ transmit antennas are used by letting $M^* = K$ when $M > K$, we shall allow in the new system (we call this pilot-aided system II) to use more than $K$ transmit antennas, even though the degrees of freedom is certainly at most $K$. By doing so, we have an opportunity to obtain fairly large power gain suggested by (17) due to transmit correlation diversity which compensates the increase in channel uses required for downlink training. To understand this, notice that using more than $K$ antennas gets less and less impact on the system multiplexing gain as $G$ and/or $T_c$ grows, as shown in (32).

To take into account the additional power gain from using more than $M^*$ transmit antennas in pilot-aided system II, we replace the optimization problem in (30) with the following one based on (17).

$$M^* = \arg\max_{M'} f(M')$$

subject to $M' \leq M$, where $f(M') = M^* \left( T_c - \frac{M'}{G} \right) \log \frac{P}{e} M' + (\frac{M'}{K} - 1) \log \frac{M'}{M' - K}$ with $M^*$ (the maximum number of degrees of freedom for the communication phase) unchanged.

Fig. 4 shows the optimum number of transmit antennas, $M^*$, for different $P$ and $\mu$. Here, $M^* = K = 20$ for $\mu = 10$ and $M^* = K = 50$ for $\mu = 4$. Therefore, if we consider not only the multiplexing gain but also the power gain due to eigen-beamforming, the optimum values of $M^*$ are shown to be quite different from $M^*$.

The high-SNR capacity of pilot-aided system II for $\mu > 1$ scales linearly in $K$ with the ratio

$$\frac{C_{\text{sum}}(P, M^*, r^*)}{K} = (1 - \nu^*) \left\{ \log \frac{\mu^* \text{SNR}}{e} + (\mu^* - 1) \log \frac{\mu^*}{\mu^* - 1} \right\} + O(1)$$

(39)
Fig. 4. Values of \( f(M') \) versus the number of transmit antennas to use, \( M' \), in pilot-aided system II when \( M > K \) (i.e., \( \mu > 1 \)), where \( M = 200, T_c = 64 \), and the ‘o’ indicates the optimum numbers of transmit antennas, \( M^* \), and the ‘x’ indicates \( M^* \).

where \( \frac{M^*}{T_G^G} = \nu^* \) and \( \mu^* = \frac{M^*}{K} \). Fig. 5 compares the asymptotic sum rates of pilot-aided system I and II when \( T_c = 32 \) and \( T_c = 128 \). It is shown that the rate gap gets larger as \( T_c \) increases, since, for large \( T_c \), the extra overhead due to training more than \( K \) antennas reduces, as mentioned earlier.

VI. Capacity Scaling with Non-ideal Transmit Correlations and Limited Feedback

So far, we have assumed the ideal unitary structure of transmit correlations, for which the transmit correlation diversity gain is maximally attainable without a loss of optimality due to the limited CSI feedback of JSDM. In a realistic system, however, we generally cannot construct such a favorable structure that users in a group have a common transmit correlation and that all groups strictly satisfy the orthogonality condition. It has been further assumed that the BS perfectly knows the transformed channel state \( \mathbf{H}_g \) fed back by users. In what follows, we review a particular scheme of JSDM, proposed in [35]. This scheme will be shown to approach close to the optimal sum-rate scaling law in Theorem 4 in realistic systems, requiring neither the unitary structure nor \( \mathbf{H}_g \).

A. Group-Specific Beamforming with Semi-Unitary User Partition

For a given distribution of transmit correlation matrices \( \mathbf{R}_g \), we need to appropriately choose \( \mathbf{U}_g^{(t)} \) to create a useful structure of transmit correlations that approximately realizes the ideal unitary structure. A natural criterion for selecting \( \mathbf{U}_g^{(t)} \) is to minimize the chordal distance between \( \mathbf{U}_i \) and its associated \( \mathbf{U}_g^{(t)} \) and to simultaneously maximize the minimum of generalized (to be defined in (67)) chordal distances between \( \mathbf{U}_g^{(t)} \)'s of \( G_t \) groups in class \( t \) in favor of the orthogonality. Roughly speaking, the latter leads to the interpretation that for each class \( t \), the resulting \( \{ \mathbf{U}_g^{(t)}, \forall g \} \) are “semi-unitary” such that \( \mathbf{U}_g^{(t)^H}\mathbf{U}_g^{(t)} \approx \mathbf{I} \) where we assumed that \( \sum_{g=1}^{G_t} r_g^{(t)} \leq M \) for all \( t \). Provided that \( \mathbf{U}_g^{(t)} \) are obtained in the above way, we do the following user partition based on fixed pre-beamforming matrices \( \mathbf{B}_g^{(t)} \).
Fig. 5. Asymptotic capacity comparison of two pilot-aided systems where $\mu = 10$, $G = 10$, and SNR = 20 dB.

1) Construction of $B_g^{(t)}$: We may let $B_g^{(t)} = U_g^{(t)}$ for the similarity or let $B_g^{(t)}$ satisfy the block diagonalization (BD) constraint \cite{25} on the set of $\{U_h^{(t)}, h \neq g\}$ to improve the orthogonality. In this section, we choose the former.

2) Class and group association: Assuming that user $i$ has the perfect knowledge on $R_i$, it chooses and feeds back to the BS the indices $(t, g)$ of class and group that maximize the following similarity measure

$$\| R_i^H B_g^{(t)} \|_F$$

or that minimize the generalized chordal distance between $U_i^*$ and $B_g^{(t)}$.

We collectively refer to user partition algorithms satisfying the above selection criterion of $U_g^{(t)}$ as the semi-unitary user partition (SUP) algorithm.

The semi-unitary structure created by the above SUP algorithm can be formally defined in the following form, where the two matrix constraints are given in terms of the similarity and the orthogonality with tolerable threshold levels $\alpha$ and $\beta$, respectively. We omit $^{(t)}$ and let $r_g = r$ for notational convenience.

**Definition 3** (Semi-Unitary Structure). For $M \geq rG$, we have a semi-unitary structure if the users $g_k$ in $G$ satisfy

$$\frac{1}{r} \| U_g^H U_{g_k}^* \|_F^2 \geq \alpha$$  \hspace{1cm} (41)$$

$$\frac{1}{r} \| U_g^H U_h^* \|_F^2 \leq \beta$$  \hspace{1cm} (42)$$

where $\alpha$ and $\beta$ are positive and nonnegative real numbers, respectively.

The semi-unitary structure is to approximate the unitary structure in Definition \[2\] for which $\alpha$ and $\beta$ become one and zero, respectively. In general, the levels depend on a number of factors including the number of classes $T$, the number of groups $G$, the effective ranks $r_g$ for the various groups, and the
systems or problems of interest. In the quest where we show that GBF-SUP achieves near-optimal sum-rate scaling for large $K$, admissible values of $\alpha$ and $\beta$ will be explicitly given in the following subsection with additional constraints.

We assume $|G_g| = O(K)$ in the large number of users regime, as before. Given a SUP algorithm satisfying the semi-unitary structure, we propose the GBF scheme, a special case of JSDM. The basic idea of GBF and its CSI feedback is based on opportunistic beamforming [22] and RBF with partial CSIT [9]. The generic transmit signal vector of JSDM for group $g$ can be written as

$$x_g = B_g P_g d_g.$$  

The GBF scheme is defined by letting $P_g$ be a beam selection matrix. If a user is selected for the $m$th beam vector (column of $B_g$) by the scheduler, the beam is mapped by the permutation matrix $P_g$ to the corresponding data element of $d_g$. The rationale behind GBF is that by virtue of the SUP algorithm, the eigenspaces of transmit correlation matrices $R_{g_k}$ in group $g$ are supposed to be fairly similar to that of $R_g$. Thus, the pre-beamforming matrix $B_g$ approximates the eigen-beamformer $U_{g_k}^*$ dedicated for each user $g_k$ in group $g$, unlike the classical opportunistic beamforming and RBF.

Since GBF treats inter-group interference as noise, the SINR of user $g_k$ with respect to the $m$th beam $b_{g,m}$ is

$$\text{SINR}_{g_k,m} = \frac{|h_{g_k}^H b_{g,m}|^2}{M + \sum_{i=1,i\neq m}^{r_g} |h_{g_k}^H b_{g,i}|^2 + \sum_{h=1,h\neq g}^{G} \sum_{j=1}^{r_g} |h_{g_k}^H b_{h,j}|^2},$$  

The semi-unitary structure created by the SUP algorithm reduces the CSI feedback overhead and ensures that GBF does not suffer from severe inter-group interference. We call this combination as group-specific beamforming with the semi-user partition algorithm (GBF-SUP).

The CSI feedback of GBF-SUP consists of statistical and instantaneous components. The statistical part is a $\lceil \log_2 \left( \sum_t G_t \right) \rceil$-bit feedback on class and group indices $(t, g)$, which are extremely slow varying, and the instantaneous part is the maximum SINR of each user, i.e.

$$\max_{1 \leq m \leq r_g} \text{SINR}_{g_k,m}$$  

where $b_g \leq r_g$, and the corresponding beam index $m$, which amounts to a $\lceil \log_2 b_g \rceil$-bit feedback. Therefore, GBF-SUP can reduce the instantaneous feedback overhead for the beam index of RBF, which amounts to $\lceil \log_2 M \rceil$ bits.

On the other hand, the companion work in [30] treats the case where all users in group $g$ have a common transmit correlation matrix $R_g$ for all $g$. This yields a much easier setup than the problem of interest in this work.

### B. Lower Bound on the SINR of GBF-SUP

For the purpose of the study on sum-rate scaling of realistic correlated fading MIMO BC, we will make use of the extreme value theory [36]–[38]. Sharif and Hassibi [9] proved the asymptotic optimality of the RBF scheme in the i.i.d. Rayleigh fading BC when $K$ is sufficiently large by using extreme value theory. In essence, they showed that SINRs resulting from RBF are i.i.d. random variables and that the distribution of SINR conforms to the sufficient conditions in [36] to apply the well-known result on the asymptotic behavior of the maximum of $n$ i.i.d. random variables for sufficiently large $n$. It seems that this two requirements have restricted the application of their approach to some special channel models. In

---

8This implies that users should immediately report the CSI measurement back to the BS within every coherence time of channel.
particular, a subsequent work in [8] captures the impact of transmit correlation on achievable throughput only in the special case where all users have a common transmit correlation matrix \( R \). This restriction was essential to let the resulting SINR i.i.d. over all users. In a realistic system, however, users have different transmit correlation matrices \( R_g \), which makes our problem much more complicate due to lack of some useful structures. As we do not assume here the ideal unitary structure on \( R_g \), it is very difficult to directly show that GBF-SUP meets the above two requirements. To be specific, SINR\(_{g_k} \) in (43) are not i.i.d. over \( g \) and \( k \). Along with the different correlation matrices \( R_g \) over \( g \) and \( k \), and the presence of inter-group interference, this makes the problem to find the distribution of SINR\(_{g_k} \) much more complicate than the common \( R \) case. Moreover, the dependance among the intended signal, intra-group interference and inter-group interference is quite burdensome as well. Therefore, we need to simplify the problem at hand to avoid such difficulties.

Prior to all the details, we present an outline as follows. We first introduce a pair of matrix constraints to lower-bound the SINR in (43) of GBF-SUP and check if the number of users satisfying the constraints can be still \( O(K) \). In particular, since the constraints are with respect to only large-scale fading component (i.e., transmit correlation) such that small-scale fading component remains an i.i.d. random variable, the resulting lower bound is much more mathematically tractable than the original SINR. Also, we use the generalized Chi-squared distribution in \([39], [40]\) to take into account the sum of independent Gaussian random variables with different variances. Then, we calculate the distribution of the lower bound on SINR to see if it satisfies the sufficient conditions, following the approach in [9]. Finally, we will show that GBF-SUP can satisfy the above two requirements for applying the well-known asymptotic behavior of the maximum of i.i.d. random variables and then prove its asymptotical optimality. This ends with the achievability proof.

Following the GBF-SUP scheme in the previous subsection, we can write its achievable throughput as

\[
\mathcal{R}_{\text{GBF-SUP}} = \mathbb{E} \left[ \sum_{g=1}^{G} \sum_{m=1}^{r_g} \log \left( 1 + \max_{g_k \in G_g} \text{SINR}_{g_k, m} \right) \right] + o(1) \tag{44}
\]

where \( o(1) \) takes into account the fact that user \( g_k \) may be the strongest user of two or more beams \( u_{g,m} \) with a vanishing probability as \( K_g \) goes to infinity.

We assume \( r_g = r \) for all \( g \) and \( r \) divides \( M \) such that \( G = M/r \) is an integer as before. In the Kronecker channel model, the channel vector \( h_{g_k} \) of user \( g_k \) can be decomposed instead of (8) as

\[
h_{g_k}^H = h_{W,g_k}^H R_{g_k}^{1/2}
\]

where \( h_{W,g_k} \in \mathbb{C}^{M \times 1} \sim \mathcal{CN}(0, I) \). We also rewrite the eigendecomposition of \( R_{g_k} \) as

\[
R_{g_k} = U_{g_k} \Lambda_{g_k} U_{g_k}^H,
\]

where \( U_{g_k} \) and \( \Lambda_{g_k} \) denote the \( M \times M \) (full-size) unitary matrix and diagonal matrix of \( R_{g_k} \) of rank \( r_{g_k} \leq M \), respectively, and the diagonal elements (eigenvalues) of \( \Lambda_{g_k} \) are in descending order. Define \( M \times M \) matrix \( A_{g_k} \) and \( M \times M \) matrix \( C_g \), respectively, as

\[
A_{g_k} = R_{g_k}^{1/2} U_{\pi}
\]

\[
C_g = U_{\pi} \Lambda_{g}^{1/2}
\]

where \( U_{\pi} \) is unitary and a permutation of columns of \( U \) defined in (11) such that \( U_{\pi} = [U_g, U_{g+1}, \ldots, U_{g-1}] \) and

\[
\Lambda_g = \begin{bmatrix}
\Lambda_{g} & 0 \\
0 & \alpha_{g} I_{M-r \times M-r}
\end{bmatrix}.
\]

By definition of \( U_{\pi} \), \( \beta \) in (42) becomes zero. Finally, let \( a_{g_k,m} \) and \( c_{g,m} \) be the \( m \)th column of \( A_{g_k} \) and \( C_g \), respectively.
1) Lower Bound: We present a lower bound of SINR\(_{g_k,m}\), which is much more mathematically tractable than the original SINR\(_{g_k,m}\). Define

\[
\text{SINR}'_{g_k,m} \triangleq \frac{\alpha |h_{W,g_k}^H c_{g,m}|^2}{M + \sum_{i=1, i \neq m}^r |h_{W,g_k}^H c_{g,i}|^2 + \sum_{j=r+1}^M |h_{W,g_k}^H c_{g,j}|^2}
\]

and let \( G'_g \) be a subset of \( G_g \), defined by the following matrix constraints.

\[
G'_g = \left\{ g_k \in G_g : a_{g_k,m}^H a_{g_k,m} \succeq \alpha c_{g,m}^H c_{g,m}, \text{ for } m = 1, \ldots, r \right\}
\]

(47)

In fact, the parameters \( \alpha \) and \( \alpha_0 \) (hidden in \( \Lambda_g \)) are penalty factors required to configure the tolerable dissimilarity level. Notice that the equalities hold with \( \alpha = 1 \) and \( \alpha_0 = 0 \) under the unitary structure. Moreover, it is not difficult to show that (47) is a sufficient condition of (41). This can be done by noticing that (47) ensures \( \text{tr}(R_{g_k}^{1/2}U_g U_g^H R_{g_k}^{1/2}) \geq \alpha \text{tr}(U_g \Lambda_g U_g^H) \) and using a well-known majorization result in Lemma 3 of Appendix A. The detail is omitted for space limitation.

We should investigate that for what values of the penalty factors \( \alpha \) and \( \alpha_0 \), the conditions (47) and (48) are met and then prove the existence of a positive real number \( \alpha \). To do so, we shall use Lemma 5 in Appendix A. Reusing some notations therein, let the rank-1 matrix \( c_{g,m}^H g_m \) decomposed as

\[
c_{g,m}^H g_m = [Q_1 \ Q_1^+] \begin{bmatrix} \lambda_{g,m} & 0 \\ 0 & 0 \end{bmatrix} [Q_1 \ Q_1^+]^H
\]

where, for notational convenience, \( Q_1 \) denotes the eigenvector corresponding to the positive eigenvalue \( \lambda_{g,m} \) equal to \( a_{g,m}^H a_{g,m} \), and the columns of \( Q_1^+ \) span the orthogonal complement of range\((c_{g,m}^H g_m)\). Also, let \( \mathcal{P} \) and \( \mathcal{P}^\perp \) be orthogonal bases of range\((Q_1^H a_{g_k,m} a_{g_k,m}^H Q_1^+)\) and ker\((Q_1^H a_{g_k,m} a_{g_k,m}^H Q_1^+)\), respectively, and let \( Q = [Q_1, Q_2, Q_3] \), where \( Q_2 = \mathcal{P} \) and \( Q_3 = \mathcal{P}^\perp \), and \( a_{ij} = Q_{ij}^H a_{g_k,m} a_{g_k,m}^H Q_j \), where \( i, j = 1, 2 \). Then using (87) and (88), we can see that \( a_{g_k,m} a_{g_k,m}^H \succeq \alpha c_{g,m}^H c_{g,m} \) is equivalent to

\[
\begin{align*}
& a_{22} > 0, \quad \alpha \leq \lambda_{g,m}^{-1}(a_{11} - a_{12}a_{22}^{-1}a_{21}) \\
& a_{22} = 0, \quad \alpha \leq \lambda_{g,m}^{-1}a_{11}.
\end{align*}
\]

(49)

To understand the above conditions on \( \alpha \), we study the second (easier than the first) case of \( a_{22} = 0 \), which happens when range\((Q_1^H a_{g_k,m} a_{g_k,m}^H Q_1^+) = \{0\} \). This simple condition can be interpreted as range\((a_{g_k,m} a_{g_k,m}^H) = \text{range}(a_{g_k,m} a_{g_k,m}^H) = \text{range}(u_{g,m}), \forall m \), i.e., the rank-1 matrix \( a_{g_k,m} a_{g_k,m}^H \) has the same eigenvector (corresponding to the non-zero eigenvalue) as \( u_{g,m} \). Notice \( a_{g_k,m} a_{g_k,m}^H u_{g,m} = (R_{g_k}^{1/2} u_{g,m} R_{g_k}^{1/2}) u_{g,m} = R_{g_k}^{1/2} (c u_{g,m}) \), where \( a_{g_k,m} = R_{g_k}^{1/2} u_{g,m} \) by definition and \( c = u_{g,m}^H R_{g_k}^{1/2} u_{g,m} \) takes a nonnegative real value since \( R_{g_k} \) is positive semi-definite. To meet the condition of \( a_{22} = 0 \), we only need to check if there exist \( r \) non-zero eigenvalues of \( R_{g_k}^{1/2} \) whose eigenvectors are \( \{u_{g,m}, \forall m\} \) such that

\[
R_{g_k}^{1/2} (c u_{g,m}) - \lambda_{g_k,s(m)} (R_{g_k}^{1/2} (c u_{g,m}) = R_{g_k}^{1/2} u_{g,m} - \lambda_{g_k,s(m)} (R_{g_k}^{1/2}) u_{g,m} = 0, \quad m = 1, \ldots, r
\]

where \( S \subseteq [1 : r_{g_k}] \) is a subset of size \( r \), with \( r_{g_k} \geq r \), and \( s(m) \) is the \( m \)th element of \( S \). This implies that the subspace spanned by \( r \) eigenmodes of \( R_{g_k} \) is equal to range\((U_g)\). Therefore, if

\[
\text{range}(R_{g_k}) \supseteq \text{range}(U_g)
\]

(50)

then we have \( a_{22} = 0 \). Note that the condition (50) does not necessarily require that \( u_{g_k,s(m)} = u_{g,m} \) for all \( m \), since eigenvectors are not unique. If we let \( Q_1 = u_{g,m} \), then \( a_{11} = \lambda_{g_k,s(m)} \), for all \( (g, k, m) \). Thus, (49) provides the upper bound on \( \alpha \)

\[
\alpha \leq \min_{g,k,m} \lambda_{g_k,s(m)}^{-1} \lambda_{g_k,s(m)}
\]

(51)
Eventually, there exists a positive real value on \( \alpha \) satisfying (47).

Next, the condition (48) can be translated as \( \lambda_{g_k,m} \leq \lambda_{g,m}, \forall m, \) and
\[
\alpha_0 \geq \max_{g,k} \lambda_{g_k,r+1}.
\]

Combining the former condition with (51), we have
\[
0 < \alpha \leq \min_{g,k,m} \lambda_{g_k,m}^{-1}\lambda_{g,m,S(m)} \leq \min_{g,k,m} \lambda_{g_k,m}^{-1} \lambda_{g,m} \leq 1.
\]

The above arguments show that given \( U_g \), there are a broad range of \( R_{g_k} \) which meets the required matrix constraints (47) and (48). We now prove that the original SINR \( g_k,m \) is lower-bounded by SINR\( g_k,m \) for any small-scale component \( h_{W,g_k} \).

**Proposition 3.** For users \( g_k \in G_g', \) where \( G_g' \) is defined in (47) and (48)
\[
\text{SINR}_{g_k,m} \geq \text{SINR}'_{g_k,m}, \; \text{for } m = 1, \ldots, r.
\]

**Proof:** We first check that for arbitrary \( h_{W,g_k} \), if
\[
\begin{align*}
|h_{W,g_k}^H a_{g_k,m}|^2 & \geq \alpha |h_{W,g_k}^H c_{g,m}|^2, \; \forall m \\
|h_{W,g_k}^H A_{g_k}|^2 & \leq |h_{W,g_k}^H C_g|^2
\end{align*}
\]
then (52) is true. This can be easily done by letting \( b_{g,m} = u_{g,m} \) for all \( g \) and \( m \) and by rewriting SINR\( g_k,m \) in (43) and SINR\( g_k,m \) in (46) respectively as
\[
\text{SINR}_{g_k,m} = \frac{|h_{W,g_k}^H a_{g_k,m}|^2}{\sum_{g_k,m} |h_{W,g_k}^H A_{g_k}|^2 - |h_{W,g_k}^H a_{g_k,m}|^2} \quad (55)
\]
\[
\text{SINR}'_{g_k,m} = \frac{\alpha |h_{W,g_k}^H c_{g,m}|^2}{\sum_{g_k,m} |h_{W,g_k}^H C_g|^2 - |h_{W,g_k}^H c_{g,m}|^2} \quad (56)
\]

Since the matrix \( a_{g_k,m} a_{g_k,m}^H - \alpha c_{g,m} c_{g,m}^H \) is positive semi-definite by the constraint in (47), it can be seen that (47) is just a sufficient condition for (53). We show next that (48) is also a sufficient condition for (54).

Observe that (54) is equivalent to
\[
R_{g_k} \preceq U^* \Lambda g \preceq U^H \Lambda g U^H \quad \iff \quad U^H g R_{g_k} U^H \preceq \Lambda g.
\]

Also, it follows from Lemma 2 in Appendix A and the fact that \( U^H \) is unitary that
\[
U^H g R_{g_k} U^H \preceq \Lambda g_k.
\]

As a result, if \( \Lambda g_k \preceq \Lambda g \), we get (54).

Finally, we need to see whether \( |G_g'| = O(K) \) for all \( g \). Assuming ULA, \( R_{g_k} \) is of algebraic full rank. Moreover, it is known to have generally full rank even in actual channel measurements, although eigenvalues except some dominant ones decay fast. Thus, we can get range\( (R_{g_k}) = C^M \supseteq \text{range}(U_g) \) with high probability for any unitary \( U \). However, in the case where \( \text{min}_{g,k} \lambda_{g_k,S(r)} \) is extremely small, the penalty factor \( \alpha \) may be made too small according to (51), which causes a significantly large penalty term \( O(1) \) in sum-rate scaling due to dissimilarity between \( U^* \) and \( U_g \). We can avoid this and reduce the \( O(1) \) term by having more classes, i.e., by increasing \( T \) to improve the similarity, since the chordal distance between \( U^* \) and \( U_g \) reduces as \( T \) increases. Additionally, we can also reduce the penalty term by allowing a small fraction of users not to satisfy (47) and (48) such that there exist some users \( g_k \notin G_g' \). Thus \( |G_g'| = (1-\epsilon)|G_g| \), where \( \epsilon \) is a small constant. The resulting multiuser diversity gain reduction would be negligible for the same reason in the achievability proof of Theorem 4. For \( M \) large, as long as the group angular region of \( U_g \) is contained in the angular region of \( R_{g_k} \) for all users in group \( g \), the
conditions (47) and (48) are naturally attained, since this situation ensures a good approximation of the unitary structure, as shown in [20].

The advantage of obtaining $\text{SINR}'_{g_k,m}$ by imposing the constraints on the large-scale component $R_{g_k}$ is that the lower bound of $\text{SINR}_{g_k,m}$ becomes i.i.d. over users in $G'_g$. Also, the conditions in $G'_g$ are only for $R_{g_k}$ and not for the small-scale component $h_{W,g_k}^H$ so that the distribution of the elements of $h_{W,g_k}$ remains i.i.d. Gaussian in $\text{SINR}_{g_k,m}$. Furthermore, the intended signal, intra-group interference and inter-group interference terms are independent. To check this, notice that

$$h_{W,g_k}^H C_g \sim \mathcal{CN}(0, \Lambda_g)$$

(59)

for all users $g_k \in G_g$ and hence the elements $h_{W,g_k}^H c_{g,m}$ are independent and non-identically distributed. As a consequent, the numerator and the denominator of $\text{SINR}_{g_k,m}'$ are independent for users with $R_{g_k}$ satisfying the covariance constraints (47) and (48). These properties of $\text{SINR}_{g_k,m}'$ are what we sought after.

For large $K_g$, we can lower-bound the expectation term in (44) as

$$\mathbb{E} \left[ \sum_{g=1}^{G} \sum_{m=1}^{r_g} \log \left( 1 + \max_{g_k \in G_g} \text{SINR}_{g_k,m} \right) \right] \geq \mathbb{E} \left[ \sum_{g=1}^{G} \sum_{m=1}^{r_g} \log \left( 1 + \max_{g_k \in G'_g} \text{SINR}_{g_k,m}' \right) \right]$$

(60)

where we used (52) and $G'_g \subseteq G_g$.

C. Distribution of $\text{SINR}'$

In order to evaluate the right-hand side of (60) and obtain a sum-rate scaling law of RBF-SUP, we should figure out the distribution of $\text{SINR}'_{g_k,m}$. Owing to the sought-after properties of $\text{SINR}_{g_k,m}'$, the calculation of probability density function (pdf) and cumulative density function (cdf) of $\text{SINR}_{g_k,m}'$ is much easier than those of the original $\text{SINR}_{g_k,m}$.

Letting the random variables $v$ and $y$ denote the numerator and denominator of $\text{SINR}'_{g_k,m}$, respectively, i.e., $v = |h_{W,g_k}^H c_{g,m}|^2$ and $y = \frac{1}{\alpha} \sum_{i=1, i \neq m}^{r} |h_{W,g_k}^H c_{g,i}|^2 + \frac{1}{\alpha} \sum_{j=r+1}^{M} |h_{W,g_k}^H c_{g,j}|^2$, respectively, we can rewrite (46) as

$$\text{SINR}'_{g_k,m} = \frac{v}{\alpha P} + y.$$  

It follows from (59) that $v$ has the $\lambda_{g,m} \cdot \chi^2(2)$ distribution and that $y$ can be rewritten as $y = \sum_{i=1}^{M-1} |z_i|^2$, where the random variables $z_i$’s are independent over $i$ and zero-mean complex Gaussian distributed with variance

$$\sigma_i^2 = \begin{cases} \frac{\lambda_{g,n(i)}}{\alpha_2} & \text{if } 1 \leq i \leq r - 1 \\ \frac{\alpha_1}{\alpha_2} & \text{otherwise} \end{cases}$$

where the subscript $n(i)$ indicates the $i$th one of $\{\lambda_{g,n}, n \neq m\}$. Thus, the random variable $y$ has the generalized Chi-squared distribution with some repeated (common) variances among $\sigma_i^2$ whose pdf is given in the field of renewal theory [39] and later in communications [40]. Since the pdf with repeated variances involves quite lengthy notations, we will instead use a rather general formula with distinct variances for notational simplicity in the sequel. Thus, the pdf of $y$ is given by

$$f_Y(y) = \sum_{i=1}^{M-1} \frac{e^{-\frac{y}{\sigma_i^2}}}{\sigma_i^2 \prod_{j=1, j \neq i}^{M-1} \left( 1 - \frac{\sigma_j^2}{\sigma_i^2} \right)}, \text{ for } y \geq 0.$$  

(61)

By incorporating the intra-group interference and inter-group interference terms into the generalized Chi-squared distributed $y$, we will see in the next subsection that the both interference terms become arbitrarily small.
D. Asymptotic Throughput of GBF-SUP

Using the auxiliary results in the previous subsections, we present the final result.

**Theorem 5.** For $M$ fixed and $K$ sufficiently large with users satisfying the semi-unitary structure in (41) and (42), the throughput of GBF-SUP scales like

\[
M \log \log K + M \log \frac{P}{M} + \frac{1}{T} \sum_{t=1}^{T} \sum_{g=1}^{G_t} \left( \max \left( 0, M \log \log K^{(t)}_g \right) + \log |\Lambda^{(t)}_g| \right) + O(1),
\]

where $O(1) = M \log \alpha + o(1)$ with $\log \alpha \leq 0$.

**Proof:** See Appendix E.

Compared to the ideal case in Theorem 4, we have the penalty term $M \log \alpha$ in (122), where $\log \alpha \leq 0$. This implies that as $K$ increases, while the penalty factor $\alpha$ due to imperfect similarity still remains to affect the system throughput, the effect of another factor $\alpha_0$ vanishes. When $G = 1$ and all $R_{gk}$’s are the same, we can easily recover the deterministic beamforming results of [8], [26].

Letting $T = 1$ for convenience, we have assumed that every user follows the semi-unitary structure in Definition 3. Not relying on this assumption, consider the more general case where we have multiple classes. Suppose that $T$ classes are served in orthogonal time/frequency slots, with equal probability of scheduling and with $K^{(t)}_g = O(K)$ for all $t$ and $g$. We can then generalize the above result in the following form.

**Proposition 4.** For finite $M$ and large $K$, the capacity of correlated Rayleigh fading MIMO BCs is

\[
C^{\text{sum}} = M \log \frac{P}{M} + \frac{1}{T} \sum_{t=1}^{T} \sum_{g=1}^{G_t} \left( \max \left( 0, M \log \log K^{(t)}_g \right) + \log |\Lambda^{(t)}_g| \right) + O(1)
\]

where $T$ takes a finite value.

The proof is straightforward by using Theorem 5 and omitted here. This corollary is valid since using the SUP algorithm, we can partition users with arbitrary transmit correlation into a finite number of subsets (classes), each satisfying the semi-unitary structure with a proper choice of $\alpha$ and $\beta$.

Proposition 4, one of the main results in this paper, provides a means to characterize the capacity of general correlated fading MIMO BCs, which to the best of our knowledge has been open for the past decade, in the large $K$ regime. It further implies that the selection of $\{\Lambda^{(t)}_g, U^{(t)}_g\}$ in user partitioning is very relevant to well characterize the capacity. Such selection is a sort of quantization of transmit correlation matrices. A selection criterion of $U^{(t)}_g$ is discussed in the next section.

VII. SOME PRACTICAL CONSIDERATIONS

A. Measures of Similarity and Orthogonality

In Sec. III, we provided the two relevant properties, similarity and orthogonality, for a favorable structure of users transmit correlations that a user partition should construct for each class. In order to measure how well the user partition meets the properties, we investigate some metrics on them.

We first review standard distance metrics between subspaces in Grassmann manifold. In the special case where $r^{(t)}_g$ are the same for all $g$, i.e., subspaces spanned by the columns of $U_g$ in a class have the same dimension, our user partition problem is closely related to the well-known subspace packing on Grassmann manifold [15], [41]. The Grassmann manifold $G(M, m)$ is the set of all $m$-dimensional
subspaces in \( C^{M \times 1} \). Let \( U_{F_1} \) and \( U_{F_2} \) be the \( m \)-dimensional column spaces of \( F_1 \) and \( F_2 \) in \( \mathcal{G}(M, m) \), respectively. The best-known distance metric between subspaces \( U_{F_1} \) and \( U_{F_2} \) is the chordal distance

\[
d_c(F_1, F_2) = \frac{1}{\sqrt{2}} \| F_1 F_1^H - F_2 F_2^H \|_F
\]

\[= \sqrt{m - \| F_1^H F_2 \|_F^2}. \tag{65}\]

The Fubini-Study distance is also defined as

\[d_{FS}(F_1, F_2) = \arccos \| F_1^H F_2 \|. \tag{66}\]

Considering the user partition as a subspace packing problem in Grassmann manifold \( \mathcal{G}(M, m) \) with well-known distance metrics may allow us to use several bounding techniques in the literature (e.g., [42]). In the context of JSDM, however, we need distance metrics between two subspaces with different dimensions. To this end, let \( U_{F_2} \) be now the \( n \)-dimensional column space of \( F_2 \) in \( \mathcal{G}(M, n) \), where \( n \geq m \) without loss of generality. The above distance metrics in (65) and (66) can be generalized as follows. The generalized chordal distance between subspaces \( U_{F_1} \) and \( U_{F_2} \) is defined as

\[
d_{gc}(F_1, F_2) = \sqrt{\frac{m + n}{2}} - \| F_1^H F_2 \|_F^2 \tag{67}\]

which follows directly from (65). This satisfies the coincidence axiom of a metric, i.e., \( d_{gc}(F_1, F_2) = 0 \) if and only if \( F_1 = F_2 \). The generalized Fubini-Study distance is defined as

\[d_{gFS}(F_1, F_2) = \arccos \| F_1^H F_2 F_2^H F_1 \| \tag{68}\]

whose values range between 0 and \( \pi/2 \).

Given the above definitions, in order to measure a distance (i.e., orthogonality) between the subspace \( U_{t,g} \) and the orthogonal complement of \( \text{span}(\{U_h : h \neq g\}) \) with dimension \( r_g = M - \sum_{h \neq g} r_h \), we use the following generalized Fubini-Study distance

\[d_{gFS}(U_g, G_g) \tag{69}\]

where \( G_g \) is an \( M \times r_g \) matrix whose columns span the orthogonal complement to \( \text{span}(\{U_h : h \neq g\}) \). This metric can be used to get a good partition criterion to maximize sum rate. Let \( \delta_{gFS} \) be the maximum over all \( G \) generalized Fubini-Study distances, defined as

\[\delta_{gFS} = \min_g d_{gFS}(U_g, G_g). \tag{70}\]

B. Partition Criteria

The SUP algorithm in Sec. VI is one of user partition methods that select \( U_{g}^{(t)} \) based on an appropriate criterion and map each user to the best pair \( (t, g) \) for class and group association. In what follows, we investigate the impact of selecting \( U_g \) on the sum rate of JSDM and propose a criterion with which we should select \{\( U_g \)\} to maximize the sum rate. The resulting criterion is based on the generalized Fubini-Study distance in (69). In order to focus on finding a relevant selection criterion for \( U_g \), we assume the perfect similarity such that \( h_{gk} = U_g \Lambda_g^{1/2} w_{gk} \). Then, we can express the achievable sum rate of JSDM denoted by \( \mathcal{R}^{\text{sum}} \) as

\[\mathcal{R}^{\text{sum}} \geq \max_{\sum_{g=1}^G \text{tr}(S_g) \leq P} \sum_{g=1}^G \log \left| I_{r_g} + \Lambda_g^{1/2} W_g S_g W_g^H \Lambda_g^{1/2} \right| + G \log \cos(\delta_{gFS}). \tag{71}\]

The proof is given in Appendix F. Comparing (71) with (13), the second term \( G \log \cos(\delta_{gFS}) \) in (125) corresponds to a sum-rate offset from the sum capacity with the unitary structure.
In conclusion, in order to reduce the rate gap from the ideal unitary case and to maximize the sum rate of JSDM, we should select \( \{ U_g \} \) to minimize the maximum of all generalized Fubini-Study distances \( d_{gFS} \) between subspace \( U_g \) and its associated orthogonal complement to the subspace spanned by the columns of all other \( U_h, h \neq g \). This result is in line with [43] in the point-to-point MIMO case, where the classical Fubini-Study distances \( \| B \| \) was used to provide an insight into how to design a good codebook in limited feedback unitary precoding problems.

C. CSI Feedback

In the GBF scheme in Sec. VI-A, we let each user feed back the maximum of SINR in (43), assuming that the BS schedules and transmits all the \( M \) beams and that all users have the equal transmit power \( P/M \). These assumptions do not make much sense in realistic systems, unless \( K \gg M \), which is even more infeasible for \( M \) large. Therefore, we should enable the scheduler to arbitrarily select 1-to-\( M \) tuple user subsets with appropriate CSI feedback. However, reporting the corresponding \( \sum_{i=1}^{M} C_i^M \) SINRs is impossible for \( M \) large, since it may consume most of uplink resources. To overcome this difficulty, we propose a new CSI feedback in the sequel.

Assume \( \sum_{g} r_g = M \) and, for notational convenience, the equal power allocation over multiple beams to transmit. Given \( B_g \subseteq [1 : r_g] \) for all \( g \), where \( B_g \) is a subset of the indices of \( r_g \) beams \( b_{g,i} \), we have \( |B_g| \) SINR

\[
\text{SINR}_{g,k,m} = \frac{|h_{g_k}^H b_{g,m}|^2}{\frac{\sigma_z^2}{P} + \sum_{i \in B_g \setminus m} |h_{g_k}^H b_{g,i}|^2 + \sum_{h \neq g} \sum_{j \in B_h} |h_{g_k}^H b_{h,j}|^2}, \quad m \in B_g
\]  

(72)

where \( \sigma_z^2 \) is the variance of Gaussian noise at user \( g_k \), which takes into account inter-cell interference in multi-cell scenarios.

For each beam \( b_{h,j} \), where \( h = 1, \cdots, G \) and \( j = 1, \cdots, r_h \), we define the corresponding interference and noise ratio (INR) as

\[
\text{INR}_{g_k,h,j} = \frac{|h_{g_k}^H b_{h,j}|^2}{\sigma_z^2}.
\]

(73)

If the above INRs are available at the BS, the scheduler selects one of INRs as the numerator of SINR in (72) and can calculate the corresponding all possible SINRs for user \( g_k \) with any \( B_g \), namely

\[
\text{SINR}_{g_k,m} = \frac{\text{INR}_{g_k,g,m}}{\frac{\sigma_z^2}{P} + \sum_{i \in B_g \setminus m} \text{INR}_{g_k,g,i} + \sum_{h \neq g} \sum_{j \in B_h} \text{INR}_{g_k,h,j}}, \quad m \in B_g
\]

which enables fully flexible scheduling over all 1-to-\( M \) tuple user subsets. Furthermore, unlike RBF, beam indices need not be additionally reported since INRs for all \( M \) beams are fed back. Finally, if a beam \( b_{g,m} \) is selected at user \( g_k \) and is reported, we may reduce the feedback overhead for INRs corresponding to inter-group interference in (72) by letting users report a single average value of \( r_h \) INRs of group \( h \) for all \( h \neq g \) at the cost of marginal performance loss, since \( r_h \) beams within a group have a similar directionality by user partitioning unless \( G \) is too small. Therefore, a user need to report at most \( (r_g + G - 1) \) real values of INR instead of \( M \) ones.

D. Dedicated Pilot Overhead

While training for the FDD downlink consists of downlink common and dedicated pilots, the TDD downlink requires uplink user-specific pilot and downlink dedicated pilot\(^9\). While downlink common and uplink user-specific pilots are used for acquiring CSIT in FDD and TDD, respectively, dedicated pilot is required for CSI at the receiver (CSIR) and demodulation.

\(^9\)In the long-term evolution (LTE) context [32], the downlink common, dedicated, and the uplink user-specific pilots correspond to CSI reference signal (CSI-RS), demodulation reference signal (DM-RS), and sounding reference signal (SRS), respectively.
Practical systems require in general much more resources for dedicated pilots than common pilot to average out channel estimation errors due to noise and interference\textsuperscript{[10]} As a result, the overhead for dedicated pilot is indeed a much more prominent limiting factor in realistic large-scale MIMO systems than common pilot. This is also the case with TDD unless $M$ is large enough to invoke the law of large number for the so-called favorable propagation property [17]. Moreover, the required number of transmit antennas gets much larger in highly correlated fading cases, where, due to the channel dimension reduction effect, the effective dimension of channel vectors reduces depending on the number of dominant eigenvalues of transmit correlation matrices as mentioned in Sec. [11].

At this point, note that the INR feedback gives us a side information as to mutual interference between users. The second but more important role of the proposed CSI feedback in (73) is that it can reduce the dedicated pilot overhead. Depending on the INR feedback, the scheduler can figure out how much the intended user $g_k$ is interfered with by the other user receiving the beam $b_{h,j}$. So, the scheduler can allocate a common resource for dedicated pilot to multiple users having sufficiently low interfering beams between them. In this way, the INR feedback allows us to achieve a large portion of the potential gain due to the pilot overhead reduction in Sec. [V] even in practice, where the ideal unitary structure is seldom attainable. Note that transmit correlation diversity enables the significant saving of dedicated pilot due to INR feedback.

In sum, the INR feedback provides not only flexible scheduling but it can also significantly reduce the dedicated pilot overhead, which increases the system multiplexing gain.

As for reducing the common pilot overhead, we may adopt multiple precoded common pilots which share a common frequency/time resource. This seems feasible in practice because the SUP algorithm can considerably reduce the number of dominant interferences between multiple common pilots and because, unlike data streams or even dedicated pilot, the remaining dominant interference is much easier to suppress with advanced receiver algorithms like successive interference cancellation (SIC) or iterative receiver.

\section*{E. Alternative to Transmit Correlation Matrix Estimation}

So far, the JSDM strategy has assumed either that the BS perfectly knows transmit correlation matrices $R_i$ (or $U_i$) of all users $i$ or that each user perfectly estimates $R_i$ as in GBF. In practice, this may be still costly and burdensome when $M$ is large or user mobility is high.

As a simple heuristic approach to avoid the difficulty, we may replace (40) with using $\arg\max_{t,g} \sum_{n=1}^{N} \| h_i^H(n) B_g^{(t)} \|_2$ in an average sense without estimating $R_i$. We can easily see that for large $N$

$$\arg\max_{t,g} \sum_{n=1}^{N} \| h_i^H(n) B_g^{(t)} \|_2 = \arg\max_{t,g} \sum_{n=1}^{N} \| w_i^H(n) \Lambda_i^{1/2} U_i^H B_g^{(t)} \|_2$$

$$\rightarrow \arg\max_{t,g} \| \Lambda_i^{1/2} U_i^H B_g^{(t)} \|_F$$

$$= \arg\max_{t,g} \| R_i^{1/2} B_g^{(t)} \|_F$$

(74)

since $w_i$ is mean-ergodic and distributed as $\mathcal{CN}(0, I)$ for all $i$. Therefore, instead of estimating and reporting the large dimensional $R_i$, each user may need only report class and group indices $(t, g)$ in a long-term time/frequency period.

\section*{F. Low-Complexity Scheduling}

By the notion of virtual sectorization approximated by the SUP algorithm, the inter-group interference term $\sum_{h=1, h \neq g}^{G} \sum_{j=1}^{r_g} | h_{g_k}^H b_{h,j} |^2$ in (43) is likely to be smaller than the intra-group interference term $\sum_{h=1, h \neq g}^{G} \sum_{j=1}^{r_g} | h_{g_k}^H b_{h,j} |^2$ in (43) is likely to be smaller than the intra-group interference term

\textsuperscript{10}In LTE, for $M = 4$, CSI-RS consumes at most 4 frequency/time resource elements (REs) every 5 time slots (5 ms). In contrast, DM-RS requires 24 REs per slot just for $s = 2$. CSI-RS (common pilot) is sparse in time/frequency slot.
Thus we schedule users independently among groups as we can treat each group as a virtual sector.

When users have unequal received SNRs, we should consider some fairness criteria over classes, groups within a class, and users within a group. The problem of interest is then turn to a complicate weighted sum-rate maximization. The scheduling of JSDM can be done in three stages with low complexity, instead of scheduling (searching) over the entire set $\mathcal{K}$. The first stage scheduling is done over users within a particular group, independently with the other groups due to the virtual sectorization. The next stage is over groups within a class and the last one is over classes. This multi-stage scheduling can significantly reduce the computational complexity especially for both $M$ and $K$ large.

We can compare GBF-SUP with RBF in terms of scheduling complexity. RBF assumes an exhaustive search over the entire user set, yielding the computational complexity of $C^K_s$ (the size of its search space) under the assumption of transmitting always $M$ beams. In contrast, GBF-SUP has a much lower complexity of $\sum_t \sum_g C^K_{s(t)}$ when $K$ is large. For example, when $M = 32, s = 32, K = 128$ and, for GBF-SUP, $T = 2, G = 8$ (yielding $K' = 8, s/G = 4$ for convenience), RBF has the size of $C^{128}_{32} \approx 1.5 \times 10^{10}$ but GBF-SUP needs only $2 \times 8 \times C^8_4 = 1120$. Therefore, GBF-SUP can dramatically reduce the scheduling complexity for large $K$ and $M$.

VIII. Numerical Results

A. One-Ring Channel Model

In the typical cellular downlink case, the BS is elevated and free of local scatterers, and the user terminals are placed at ground level and are surrounded by local scatterers. Thus, the channel in the form of (4) may reduce to the one-ring model [3], for which we have $\mathbf{H} = I$. For the one-ring model, a user located at azimuth angle $\theta$ and distance $s$ is surrounded by a ring of scatterers of radius $r$ such that $\Delta \approx \arctan(r/s)$. Assuming the ULA with a uniform distribution of the received power from planar waves impinging on the BS array, we have that the correlation coefficient between BS antennas $1 \leq p, q \leq M$ is given by

$$[\mathbf{R}]_{p,q} = \frac{1}{2\Delta} \int_{-\Delta}^{\Delta} e^{j2\pi D(p-q)\sin(\alpha+\theta)} d\alpha$$

where $D$ is the normalized distance between antenna elements by the wavelength.

B. System Throughput in the Large $K$ Regime

Comprehensive sum-rate analysis and numerical results of JSDM in the large system ($M$) regime was already given in [20]. We will present here the achievable throughput of GBF-SUP and show how well Proposition 4 characterizes the sum capacity in a correlated fading MIMO BC, where the BS has $M = 4$ and ULA with $D = 1/2$, and users are uniformly distributed over the range $[-60^\circ, 60^\circ]$ with $\Delta_i$ uniformly distributed in the range $[5^\circ, 15^\circ]$, where $\Delta_i$ is AS of user $i$. The users transmit correlation matrices are generated by the one-ring channel model (75). Let $T = 2, G = 4, \Delta_g = 10^\circ$ for all $g$, where $\Delta_g$ is AS of group $g$, and also let two AoD sets for each class be $\{-57.5^\circ, -23^\circ, 7.5^\circ, 41.5^\circ\}$ and $\{-41.5^\circ, -7.5^\circ, 23^\circ, 57.5^\circ\}$, respectively. This choice of AoDs was done such that given $\Delta_g$, we maximize the minimum chordal distance between any two of the set $\{U_{g}^{(t)}\}$ for each class. Then, we can generate $\mathbf{R}_g^{(t)}$ for all $(t, g)$ according to (75).

Fig. 6 compares the sum rates of DPC with greedy user selection [44] (denoted by capacity), ZFBF-SUS [23], RBF, GBF-SUP, GBF-SUP-INR for $M = 4$, where GBF-SUP-INR denotes GBF-SUP with INR feedback. While DPC and ZFBF-SUS assume the perfect CSIT, GBF-SUP needs the SINR feedback and 3-bit long-term CSI feedback indicating one of 8 groups, requiring no instantaneous feedback on beam index in this case. The INR feedback consists of 4 INRs instead of a single SINR.
It has been well observed that the throughput of RBF is quite smaller than ZFBF with perfect CSIT \cite{23} and even with imperfect CSIT (e.g., \cite{45}). One main reason is that RBF always transmits $M$ orthogonal beams, even though no users have a good match with a fraction of $M$ random beam vectors when $K$ is small. In contrast, the SUP algorithm leaves some groups empty to which no one has a similar transmit correlation and only nonempty groups are scheduled for each class. Therefore, GBF-SUP does not always transmit all the $M$ beams. This is also the case with the SUS algorithm, which schedules $M$ users as long as there are at least $M$ semi-orthogonal users. In fact, the SUP algorithm plays a role of user selection, but it depends only on long-term channel statistics rather than instantaneous channel realizations. This explains why GBF-SUP outperforms RBF, compared to ZFBF-SUS, for $M = 4$ with $G = 4$ and $r = 1$, in which case one may consider GBF-SUP analogous to RBF. Furthermore, GBF approximates the eigenbeamformer for users in a certain group.

We can also see that the INR feedback improves the throughput of GBF-SUP due to fully flexible scheduling especially when $K$ is small. The flexible scheduling gain is readily expected to become larger as $M$ increases, at the cost of a moderate increase in CSI feedback. It seems surprising that at SNR = 10 dB, GBF-SUP-INR shows a competitive performance when $K$ is small and even noticeably outperforms ZFBF-SUS as $K$ grows. This is in sharp contrast to RBF which requires a large number of users, $K \gg M$, but whose individual rate per user decays as $O\left(\frac{\log \log K}{K}\right)$, the proposed scheme is shown to get a good system throughput even for not-so-large $K$. As $M$ or SNR increases, however, GBF-SUP and GBF-SUP-INR are shown to require gradually much larger $K$ to approach the performance of ZFBF-SUS, as shown in \cite{30}.

The asymptotic sum rate scaling law in Proposition \ref{prop:sum_rate} is denoted by large $K$ analysis in Fig. \ref{fig:sum_rate}. The gap between the sum rate scaling \eqref{eq:sum_rate} and the sum capacity is shown to be not large, unless $K$ is too small. Also recall that the approximate capacity in Fig. \ref{fig:capacity} was given by a greedy user selection due to computational complexity so that the exact capacity would be somehow larger. Therefore, with the above choice of $R_{(t)}^g$, the asymptotic result seems to well characterize the sum capacity of the correlated Rayleigh fading MIMO BC.

IX. CONCLUDING REMARKS

In this paper, we investigated the impact of transmit correlation and the fundamental limits on the capacity of correlated fading MIMO BCs. To clearly show the potential gains of transmit correlation diversity, we used the ideal unitary structure across users. Under this assumption, we found that multiplexing gain can continue growing as the number of antennas and the number of users increase, as long as transmit correlations among users are sufficiently high. We observed that the fundamental limits and asymptotic behaviors of correlated fading MIMO BCs are not upper-bounded any longer by those of MIMO point-to-point channels. Furthermore, we compactly characterized the capacity of the ideal correlated fading BCs in various regimes of interest and also the capacity of general correlated fading BCs in the large number of users regime. Not only presenting theoretical results, we also provided a practical JSDM scheme and some remarks to realize the potential gain of transmit correlation diversity in realistic systems.

Transmit correlation has been traditionally considered as a harmful effect for MIMO wireless networks, but it can be indeed used as a highly beneficial resource. The most remarkable result may be summarized as: exploiting transmit correlation can increase the multiplexing gain of MU-MIMO systems by a factor of the degrees of transmit correlation diversity.

This article ends up with presenting some interesting topics for further study.

- **Applicability to various MIMO networks**: The notion of transmit correlation diversity can be used in various forms of MIMO wireless networks including MIMO multiple-access channels (MACs), multi-cell MU-MIMO systems, and wireless interference networks.
- **Distributed antennas systems**: In the typical distributed antennas at the transmitter, we can exploit receive correlation diversity. A single receiver with multiple antennas may have different receive
correlations on the multiple links between each distributed transmit array and the receive array. In fact, this reduces to MIMO MACs, but not uplink systems. With distributed antennas at the receiver, we can use transmit correlation diversity even in a single receiver. This may be viewed as a sort of partial cooperation in Sec. IV-D. The recent research topic on small cells with wireless backhaul is the case, if a small cell has distributed antennas. This is neither a typical point-to-point MIMO nor multiuser communications. Thus, fundamental limits on MIMO wireless networks can be refined in the distributed antenna context.

- **High-frequency channels:** Another interesting application of JSDM is mm-Wave channels. In [46], the authors already demonstrated the applicability of JSDM in actual measured channels.

### APPENDIX A

**USEFUL LEMMAS**

We collect here some useful lemmas which are needed in various proofs in this work.

**Lemma 1.** A central Wishart matrix $WW^H$ with $W$ the $m \times n$ matrix, where $n \geq m$, satisfies [47]

$$
\mathbb{E} \left[ \ln |WW^H| \right] = \sum_{\ell=0}^{m-1} \psi(n - \ell)
$$

(76)

where

$$
\psi(n) = -\gamma + \sum_{\ell=1}^{n-1} \frac{1}{\ell}
$$

(77)
is the Euler’s digamma function with $\gamma \approx 0.5572$ the Euler-Mascheroni constant. With the ratio $\eta = \frac{n}{m}$, for $m$ large, the random matrix $WW^H$ also shows the asymptotic behavior
\[
\frac{1}{m} \mathbb{E} \left[ \ln |WW^H| \right] = (\eta - 1) \ln \frac{\eta}{\eta - 1} + \ln n - 1 + O(m^{-1}).
\] (78)

The proof of (78) can be immediately given by applying
\[
\frac{1}{k} \sum_{\ell=1}^{k} \psi(\ell) = \psi(k + 1) - 1.
\] (79)

and by using the fact that $\psi(k)$ behaves as
\[
\lim_{k \to \infty} \psi(k) = \ln k + O(k^{-1})
\] (80)
due to
\[
\lim_{k \to \infty} \sum_{n=1}^{k} \frac{1}{n} - \ln k = \gamma.
\]

The following lemma is derived from the well-known Weyl theorem [48].

**Lemma 2** (49). Let $A$ and $B$ be $n \times n$ matrices with $A$ Hermitian. For $k \in [1 : \pi(A)]$
\[
\max_{1 \leq i \leq n} \left\{ \lambda_i(A)\lambda_{n+k-i}(BB^H) \right\} \leq \lambda_k(BAB^H) \leq \min_{1 \leq i \leq k} \left\{ \lambda_i(A)\lambda_{k+1-i}(BB^H) \right\}
\] (81)
where $\pi(A)$ is the number of positive eigenvalues of $A$.

We use the following well-known result from majorization theory.

**Lemma 3** (50). Let $A$ and $B$ be $n \times n$ positive semi-definite Hermitian matrices. Then
\[
\text{tr}(AB) \leq \sum_{i=1}^{n} \lambda_i(A)\lambda_i(B).
\] (82)

We need the following lemma to derive Lemma 5.

**Lemma 4** (Schur complement lemma). Let $X$ be a Hermitian matrix partitioned as
\[
X = \begin{bmatrix} A & B \\ B^H & C \end{bmatrix}
\]
where $A$ and $C$ are Hermitian. The Schur complement of $C$ in $X$ is defined as $S = A - BC^{-1}B^H$. Then, we have the following equivalences:
\[
X \succ 0 \iff C \succ 0, \ S \succ 0.
\] (83)

If $C \succ 0$, then
\[
X \succeq 0 \iff S \succeq 0.
\] (84)

If $C$ is singular, then
\[
X \succeq 0 \iff C \succeq 0, \ (I - CC^H)B = 0, \ A - BC^HB^H \succeq 0.
\] (85)

This lemma is an alternative form of the lemma in [51], where the Schur complement of $A$ in $X$ is used instead of $S$.

**Lemma 5.** Suppose that $X$ and $Y$ are Hermitian matrices of the same size and that $Y$ is positive semi-definite and decomposed as
\[
Y = \begin{bmatrix} Q_1 & Q_1^+ \end{bmatrix} \begin{bmatrix} \Sigma_Y & 0 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} Q_1 & Q_1^+ \end{bmatrix}^H
\]
where $Q_1$ consists of eigenvectors that span the subspace $\text{range}(Y)$, the columns of $Q_1^\perp$ span $\text{ker}(Y)$, and $\Sigma_Y$ is a diagonal matrix whose diagonal elements are the non-zero eigenvalues of $Y$. Let $P$ and $P^\perp$ consist of orthogonal bases of $\text{range}(Q_1^\perp X Q_1^\perp)$ and $\text{ker}(Q_1^\perp X Q_1^\perp)$, respectively. Also, let $Q = [Q_1, Q_2, Q_3]$, where $Q_2 = Q_1 P$ and $Q_3 = Q_1 P^\perp$, and $X_{ij} = Q_i^H X Q_j$, where $i, j = 1, 2$. Then the following properties are equivalent:

\[
X \succeq zY \iff \{X_{22} \succeq 0, \quad (I - X_{22} X_{22}^\perp) X_{12} = 0, \quad z \leq \lambda_{\min}(\Sigma_Y^{-1/2} (X_{11} - X_{12} X_{22}^\perp X_{12}^H) \Sigma_Y^{-1/2})\}. \tag{86}
\]

In particular, if $X_{22} = 0$, the right-hand side of (86) reduces to

\[
z \leq \lambda_{\min}(\Sigma_Y^{-1/2} X_{11} \Sigma_Y^{-1/2}). \tag{87}
\]

Also if $X_{22} \succ 0$, (86) reduces to

\[
z \leq \lambda_{\min}(\Sigma_Y^{-1/2} (X_{11} - X_{12} X_{22}^{-1} X_{12}^H) \Sigma_Y^{-1/2}). \tag{88}
\]

**Proof:** With the matrix $Q$ defined above, we have

\[
Q^H X Q = \begin{bmatrix}
X_{11} & X_{12} & 0 \\
X_{12}^H & X_{22} & 0 \\
0 & 0 & 0
\end{bmatrix}.
\]

Since $Q$ is non-singular and it represents an orthogonal change of coordinates, we can see that $X \succeq zY$ is equivalent to

\[
\begin{bmatrix}
X_{11} & X_{12} \\
X_{12}^H & X_{22}
\end{bmatrix} \succeq \begin{bmatrix}
\Sigma_Y & 0 \\
0 & 0
\end{bmatrix}.
\tag{89}
\]

This implies $X_{22}$ must be positive semi-definite.

As a special case, if $X_{22} = 0$ (i.e., $\text{range}(X) \cap \text{ker}(Y) = 0$), (89) reduces to $X_{11} \succeq z \Sigma_Y$, which is equivalent to

\[
z I \preceq \Sigma_Y^{-1/2} X_{11} \Sigma_Y^{-1/2}.
\]

If $X_{22} \succ 0$, applying the non-singular $\begin{bmatrix}
\Sigma_Y^{-1/2} & 0 \\
0 & I
\end{bmatrix}$ to both sides of (89) leads to

\[
\begin{bmatrix}
\Sigma_Y^{-1/2} X_{11} \Sigma_Y^{-1/2} & \Sigma_Y^{-1/2} X_{12} \\
X_{12}^H \Sigma_Y^{-1/2} & X_{22}
\end{bmatrix} \succeq \begin{bmatrix}
I & 0 \\
0 & 0
\end{bmatrix}
\iff\begin{bmatrix}
\Sigma_Y^{-1/2} X_{11} \Sigma_Y^{-1/2} - zI & \Sigma_Y^{-1/2} X_{12} \\
X_{12}^H \Sigma_Y^{-1/2} & X_{22}
\end{bmatrix} \succeq 0 \tag{90}
\iff (a)\quad z I \preceq \Sigma_Y^{-1/2} (X_{11} - X_{12} X_{22}^{-1} X_{12}^H) \Sigma_Y^{-1/2}.
\]

In $(a)$ we used (84) in Lemma 4.

Finally, if $X_{22}$ is singular, it immediately follows from (85) in Lemma 4 that (90) becomes equivalent to

\[
X_{22} \succeq 0, \quad (I - X_{22} X_{22}^\perp) X_{12} = 0, \quad z I \preceq \Sigma_Y^{-1/2} (X_{11} - X_{12} X_{22}^\perp X_{12}^H) \Sigma_Y^{-1/2}.
\]

This completes the proof. ■
Lemma 6 ([52]). Let \( A \) and \( B \) be Hermitian matrices with eigenvalues \( \lambda_1(A) \geq \lambda_2(A) \geq \cdots \geq \lambda_n(A) \) and \( \lambda_1(B) \geq \lambda_2(B) \geq \cdots \geq \lambda_n(B) \), respectively. Then
\[
\min_{\pi} \prod_{i=1}^{n} (\lambda_i(A) + \lambda_{\pi(i)}(B)) \leq |A + B| \leq \max_{\pi} \prod_{i=1}^{n} (\lambda_i(A) + \lambda_{\pi(i)}(B)) \tag{91}
\]
where \( \pi \) denotes a permutation of indices 1, 2, \cdots, \( n \). In particular, if \( \lambda_n(A) + \lambda_n(B) \geq 0 \), then
\[
\prod_{i=1}^{n} (\lambda_i(A) + \lambda_i(B)) \leq |A + B| \leq \prod_{i=1}^{n} (\lambda_i(A) + \lambda_{n-i+1}(B)). \tag{92}
\]

**Appendix B**

**Proof of Theorem 2**

We first prove the case of \( K_g > r_g \). Provided the unitary structure is available, the sum rate of the \( g \)th dual MAC subchannel in (13) can be rewritten as
\[
\log |I + \Lambda_g^{1/2} W_g S_g W_g^H \Lambda_g^{1/2}| = \log |\Lambda_g^{-1} + W_g S_g W_g^H| + \log |\Lambda_g|. \tag{93}
\]
By allowing the receiver cooperation within each group, the capacity region of the dual MAC subchannel is outer-bounded by that of the corresponding cooperative MIMO system. Given the perfect CSIT and at high SNR, the asymptotic optimal input \( \mathbf{X}_g \) in the cooperative MIMO system is the uniform power allocation over \( r_g \) eigenmodes of \( W_g W_g^H \) with \( \sum_{g} \text{tr}(\mathbf{X}_g) \leq P \), since the Wishart matrix \( W_g W_g^H \) is well conditioned with high probability for all \( g \). Then, we have at high SNR (i.e., \( P \))
\[
\log |\Lambda_g^{-1} + W_g S_g W_g^H| \leq \log |\Lambda_g^{-1} + W_g X_g W_g^H|
\simeq \log |\Lambda_g^{-1} + \frac{P}{P} W_g W_g^H|
\simeq \log |W_g W_g^H| + r_g \log \frac{P}{P} \tag{94}
\]
where \( \simeq \) denotes the asymptotic equivalence and we used the fact that \( \lambda_{g,i}^{-1} < \infty \) for all \( i \) since \( \lambda_{g,i} \) are dominant eigenvalues as defined in [5]. As a consequence, when \( K_g > r_g \), the sum capacity is upper-bounded as
\[
C^{\text{sum}}(P) \leq M \log \frac{P}{P} + \sum_{g=1}^{G} \mathbb{E} \left[ \log |W_g W_g^H| \right] + \sum_{g=1}^{G} \log |\Lambda_g| + o(1) \tag{95}
\]
\[
= M \log \frac{P}{P} + \log e \sum_{g=1}^{G} \left( K_g \psi(K_g + 1) - (K_g - r_g) \psi(K_g - r_g + 1) - r_g \right) + \sum_{g=1}^{G} \log |\Lambda_g| + o(1)
\]
\[
= M \log \frac{P}{P} + \log e \sum_{g=1}^{G} r_g \left( -\gamma + \sum_{i=2}^{K_g} \frac{1}{i} + \frac{K_g - r_g}{r_g} \sum_{\ell=K_g-r_g+1}^{K_g} \frac{1}{\ell} \right) + \sum_{g=1}^{G} \log |\Lambda_g| + o(1) \tag{96}
\]
where \( o(1) \) goes to zero as \( P \to \infty \), and we used the well-known result of random matrix theory [47] in Lemma 1 of Appendix A namely
\[
\mathbb{E} \left[ \ln |W_g W_g^H| \right] = \sum_{\ell=0}^{r_g-1} \psi(K_g - \ell) \tag{97}
\]
where \( W_g W_g^H \) is almost surely nonsingular and \( \psi(n) \) is defined in [77]. From (79), the second equality in (96) immediately follows.
The achievability of (14) is given by simply letting the diagonal input matrix $S_g$ as

$$S_g = \begin{bmatrix} P \frac{I_{r_g}}{M} & 0 \\ 0 & 0_{K_g-r_g} \end{bmatrix}.$$  

The resulting $W_gS_gW_{g}^H = \frac{P}{M} \tilde{W}_g \tilde{W}_g^H$ is also a Wishart matrix with $r_g$ degrees of freedom and hence

$$\log |\Lambda_g^{-1} + W_gS_gW_{g}^H| \geq \log |\tilde{W}_g \tilde{W}_g^H| + r_g \log \frac{P}{M} + o(1) \quad (98)$$

Using (76) in Lemma 1 we have

$$C^{\text{sum}}(P) \geq M \log \frac{P}{M} + \log e \sum_{g=1}^{G} r_g \left(-\gamma + \sum_{\ell=2}^{r_g} \frac{1}{\ell} \right) \geq \sum_{g=1}^{G} \log |\Lambda_g| + o(1). \quad (99)$$

Then, we see that the achievable rate approaches the upper bound in (96) within $O(1)$ being $- \log e \sum_{g=1}^{G} \left( \sum_{\ell=r_g+1}^{K_g} \frac{1}{\ell} + \frac{K_g - r_g}{r_g} \sum_{\ell=K_g-r_g+1}^{K_g} \frac{1}{\ell} \right) + o(1) \leq O(1) \leq o(1)$, which yields (14).

Next, we consider the second case of $K_g \leq r_g$. When the number of transmit antennas is greater than or equal to the total number of receive antennas in an MIMO BC, the sum capacity of the dual MAC is well known [24] to be equivalent at high SNR to that of the corresponding cooperative MIMO system. This also implies that, for $K_g \leq r_g$, uniform power allocation across $K_g$ eigenmodes in the $g$th dual MAC (13) is asymptotically optimal, yielding $S_g = \frac{P}{K} I_{K_g}$ for all $g$. Therefore, for sufficiently large $P$, we have the upper bound

$$\log |\Lambda_g^{-1} + W_gS_gW_{g}^H| \simeq \log |\Lambda_g^{-1} + \frac{P}{K} W_g W_{g}^H| \leq \log \prod_{i=1}^{r_g} \left( \frac{\lambda^{-1}_{g,i} + \frac{P}{K} \lambda_i(W_g W_{g}^H)}{r_g} \right) = \log \prod_{i=1}^{K_g} \left( \frac{\lambda^{-1}_{g,i} + \frac{P}{K} \lambda_i(W_g W_{g}^H)}{r_g} \right) + \log \prod_{i=K_g+1}^{r_g} \lambda^{-1}_{g,i} \simeq \log |W_g W_{g}^H| + r_g \log \frac{P}{K} + \log \prod_{i=K_g+1}^{r_g} \lambda^{-1}_{g,i} \quad (100)$$

where $(a)$ follows from Lemma 2 in Appendix A. Using Lemma 2 again, we can get the lower bound

$$\log |\Lambda_g^{-1} + W_gS_gW_{g}^H| \geq \log |W_g W_{g}^H| + \log \prod_{i=K_g+1}^{r_g} \lambda^{-1}_{g,i} \quad (101)$$

Plugging (100) and (101) into (93), using the fact that for $K_g \leq r_g$, the Wishart matrix $W_g W_{g}^H$ is almost surely nonsingular, and invoking (76) again, we have

$$K \log \frac{P}{K} + \log e \sum_{g=1}^{G} K_g \left(-\gamma + \sum_{\ell=2}^{r_g} \frac{1}{\ell} + \frac{r_g - K_g}{K_g} \sum_{\ell=K_g-r_g+1}^{K_g} \frac{1}{\ell} \right) + \sum_{g=1}^{G} \log \prod_{i=1}^{K_g} \lambda_{g,i} + o(1) \leq C^{\text{sum}}(P) \leq K \log \frac{P}{K} + \log e \sum_{g=1}^{G} K_g \left(-\gamma + \sum_{\ell=2}^{r_g} \frac{1}{\ell} + \frac{r_g - K_g}{K_g} \sum_{\ell=K_g-r_g+1}^{K_g} \frac{1}{\ell} \right) + \sum_{g=1}^{G} \log \prod_{i=1}^{K_g} \lambda_{g,r_g-i+1} + O(1). \quad (102)$$

With $O(1)$ being $\sum_{g=1}^{G} \sum_{i=1}^{K_g} \log \frac{\lambda g_{g,r_g-i+1}}{\lambda g_{g,i}} + o(1) \leq O(1) \leq o(1)$, we have (15). This concludes the proof.
Beside the above proof, an alternative expression of (15) can be found as follows. For sufficiently large \( P \), we have
\[
\log |I + \Lambda_1^{1/2}W_g S_g W_g^H| = \log \left| I + \frac{P}{M} \Lambda_1^{1/2}W_g W_g^H \right| \\
= \log \left| I + \frac{P}{M} W_g^H \Lambda_g W_g \right| \\
\simeq \log |W_g^H \Lambda_g W_g| + r_g \log \frac{P}{M}.
\]
(103)

Using Lemma 2 in [7] (See also [27]), we can get
\[
\mathbb{E} \left[ \log |W_g^H \Lambda_g W_g| \right] = \log e \sum_{g=1}^{G} \frac{\Upsilon_g}{|\Omega_g|} \sum_{k=1}^{K_g} |\Psi_{g,k}| \tag{104}
\]
where \( \Psi_{g,k} \) is an \( K_g \times K_g \) matrix whose \((i, j)\) element is
\[
(\Psi_{g,k})_{i,j} = \nu_{r_g - K_g + i} r_g - K_g - 1 + j - \sum_{d=1, q=1}^{r_g - K_g} \nu_q (\Upsilon_g^{-1})_{d,q} \lambda_d^{-1} \lambda_g, r_g - K_g + i \lambda_q, r_g - K_g - 1 + j.
\]
(105)

where \( \nu_q = \psi(\ell) + \ln \lambda_q \) for \( \ell = k \); otherwise, \( \nu_q = 1 \), \( \Omega_g \) is the Vandermonde matrix
\[
\Omega_g = \begin{bmatrix}
1 & \lambda_{g,1} & \cdots & \lambda_{g,r_g - 1} \\
1 & \lambda_{g,2} & \cdots & \lambda_{g,r_g - 1} \\
\vdots & \vdots & \ddots & \vdots \\
1 & \lambda_{g,r_g} & \cdots & \lambda_{g,r_g}
\end{bmatrix}
\]
(106)

and \( \Upsilon_g \) is the \((r_g - K_g) \times (r_g - K_g)\) principle submatrix of \( \Omega_g \). This yields
\[
C_{\text{sum}}(P) = M \log \frac{P}{M} + \log e \sum_{g=1}^{G} \frac{\Upsilon_g}{|\Omega_g|} \sum_{k=1}^{K_g} |\Psi_{g,k}| + o(1).
\]
(107)

**APPENDIX C**

**PROOF OF THEOREM 3**

The proof begins with the dual MAC in (13)
\[
\sum_{g=1}^{G} \log \left| I + \Lambda_1^{1/2}W_g S_g W_g^H \right| = \sum_{g=1}^{G} \log |\Lambda_1^{-1} + W_g S_g W_g^H| + G \log |\Lambda|
\]
(108)

where the equality is given by (95) and the assumptions. Taking expectation on the first term in the right-hand side of (108) and dividing by \( M \), for \( \mu < 1 \) (equivalently, \( R' > r \)), at high SNR (\( P \)), we have
the upper bound
\[
\frac{1}{M} \mathbb{E} \left[ \sum_{g=1}^{G} \log |\Lambda_1^{-1} + W_g S_g W_g^H| \right] \overset{(a)}{\leq} \frac{1}{r} \mathbb{E} \left[ \log |W_g W_g^H| \right] + \log \frac{P}{M}
\]
\[
\overset{(b)}{=} \log e \left\{ (\mu^{-1} - 1) \ln \frac{\mu^{-1}}{\mu^{-1} - 1} + \ln K' - 1 + O(r^{-1}) \right\} + \log \frac{P}{M}
\]
\[
= \log \frac{P}{e\mu G} + \left( \frac{1 - \mu}{\mu} \right) \log \frac{1}{1 - \mu} + O(r^{-1})
\]
(109)
where (a) follows from (94) and (b) follows from (78) in Lemma 1. From (98) and (78), we also get the lower bound
\[
\frac{1}{M} \mathbb{E} \left[ \sum_{g=1}^{G} \log |\Lambda^{-1} + W_g S_g W_g^H| \right] \geq \log \frac{P}{eG} + O(r^{-1}). \tag{110}
\]
Noticing
\[
\frac{1}{r} \log |\Lambda| \leq \frac{1}{r} \log \left( \frac{M}{r} \right)^r = \log G \tag{111}
\]
due to \( \text{tr}(R_g) = \text{tr}(\Lambda) = M \) and using the fact that \( \lambda_i(\Lambda) \) are dominant eigenvalues defined in (5), we have
\[
\frac{1}{r} \log |\Lambda| = \log G + c
\]
where \( c \leq 0 \) is a constant which does not depend on both \( r \) and \( G \). Substituting (109) and (110) to (108) (divided by \( M \)) yields (16) with \( O(1) \) being \( \log \mu - \left( \frac{1}{\mu} \right) \log \frac{1}{1-\mu} + c + o(1) \). For \( \mu \geq 1 \) (equivalently, \( K' \leq r \)), similar to the above steps with
\[
\frac{1}{K'} \sum_{i=1}^{K'} \log \lambda_{g,i} = \log G + c \tag{112}
\]
we can obtain (17) by using (100), (101) and (78). The remaining detail is omitted due to space limitation.

APPENDIX D
PROOF OF THEOREM 4

The achievability proof of (23) begins with (93) and Corollary 1 with the uniform power allocation over groups such that \( S_g = \frac{P}{M} I_{r_g} \), yielding
\[
\sum_{g=1}^{G} r_g \log \log K_g + M \log \frac{P}{M} + \sum_{g=1}^{G} \log |\Lambda_g| + o(1). \tag{113}
\]
Compared to \( M \log \log K \) in the i.i.d. Rayleigh fading case, the multiuser diversity gain reduces to \( \sum_{g=1}^{G} r_g \log \log K_g \). To show that this diversity gain reduction vanishes for sufficiently large \( K_g \), we use the logarithmic identity
\[
\log_c(a \pm b) = \log_c a + \log_c \left( 1 \pm \frac{b}{a} \right) \tag{114}
\]
where \( a \) and \( b \) are nonnegative. Then, we get
\[
\sum_{g=1}^{G} r_g \log \log K_g = M \log \log K + o(1)
\]
for large \( K_g \). This proves the achievability.

The converse part of the proof is given next. By defining the \( M \times M \) aggregated matrices \( \Lambda = \text{diag} \{ \Lambda_1, \cdots, \Lambda_G \} \), \( \overline{W} = \text{diag} \{ W_1, \cdots, W_G \} \), and \( \overline{S} = \text{diag} \{ S_1, \cdots, S_G \} \), the sum of the first log det term in (93) can be rewritten as
\[
\sum_{g=1}^{G} \log |\Lambda_g^{-1} + W_g S_g W_g^H| = \log |\Lambda^{-1} + \overline{W} \overline{S} \overline{W}^H|. \tag{115}
\]
Similar to the proof of [9] Lem. 1] based on the extreme value theory, for sufficiently large $K$, the right-hand side above can then be upper-bounded as

$$
\mathbb{E} \left[ \log |A^{-1} + WSW^H| \right] \overset{(a)}{\leq} \mathbb{E} \left[ M \log \left( \frac{\text{tr}(A^{-1})}{M} + \frac{1}{M} \text{tr}(W^H WS) \right) \right] \\
\overset{(b)}{\leq} \mathbb{E} \left[ M \log \left( \frac{\text{tr}(A^{-1})}{M} + \frac{P}{M} \max_{g,k} \|w_{g,k}\|^2 \right) \right] \\
\overset{(c)}{=} M \log \left( \frac{P}{M} \log K + \frac{\text{tr}(A^{-1})}{M} + O(\log \log K) \right) \\
\overset{(d)}{=} M \log \left( \frac{P}{M} \log K \right) + M \log \left( 1 + \frac{\text{tr}(A^{-1})}{P \log K} + O\left( \frac{\log \log K}{\log K} \right) \right) \\
= M \log \log K + M \log \frac{P}{M} + o(1)
$$

(116)

where $(a)$ follows from the geometric-arithmetic mean inequality $|A| \leq \left( \frac{\text{tr}(A)}{r} \right)^r$ with $r$ the rank of $A$ and $(b)$ follows from $\text{tr}(W^H WS) \leq P \max_{g,k} \|w_{g,k}\|^2$, where $\text{tr}(S) = P$, since $S$ is diagonal, $(c)$ follows from $\|w_{g,k}\|^2 \sim \chi^2(2M)$ and the fact that the maximum of $n$ i.i.d. $\chi^2(2m)$ random variables behaves with high probability like $\log n + O(\log \log n)$, and $(d)$ follows from (114). Plugging (116) into (93), we establish the upper bound and obtain (23).

APPENDIX E
PROOF OF THEOREM 5

We first calculate the pdf and cdf of $\text{SINR}$. Since the random variables $v$ and $y$ in Sec. VI-C are independent, the pdf of $\text{SINR}_{g,k,m}$ can be calculated, conditioned on $y$, by

$$
f(x) = \int_0^\infty \left( \gamma_{g,m} + \frac{y}{\lambda_{g,m}} \right) e^{-\left(\gamma_{g,m} + \frac{y}{\lambda_{g,m}}\right)x} \sum_{i=1}^{M-1} \eta_i e^{-\frac{x}{\sigma_i^2}} dy \\
= \sum_{i=1}^{M-1} \gamma_{g,m} \eta_i e^{-\gamma_{g,m}x} \int_0^\infty e^{-\left(\sigma_i^{-2} + \frac{x}{\lambda_{g,m}}\right)y} dy + \sum_{i=1}^{M-1} \frac{\eta_i}{\lambda_{g,m}} e^{-\gamma_{g,m}x} \int_0^\infty ye^{-\left(\sigma_i^{-2} + \frac{x}{\lambda_{g,m}}\right)y} dy \\
= e^{-\gamma_{g,m}x} \sum_{i=1}^{M-1} \eta_i \left( \gamma_{g,m} \left( \frac{1}{\sigma_i^2} + \frac{x}{\lambda_{g,m}} \right) \right)^{-1} \left( \frac{1}{\sigma_i^2} + \frac{x}{\lambda_{g,m}} \right)^{-2} \left( \frac{x}{\lambda_{g,m}} \right)^{-1} \\
(117)
$$

where $\gamma_{g,m} = \frac{M}{a \lambda_{g,m} P}$ and $\eta_i = (\sigma_i^{-2} \prod_{j=1,j \neq i}^{M-1} (1 - \sigma_j^{-2}/\sigma_i^{-2}))^{-1}$. The cdf of $\text{SINR}_{g,k,m}$ can then be written as

$$
F(x) = \int_0^x e^{-\gamma_{g,m}x} \sum_{i=1}^{M-1} \eta_i \left( \gamma_{g,m} \left( \frac{1}{\sigma_i^2} + \frac{x}{\lambda_{g,m}} \right) \right)^{-1} \left( \frac{1}{\sigma_i^2} + \frac{x}{\lambda_{g,m}} \right)^{-2} \left( \frac{x}{\lambda_{g,m}} \right)^{-1} dx \\
= \sum_{i=1}^{M-1} \eta_i \left( \frac{1}{\sigma_i^2} + \frac{x}{\lambda_{g,m}} \right)^{-1} \\
= 1 - e^{-\gamma_{g,m}x} \sum_{i=1}^{M-1} \eta_i \left( \frac{1}{\sigma_i^2} + \frac{x}{\lambda_{g,m}} \right)^{-1} \left( \frac{x}{\lambda_{g,m}} \right)^{-1} \left( \frac{x}{\lambda_{g,m}} \right)^{-1} \\
(118)
$$

where the last equality follows from the fact that $\int_0^\infty f(x) dx = 1$.

Given the distribution of $\text{SINR}$, we next find the asymptotic behavior of the maximum of $\text{SINR}_{g,k,m}$ for users $g \in G', \gamma_{g,k,m}$ by using the result in [36] on extreme value theory. In a way to apply the known result,
we verify the two conditions in [9, Corollary A.1]. So, let us evaluate the following growth function
\[ g(x) = 1 - F(x) \] to see if it is a positive constant in the limit of \( x \to \infty \), which is the first condition.

\[
\lim_{x \to \infty} g(x) = \lim_{x \to \infty} e^{-\gamma_{g,m} x} \sum_{i=1}^{M-1} \eta_i \left( \frac{1}{\sigma_i^2 + \frac{x}{\lambda_{g,m}}} \right)^{-1} - \left( \frac{1}{\sigma_i^2 + \frac{x}{\lambda_{g,m}}} \right)^{-2}
\]

\[
= \lim_{x \to \infty} \left( \gamma_{g,m} + \sum_{i=1}^{M-1} \eta_i \left( \frac{\sigma_i^2 \lambda_{g,m} + \sigma_i^2 x}{\lambda_{g,m} + \sigma_i^2 x} \right) \right)^{-1} - \left( \frac{1}{\gamma_{g,m} + \sum_{i=1}^{M-1} \eta_i \left( \frac{\sigma_i^2 \lambda_{g,m} + \sigma_i^2 x}{\lambda_{g,m} + \sigma_i^2 x} \right)} \right)^{-1}
\]

\[
= \gamma_{g,m} > 0.
\]

(119)

Thus, the first condition is met. As to the second condition, we need to find \( u \) to satisfy

\[
1 - F(u) = e^{-\gamma_{g,m} u} \sum_{i=1}^{M-1} \eta_i \left( \frac{1}{\sigma_i^2 + \frac{u}{\lambda_{g,m}}} \right)^{-1} = \frac{1}{K'_g}
\]

(120)

where \( K'_g = |G'_g| \). By rewriting (120) as

\[
\gamma_{g,m} u = \log K'_g + \sum_{i=1}^{M-1} \eta_i \left( \frac{1}{\sigma_i^2 + \frac{u}{\lambda_{g,m}}} \right)^{-1}
\]

and if we suppose \( u = O(\log K'_g) \), it can be easily seen that

\[
u = \gamma_{g,m}^{-1} \log K'_g - \gamma_{g,m}^{-1} \log \log K'_g + O(1)
\]

(121)

\[
u = \gamma_{g,m}^{-1} \log K - \gamma_{g,m}^{-1} \log \log K + O(1), \text{ for large } K
\]

where we used the fact that \( K'_g = O(K'_{g}) = O(K) \).

It follows from [9, Corollary A.1] and (121) that \( \max_{gk \in G'_g} \text{SINR}_{gk,m} \) behaves with high probability like \( \gamma_{g,m}^{-1} \log K + O(\log \log K) \) for sufficiently large \( K \). Applying this to (60), we can rewrite (44) as

\[
\mathcal{R}_{\text{GBF-SUP}} \geq \mathbb{E} \left[ \sum_{g=1}^{G} \sum_{m=1}^{r_g} \log \left( 1 + \max_{gk \in G'_g} \text{SINR}_{gk,m} \right) \right] + o(1)
\]

\[
= \sum_{g=1}^{G} \sum_{m=1}^{r_g} \log \left( 1 + \gamma_{g,m}^{-1} \log K + O(\log \log K) \right) + o(1)
\]

\[
= \sum_{g=1}^{G} \sum_{m=1}^{r_g} \left( \log \log K - \log \gamma_{g,m} \right) + o(1)
\]

\[
= M \log \log K + M \log \frac{P}{M} + \sum_{g=1}^{G} \sum_{m=1}^{r_g} \log \lambda_{g,m} + M \log \alpha + o(1).
\]

(122)

This completes the proof.
Assuming the perfect similarity, we have

\[
\mathcal{R}^{\text{sum}} = \max_{\sum_{g=1}^{G} \sum_{k=1}^{h_g} \text{tr}(P_{g_k}) \leq P} \log \left| I + \sum_{g=1}^{G} \sum_{k=1}^{K_g} U_g A_g^{1/2} w_{g_k} P_{g_k} w_{g_k}^H A_g^{1/2} U_g^H \right|
\]

\[
= \max_{\sum_{g=1}^{G} \text{tr}(S_g) \leq P} \log \left| I_M + \sum_{g=1}^{G} U_g K_g U_g^H \right|
\]

\[
\geq \max_{\sum_{g=1}^{G} \text{tr}(S_g) \leq P} \sum_{g=1}^{G} \log \left| I_{r-g} + G_g^H U_g K_g U_g^H G_g \right| \tag{123}
\]

where \( K_g = \sum_{k=1}^{K_g} A_g^{1/2} w_{g_k} P_{g_k} w_{g_k}^H A_g^{1/2} \) and the inequality follows from the data processing inequality \([53]\) since we used \( G_g \) in \([69]\) as a receiver filter for users in group \( g \) and from the fact that \( G_h^H U_h = 0, h \neq g \), by definition.

Using Lemma \(6\) in Appendix A to further bound \(123\), we have

\[
\left| I_{r-g} + G_g^H U_g K_g U_g^H G_g \right| = \left| I_{r_g} + U_g^H G_g G_g^H U_g K_g \right|
\]

\[
= \prod_{i=1}^{r_g} \left( 1 + \lambda_i(U_g^H G_g G_g^H U_g K_g) \right)
\]

\[
\leq \prod_{i=1}^{r_g} \left( \lambda_i(U_g^H G_g G_g^H U_g) + \lambda_{r+1}(U_g^H G_g G_g^H U_g K_g) \right)
\]

\[
\geq \left| U_g^H G_g G_g^H U_g + U_g^H G_g G_g^H U_g K_g \right|
\]

\[
= \left| U_g^H G_g G_g^H U_g \right| \left| I_{r_g} + K_g \right| \tag{124}
\]

where \( a \) follows from the fact that \( 0 \leq \lambda_i(U_g^H G_g G_g^H U_g) \leq 1 \) for all \( i \) since \( \lambda_i(U_g^H G_g G_g^H U_g) = \cos^2(\theta_i) \) (e.g., see \([42], [54]\)), where \( \theta_i \) is the \( i \)th principal angle between two subspaces \( \mathcal{U}_{U_g} \) and \( \mathcal{U}_{G_g} \), and \( b \) follows from \([91]\). Then, we obtain

\[
\mathcal{R}^{\text{sum}} \geq \sum_{g=1}^{G} \max_{\sum_{g=1}^{G} \text{tr}(S_g) \leq P} \sum_{g=1}^{G} \log \left| U_g^H G_g G_g^H U_g \right| \left| I_{r_g} + K_g \right|
\]

\[
\geq \max_{\sum_{g=1}^{G} \text{tr}(S_g) \leq P} \sum_{g=1}^{G} \log \left| I_{r_g} + \Lambda_g^{1/2} W_g S_g W_g^H \Lambda_g^{1/2} \right| + G \log \cos(\delta_{FS}) \tag{125}
\]

where the second inequality follows from \([70]\).

### References

1. G. J. Foschini, “Layered space-time architecture for wireless communication in a fading environment when using multi-element antennas,” Bell Labs Tech. J., vol. 1, no. 2, pp. 41–59, 1996.
2. I. E. Telatar, “Capacity of multi-antenna Gaussian channels,” Europ. Trans. Telecomm., vol. 10, pp. 585–595, 1999.
3. D. Shiu, G. Foschini, M. Gans, and J. Kahn, “Fading correlation and its effect on the capacity of multielement antenna systems,” Communications, IEEE Transactions on, vol. 48, no. 3, pp. 502–513, 2000.
4. C. Chuah, D. Tse, J. Kahn, and R. Valenzuela, “Capacity scaling in MIMO wireless systems under correlated fading,” Information Theory, IEEE Transactions on, vol. 48, no. 3, pp. 637–650, 2002.
[39] E. Scheuer, “Reliability of an m-out-of-n system when component failure induces higher failure rates in survivors,” IEEE Trans. Rel., vol. 37, no. 1, pp. 73–74, 1988.

[40] B. O. E. Bjørnson, D. Hammarwall, “Exploiting quantized channel norm feedback through conditional statistics in arbitrarily correlated MIMO systems,” IEEE Trans. on Sig. Proc., vol. 57, pp. 4027–4041, 2009.

[41] J. H. Conway, R. H. Hardin, and N. J. A. Sloane, “Packing lines, planes, etc.: Packings in grassmannian spaces,” Exper. Math., vol. 5, no. 2, pp. 139–159, 1996.

[42] A. Barg and D. Y. Nogin, “Bounds on packings of spheres in the grassmannian manifold,” IEEE Trans. on Inform. Theory, vol. 48, no. 9, pp. 2450–2454, 2002.

[43] D. J. Love and R. W. Heath, “Limited feedback unitary precoding for spatial multiplexing systems,” IEEE Trans. on Inform. Theory, vol. 51, no. 8, pp. 2967–2976, 2005.

[44] G. Dimic and N. Sidiropoulos, “On downlink beamforming with greedy user selection: performance analysis and a simple new algorithm,” IEEE Trans. on Sig. Proc., vol. 53, no. 10, pp. 3857–3868, 2005.

[45] N. Ravindran and N. Jindal, “Multi-user diversity vs. accurate channel feedback for MIMO broadcast channels,” Proc. IEEE Int. Conf. on Commun. (ICC), pp. 3684–3688, May 2008.

[46] A. Adhikary, E. A. Safadi, M. Samimi, R. Wang, G. Caire, T. S. Rappaport, and A. F. Molisch, “Joint spatial division and multiplexing for mm-Wave channels,” submitted to IEEE J. Select. Areas Commun., 2013. [Online]. Available: [http://arxiv.org/abs/1312.2045](http://arxiv.org/abs/1312.2045)

[47] A. M. Tulino and S. Verdú, “Random matrix theory and wireless communications,” Foundations and Trends in Commun. and Inf. Theory, vol. 1, no. 1, pp. 1–182, 2004.

[48] R. A. Horn and C. R. Johnson, Matrix Analysis. Cambridge University Press, 1990.

[49] L. Lu and C. Pearce, “Some new bounds for singular values and eigenvalues of matrix products,” Ann. Oper. Res., vol. 98, pp. 141–148, 2000.

[50] A. W. Marshall, I. Olkin, and B. C. Arnold, Inequalities: Theory of Majorization and its Applications. Springer, 2011.

[51] S. Boyd and L. Vandenberghe, Convex Optimization. Cambridge University Press, New York, NY, 2004.

[52] M. Fiedler, “Bounds for the determinant of the sum of hermitian matrices,” Proc. AMS, vol. 30, no. 1, pp. 27–31, 1971.

[53] T. M. Cover and J. A. Thomas, Elements of Information Theory. New York: Wiley, 2006.

[54] A. Björck and G. Golub, “Numerical methods for computing angles between linear subspaces,” Mathematics of Computation, vol. 27, no. 123, pp. 579–594, 1973.