

Joint Spatial Division and Multiplexing – Benefits of Antenna Correlation in Multi-User MIMO

Junyoung Nam and Jae-Young Ahn

Mobile Communications Division

Electronics Telecommunications Research Institute (ETRI), Daejeon, Korea, *email:*{jynam, jyahn}@etri.re.kr

Abstract—We show that the sum capacity of “spatially well-correlated” Gaussian MIMO broadcast channel (BC) is rather larger than that of the uncorrelated (i.i.d. Rayleigh fading) Gaussian MIMO BC, assuming a special structure of transmit correlation of users. Moreover, a potential sum-rate gap of an ideal correlated Rayleigh fading channel is shown to be $M \log M$ over the uncorrelated case, where M is the number of transmit antennas. Finally, we propose a joint spatial division and multiplexing (JSDM) scheme based on opportunistic beamforming that can achieve the optimal multiuser diversity gain with very limited channel state information (CSI) feedback in realistic scenarios, not requiring the special structure of transmit correlation.

I. INTRODUCTION

In the point-to-point case, the fundamental limits and optimal precoding of spatially correlated fading MIMO channels have been well characterized in a variety of transmit correlation models. On the contrary, the fundamental limits of spatially correlated multi-user (MU) MIMO channels have not been fully understood in general models on transmit correlation matrices \mathbf{R}_i of users $i = 1, \dots, K$. In particular, the previous results on sum-rate scaling (growth) well characterized some special cases with the aid of two main tools: large-dimensional random matrix theory and extreme value theory. While a large system analysis based on the former might be initiated by [1], where system parameters M and K go to infinity with fixed ratio $1 \leq \frac{M}{K} < \infty$, the latter was exploited to explicitly reveal the multi-user diversity gain by [2], where M is fixed and only K goes to infinity.

The impact of transmit correlation on the capacity of MU-MIMO systems has not received much attention and it has not been well characterized yet. In particular, the work of [3] extended the sum-rate scaling result of [2] to the special case where all users have the same transmit correlation such that $\mathbf{R}_i = \mathbf{R}, \forall i$, which concludes that transmit correlation has a fairly *detrimental* impact on the sum capacity of MU-MIMO downlink systems, compared with the uncorrelated channel case. This coincides with the traditional view on the impact of transmit correlation on capacity. However, in more general transmit correlation models where users have different \mathbf{R}_i , there are a number of works that exploit transmit correlation and address its potential benefits in several aspects. The benefits of transmit correlation have been observed in

terms of CSI feedback overhead, scheduling, codebook design, and throughputs of linear precoding schemes (e.g., [4], [5]).

A couple of fundamental questions on transmit antenna correlation naturally arise: *What is an optimal sum-rate scaling law of MU-MIMO downlink systems where users have different transmit correlation matrices? Is transmit correlation still detrimental to the capacity of this more realistic MU-MIMO systems? If it is even beneficial to the capacity, how much a potential gain is over the uncorrelated case?*

To address the above issues, we make use of the notion of JSDM and an important structure of users' transmit correlation called the (tall) unitary structure. The idea of JSDM was originally proposed by the authors in [6] as a feasible solution to apply a large-scale array to FDD downlink systems and to achieve the full-CSI capacity with limited feedback. In order to leverage benefits of transmit correlation and to explore an impact of transmit correlation on the capacity of MU-MIMO downlink systems, we focus here on the regime where M is fixed and K is large. The interested readers in the other regime where both M and K are large with $M \geq K$ are encouraged to refer to our companion paper [7].

In this paper, we first show that the sum capacity of correlated Gaussian MIMO BC may be significantly larger than that of uncorrelated Gaussian MIMO BC, under the unitary structure of transmit correlation of users. Moreover, the potential sum-rate gap of a correlated fading channel is $M \log M$ over the uncorrelated case, in some circumstances. Not relying on the unitary structure unusual in practice, we propose a JSDM scheme with opportunistic beamforming [8] and prove that for K large, it can achieve the optimal sum-rate scaling with minimal feedback. Finally, simulation results show that when $K = 100$, the proposed scheme already realizes a large portion of multi-user diversity gain for $M = 4$.

II. SYSTEM DESCRIPTION

Consider the downlink of a single-cell FDD system with a base station (BS) with M antennas serving K UTs (user terminals) equipped with a single antenna each, whose channels obey the one-ring channel model [9], a special case of the well-known Kronecker correlation model with no receive antenna correlation. Let \mathbf{R} denote the transmit correlation matrix, define $r = \text{rank}(\mathbf{R})$ and also let r^* denote the number of *dominant* (non-negligible) eigenvalues of \mathbf{R} . By using the Karhunen-Loeve transform, the channel vector from the BS

This work was supported by the IT R&D program of MKE/KEIT in Korea [Development of beyond 4G technologies for smart mobile services].

antenna array to a UT can then be expressed as $\mathbf{h} = \mathbf{U}\mathbf{\Lambda}^{\frac{1}{2}}\mathbf{w}$, where $\mathbf{w} \in \mathbb{C}^{r \times 1} \sim \mathcal{CN}(\mathbf{0}, \mathbf{I})$, $\mathbf{\Lambda}$ 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.

In what follows, we briefly review the idea of JSDM. More details can be found in [6], [7]. JSDM is based on a structured (two-stage) precoding given by the transmit signal $\mathbf{x} = \mathbf{B}\mathbf{P}\mathbf{d}$, where $\mathbf{B} \in \mathbb{C}^{M \times b}$ is a *pre-beamforming* matrix that depends only on the channels second-order statistics, i.e., on set $\{\mathbf{U}_g, \mathbf{\Lambda}_g\}$, or on some directional information extracted from \mathbf{R} , and $\mathbf{P} \in \mathbb{C}^{b \times S}$ is a beamforming matrix that depends on either the instantaneous realization of channel vectors linearly transformed by \mathbf{B} or any other instantaneous¹ CSI feedback with reduced dimensionality.

In order to embed a useful structure into users' transmit correlation for arbitrarily distributed users and scatterer geometry, we have to elaborately partition the entire set of users, $\mathcal{K} = \{1, 2, \dots, K\}$ into T non-overlapping subsets, $\mathcal{T}^{(1)}, \mathcal{T}^{(2)}, \dots, \mathcal{T}^{(T)}$, which we call *classes* in this work. Users associated to different classes are served with different downlink resources. Class t has G_t groups. For simplicity, however, we may assume a symmetric system where each class has G groups and each group has the same number of dominant eigenvalues such that $r_g^* = r^*$ for all g . Let the integer $K_g^{(t)}$ denote the number of users assigned to group g of class t such that $K = \sum_t \sum_g K_g^{(t)}$ to take the notion of class into account. Then, the t th class can be written as $\mathcal{T}^{(t)} = \{\mathcal{G}_1^{(t)}, \mathcal{G}_2^{(t)}, \dots, \mathcal{G}_G^{(t)}\}$, where $\mathcal{G}_g^{(t)} = \{g_1^{(t)}, g_2^{(t)}, \dots, g_{K_g^{(t)}}^{(t)}\}$ with $g_k^{(t)} = \sum_{i=1}^{t-1} \sum_{j=1}^{G-1} K_j^{(i)} + k$. Throughout this paper, the superscript (t) for classes is omitted unless necessary. Then, we let $g_k = (g-1)K_g + k$.

Once user partitioning is done, we let \mathbf{H}_g denote the aggregate channel matrix of group g and $\mathbf{B}_g \in \mathbb{C}^{M \times b_g}$ denote pre-beamforming matrices, where b_g is closely related to r_g^* and not smaller than the number of data streams for group g , denoted by S_g . In case of treating signals from other groups as interference, the precoding matrix takes on the block-diagonal form $\mathbf{P} = \text{diag}(\mathbf{P}_1, \dots, \mathbf{P}_G)$, where $\mathbf{P}_g \in \mathbb{C}^{b_g \times S_g}$, yielding the vector broadcast plus interference Gaussian channel

$$\mathbf{y}_g = \mathbf{H}_g^H \mathbf{B}_g \mathbf{P}_g \mathbf{d}_g + \sum_{h \neq g} \mathbf{H}_g^H \mathbf{B}_h \mathbf{P}_h \mathbf{d}_h + \mathbf{z}_g, \text{ for all } g. \quad (1)$$

The structure of transmit correlation can be characterized by the two properties defined as follows.

- **Similarity:** the eigenvector matrices $\{\mathbf{U}_{g_k}\}_{k=1}^{K_g}$ of users in group g are sufficiently similar to the group eigenvector matrix \mathbf{U}_g . A strict case is that $\mathbf{U}_{g_k} = \mathbf{U}_g$ for all (g, k) .
- **Orthogonality:** the eigenvector matrices $\{\mathbf{U}_g\}_{g=1}^G$ of groups in a class are such that $\mathbf{U}_g^H \mathbf{U}_h = \mathbf{0}$, for all $h \neq g$.

In order to elucidate a potential benefit of JSDM, we now define an ideal structure of the users' transmit correlation

¹This implies that UTs should immediately feed the CSI back to the BS within every coherence time of channel.

matrices that strictly satisfies the above properties. For this special case, we assume that G groups are formed in the symmetric manner, i.e., $r_g = r$ for all g .

Definition 1 (Tall Unitary Structure): For $M \geq rG$, we have a tall unitary structure if $\mathbf{U}_{g_k} = \mathbf{U}_g$ for all k and if the $M \times rG$ matrix $\underline{\mathbf{U}} = [\mathbf{U}_1, \dots, \mathbf{U}_G]$ is *tall unitary* such that $\underline{\mathbf{U}}^H \underline{\mathbf{U}} = \mathbf{I}$.

For $M = rG$, we call this the unitary structure since $\underline{\mathbf{U}}$ is now unitary. When the tall unitary structure in Definition 1 is available, we simply let $\mathbf{B} = \underline{\mathbf{U}}$. In this case, the MU-MIMO channel (1) takes on the form

$$\mathbf{y}_g = \mathbf{H}_g^H \mathbf{P}_g \mathbf{d}_g + \mathbf{z}_g = \mathbf{W}_g^H \mathbf{\Lambda}_g^{1/2} \mathbf{P}_g \mathbf{d}_g + \mathbf{z}_g, \text{ for all } g \quad (2)$$

where $\mathbf{H}_g = \mathbf{B}_g^H \mathbf{H}_g$ denotes the *transformed channel*, and \mathbf{W}_g is a $r \times K_g$ i.i.d. matrix with elements $\sim \mathcal{CN}(0, 1)$, for all g . The next results show that, under these assumptions, JSDM achieves the same sum capacity of the MU-MIMO downlink with full CSI.

Theorem 1 ([6]): Under the tall unitary structure, the sum capacity of the MU-MIMO downlink channel with full CSI is equal to the sum capacity of parallel channels (2), given by

$$\mathcal{C}^{\text{sum}}(\underline{\mathbf{H}}; P) = \max_{\sum_g \text{tr}(\mathbf{S}_g) \leq P} \sum_{g=1}^G \log \left| \mathbf{I} + \mathbf{\Lambda}_g^{1/2} \mathbf{W}_g \mathbf{S}_g \mathbf{W}_g^H \mathbf{\Lambda}_g^{1/2} \right|$$

where \mathbf{S}_g denotes the diagonal $K_g \times K_g$ input covariance matrix for group g in the dual MAC channel.

III. IMPACT OF TRANSMIT CORRELATION ON THE CAPACITY OF MULTI-USER MIMO SYSTEMS

Traditionally, transmit correlation has been considered to be a detrimental source to the capacity of MIMO wireless channels. In the context of multi-user MIMO scheduling, a traditional view on the impact of transmit antenna correlation on the sum capacity of MIMO BC is well represented by [3], which shows that transmit correlation is always harmful to the sum-rate scaling law. This result is revisited in a general way.

For i.i.d. Rayleigh fading channels, it is well known from Sharif and Hassibi [2] that the sum capacity of MIMO Gaussian BCs scales like

$$\mathcal{C}_{\text{i.i.d.}}^{\text{sum}} = M \log \log K + M \log \frac{P}{M} + o(1) \quad (3)$$

where $o(1)$ goes to zero as $K \rightarrow \infty$. Furthermore, the same scaling law was shown to be achievable by the random beamforming (RBF), which requires only partial CSI.

If channels have transmit antenna correlation, we can find some partial results on scaling laws for correlated MIMO BCs. In particular, when all users have both the same SNR and the common transmit correlation matrix \mathbf{R} of full rank, [3] proved that the sum capacity scales like $M \log \log K + M \log \frac{P}{M} + \log |\mathbf{R}| + o(1)$. This implies that the transmit correlation is purely detrimental to the asymptotic capacity since $|\mathbf{R}| \leq 1$. As an immediate consequence, we have the general form:

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

$$r \log \log K + r \log \frac{P}{r} + \log |\mathbf{A}| + o(1). \quad (4)$$

The above assumption that all users have the same transmit correlation is far from realistic transmit correlation of UTs in a cellular system. In what follows, we present a scaling law for a more general MIMO BC, in which users have different transmit correlation with the desirable unitary structure in Definition 1, as an intermediate step to a realistic case in Sec. V. As the sum capacity of a correlated channel under the unitary structure depends largely on r_g , we let $\mathcal{C}_{\mathbf{r}}^{\text{sum}}$ denote the sum capacity of this channel, where $\mathbf{r} = (r_1, \dots, r_G)$.

Theorem 2: For fixed M and sufficiently large K , the ergodic sum capacity of Gaussian MIMO BC under the unitary structure with \mathbf{R}_g of rank r_g is

$$\mathcal{C}_{\mathbf{r}}^{\text{sum}} = M \log \log K + M \log \frac{P}{M} + \sum_{g=1}^G \sum_{m=1}^{r_g} \log \lambda_{g,m} + o(1)$$

where $\lambda_{g,m}$ is the m th eigenvalue of \mathbf{A}_g .

Proof: Owing to Theorem 1, the sum capacity of Gaussian MIMO BC under the unitary structure can be written as

$$\mathbb{E} \left[\max_{\sum_g \text{tr}(\mathbf{S}_g) \leq P} \sum_{g=1}^G \log \left| \mathbf{A}_g^{-1} + \mathbf{W}_g \mathbf{S}_g \mathbf{W}_g^H \right| + \sum_{g=1}^G \log |\mathbf{A}_g| \right].$$

The achievability part of the proof then immediately follows from Corollary 1 with $\text{tr}(\mathbf{S}_g) = \frac{r_g P}{M}$ for all g .

The converse part is given as follows. By defining the $M \times M$ aggregated matrices $\underline{\mathbf{A}} = \text{diag} \{ \mathbf{A}_1, \dots, \mathbf{A}_G \}$, $\underline{\mathbf{W}} = \text{diag} \{ \mathbf{W}_1, \dots, \mathbf{W}_G \}$, and $\underline{\mathbf{S}} = \text{diag} \{ \mathbf{S}_1, \dots, \mathbf{S}_G \}$, the first sum of log det terms in the above equation can be rewritten as $\sum_{g=1}^G \log \left| \mathbf{A}_g^{-1} + \mathbf{W}_g \mathbf{S}_g \mathbf{W}_g^H \right| = \log \left| \underline{\mathbf{A}}^{-1} + \underline{\mathbf{W}} \underline{\mathbf{S}} \underline{\mathbf{W}}^H \right|$. Similar to the proof of [2, Lem. 1] based on the extreme value theory, for sufficiently large K , the right-hand side of the above can be upper-bounded as

$$\mathbb{E} \left[\log \left| \underline{\mathbf{A}}^{-1} + \underline{\mathbf{W}} \underline{\mathbf{S}} \underline{\mathbf{W}}^H \right| \right] \leq M \log \log K + M \log \frac{P}{M} + o(1)$$

where we used the geometric-arithmetic mean inequality $|\mathbf{A}| \leq \left(\frac{\text{tr}(\mathbf{A})}{r} \right)^r$ with r the rank of \mathbf{A} , 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 the logarithmic identity $\log_c \sum_{i=1}^n a_i = \log_c a_1 + \log_c \left(1 + \sum_{i=2}^n \frac{a_i}{a_1} \right)$, where a_i is non-negative for all i . This completes the proof. ■

An important implication of the above theorem is that the sum capacity of “spatially well-structured” Gaussian MIMO BC may be much larger than that of “spatially white” (uncorrelated) Gaussian MIMO BC, by exploiting the spatial structure of transmit correlation of users. The following corollary shows this sum-rate gap becomes considerably huge as the angular spread of group g denoted as Δ_g goes to zero for all g and M (and hence G) is large.

Corollary 2: If Δ_g is sufficiently small such that r_g goes to one for all g and there is no dominant line-of-sight (LoS) component so that Rayleigh fading is still valid, then

$$\lim_{\mathbf{r} \rightarrow (1, \dots, 1)} \mathcal{C}_{\mathbf{r}}^{\text{sum}} - \mathcal{C}_{\text{i.i.d.}}^{\text{sum}} = M \log M. \quad (5)$$

Proof: In the case of $\Delta_g \rightarrow 0$ and equivalently $r_g \rightarrow 1$, we have $G = M$ and $\lambda_{g,1} = M$ due to the fact that $\text{tr}(\mathbf{A}_g) = M$ for all g . By Theorem 2, the sum-rate scaling of this ideal MIMO BC under the unitary structure is $M \log \log K + M \log \frac{P}{M} + M \log M + o(1)$. Comparing this with (3), we conclude (5). ■

This corollary shows the surprising result that the sum-rate gap of the ideal correlated fading channel under the unitary structure is $M \log M$ over the uncorrelated fading channel case, previously viewed as the best channel condition for the capacity of MU-MIMO downlink. This result shifts the traditional paradigm on transmit antenna correlation from detrimental impact to beneficial impact on the capacity of wireless multi-user systems. Furthermore, this provides an insight into how to design a cellular system since we can control transmit antenna correlation with antenna spacing and/or to some extent by adjusting antenna height.

IV. GROUP-SPECIFIC BEAMFORMING (GBF) WITH SEMI-UNITARY USER PARTITION (SUP)

In Theorem 2, we imposed the strong assumption that all users satisfy the unitary structure. Moreover, it is assumed the BS perfectly knows the transformed channel state \mathbf{H}_g fed back by UTs. In what follows, we will show that the same scaling law in Theorem 2 can be achieved in realistic situations, requiring neither the unitary structure nor \mathbf{H}_g . We will omit the superscript $*$ on all related notations for convenience throughout the rest of this paper, e.g., \mathbf{U}_g implies \mathbf{U}_g^* hereafter.

We propose a semi-static beamforming called group-specific beamforming. The basic idea of GBF and its CSI feedback is based on opportunistic beamforming [8] and RBF with partial CSI [2]. In order to enable the GBF, we introduce a user partition algorithm that approximately realizes the (tall) unitary structure of users’ transmit correlation, which we will call the SUP algorithm. We describe the SUP algorithm based on fixed pre-beamforming matrices $\mathbf{B}_g^{(t)}$ in a compact form.

- 1) *Construction of $\mathbf{B}_g^{(t)}$:* Select $\mathbf{U}_g^{(t)}$ based on a selection criterion such that the orthogonality between $\mathbf{U}_g^{(t)}$ is approximated. Then, impose the BD constraint on $\{\mathbf{U}_g^{(t)}\}_{g=1}^{G_t}$ for each class, which yields $\mathbf{B}_g^{(t)}$.
- 2) *Class and group association:* UT i chooses and feeds back to the BS its class(es) and group(s) by using the following similarity measure.

$$(t^*, g^*) = \arg \max_{t,g} \|\mathbf{U}_i^H \mathbf{B}_g^{(t)}\|_F. \quad (6)$$

- 3) *Class size reduction:* If the number of users associated to a class is zero or notably small, discard that class and associate its users to the other classes according to UT’s alternatives in Step 2). This step is relevant for K small.

The above structure of users transmit correlation can be more strictly defined in the following form, where the first and second conditions are given in terms of the similarity and the orthogonality (in Sec. II) with threshold levels α_{LT} and β_{LT} , respectively.

Definition 2 (Semi-Unitary Structure): For $M \geq \sum_{g=1}^G r_g^*$, we have a semi-unitary structure if we can write \mathcal{G}_g as the following subset of \mathcal{K} :

$$\mathcal{G}_g = \left\{ i \in \mathcal{K} : \mathbf{A}_i^H \mathbf{A}_i \succeq \alpha_{\text{LT}} \mathbf{A}_g^*, \mathbf{C}_{i,h}^H \mathbf{C}_{i,h} \preceq \beta_{\text{LT}} \mathbf{I} \right\} \quad (7)$$

where $\mathbf{A}_i = \mathbf{\Lambda}_i^{1/2} \mathbf{U}_i^H \mathbf{U}_g^*$ and $\mathbf{C}_{i,h} = \mathbf{R}_i^{1/2} \mathbf{U}_h^*$.

The general transmit signal vector of JSDM for group g can be written as $\mathbf{x}_g = \mathbf{B}_g \mathbf{P}_g \mathbf{d}_g$. The GBF is defined by letting \mathbf{P}_g be a beam selection matrix. If a UT is selected for the m th beam by the BS scheduler, the beam is mapped by a column of \mathbf{P}_g to the UT's element of \mathbf{d}_g . The rationale behind the GBF is that by virtue of the SUP algorithm, the eigenmodes of users' transmit correlation \mathbf{R}_{g_k} in group g are supposed to be similar to those of \mathbf{R}_g . Thus, the pre-beamforming matrix \mathbf{B}_g closely approximates eigen-beamformers for users in group g .

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

$$\text{SINR}_{g_k,m} = \frac{|\mathbf{h}_{g_k}^H \mathbf{b}_{g,m}|^2}{\frac{M}{P} + \sum_{i=1, i \neq m}^{r_g} |\mathbf{h}_{g_k}^H \mathbf{b}_{g,i}|^2 + \sum_{h=1, h \neq g}^G \sum_{j=1}^{r_g} |\mathbf{h}_{g_k}^H \mathbf{b}_{h,j}|^2}. \quad (8)$$

The CSI feedback of GBF-SUP is the maximum SINR of each user, i.e. $\max_{1 \leq m \leq r_g} \text{SINR}_{g_k,m}$. Then, the BS schedules each beam $\mathbf{b}_{g,m}$ to the UT in group g with the highest SINR, i.e., $\max_{1 \leq k \leq K_g} \text{SINR}_{g_k,m}$.

While RBF requires $\lceil \log_2 M \rceil$ -bit instantaneous feedback per UT on beam index, the CSI feedback of GBF-SUP consists of long-term and instantaneous components. The long-term part is $\lceil \log_2 \sum_t G_t \rceil$ -bit feedback on class and group indices from UTs, which are extremely slow varying, and the instantaneous part is $\lceil \log_2 b_g \rceil$ -bit feedback on beam index.

V. ASYMPTOTIC OPTIMALITY OF GBF-SUP

Sharif and Hassibi [2] proved the asymptotic optimality of the RBF scheme in i.i.d. Rayleigh fading channels when K is sufficiently large by using a novel approach based on extreme value theory [10]. In essence, the authors showed that SINR's resulting from RBF are i.i.d. random variables and that the distribution of SINR conforms to sufficient conditions in [10] to apply some well-known results on extreme value theory. Unfortunately, this two requirements have restricted the application of their approach to more general channel models. In particular, [3] captures the impact of transmit correlation on achievable throughput only in the special case where all users have a common transmit correlation matrix \mathbf{R} . In a realistic system, users have different transmit correlation matrices \mathbf{R}_i and the dependance among the intended signal, intra-group interference and inter-group interference is quite burdensome. As we do not assume here the (tall) unitary structure on

\mathbf{R}_i , the SINR (8) of GBF-SUP can not meet the above two requirements. In order to simplify the problem and avoid the difficulty, we introduce a pair of covariance constraints to reformulate and lower-bound the SINR of GBF-SUP. Moreover, we employ a generalized Chi-squared distribution in [11] to calculate the distribution of SINR and to see if it satisfies the sufficient conditions.

As a complete proof of the asymptotic optimality of GBF-SUP is quite lengthy, we shall present only an outline as follows. We first have to derive a mathematically tractable lower bound of SINR (8). Since $\text{SINR}_{g_k,m}$ is very intricate due to lack of some structures in inter-user and inter-group interferences with arbitrary transmit correlations, it is in fact infeasible to establish such a lower bound for all users (g, k) . Therefore, we restrict our attention to a set of "opportunistic" users who satisfy the semi-unitary structure, given by covariance constraints in Definition 2, and some short-term constraints. While the semi-unitary structure can be attained by the SUP algorithm in the previous section, the latter constraints are introduced to guarantee that for the set of opportunistic users, a reformulated SINR (denoted by $\text{SINR}'_{g_k,m}$) is upper bounded by $\text{SINR}_{g_k,m}$.

Two important issues in this step should be cleared. One is that imposing the short-term constraints on \mathbf{h}_{g_k} shrinks the size of $|\mathcal{G}_g| = K_g$, where \mathcal{G}_g is given in (7), and hence it may reduce the multiuser diversity gain. The other is that we introduce two penalty factors denoted by α_{g_k} and β_{g_k} to satisfy the short-term constraints and then we have to check if the penalty factors are bounded appropriately by non-zero α and β , respectively. If not so, the resulting lower bound of $\text{SINR}_{g_k,m}$ may become arbitrarily small. The first issue can be solved by following the same line of [12]. We derive a lower bound on the cardinality of $|\mathcal{G}'_g|$, where \mathcal{G}'_g denotes a subset of users g_k in \mathcal{G}_g who meet the constraints, and show that it is still $O(K_g)$. Eventually, this multiuser diversity gain reduction is shown to vanish as K_g goes to infinity. Also, we could prove that the penalty factors are appropriately bounded after some manipulations based on the Cauchy-Schwarz inequality.

Letting $\alpha_{g_k} = \alpha$ and $\beta_{g_k} = \beta$ to satisfy the covariance (long-term) and short-term constraints, we now have the asymptotic lower bound of the original $\text{SINR}_{g_k,m}$

$$\text{SINR}'_{g_k,m} = \frac{|\mathbf{w}_{g_k}^H \boldsymbol{\lambda}_{g,m}^{1/2}|^2}{\frac{M}{\alpha P} + \frac{1}{\alpha} \sum_{i \neq m} |\mathbf{w}_{g_k}^H \boldsymbol{\lambda}_{g,i}^{1/2}|^2 + \frac{\beta}{\alpha} \sum_{h \neq g} \sum_{j=1}^{r_h^*} |\mathbf{h}_{W,g_k}^H \mathbf{u}_{g_k,h,j}|^2} \quad (9)$$

where $\boldsymbol{\lambda}_{g,m}$ and $\mathbf{u}_{g_k,h,m}$ are the m th column of $\mathbf{\Lambda}_g$ and $\mathbf{U}_{g_k,h}$, respectively, and $\mathbf{h}_{W,g_k} \in \mathbb{C}^{M \times 1} \sim \mathcal{CN}(\mathbf{0}, \mathbf{I})$. Here, $\mathbf{U}_{g_k,h}$ consists of r_h columns of the $M \times M$ (full-size) unitary matrix \mathbf{R}_{g_k} with rank $r_{g_k} \leq M$, whose subspace is similar (in terms of similarity measure) to the subspace spanned by the columns of \mathbf{U}_h of group h .

The advantage of imposing the covariance and short-term constraints is that the reformulated SINR's are now i.i.d. for all users $g_k \in \mathcal{G}'_g$. Furthermore, the intended signal,

intra-group interference and inter-group interference terms are independent.

In order to evaluate a lower bound of an achievable throughput of GBF-SUP in terms of $\text{SINR}'_{g_k,m}$ and obtain the scaling law of GBF-SUP, we should figure out the distribution of $\text{SINR}'_{g_k,m}$. Owing to the sought-after properties above of $\text{SINR}'_{g_k,m}$ and the generalized Chi-squared distribution [11], 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 $\text{SINR}_{g_k,m}$. For the space limitation, the details are omitted here. With the distributions of $\text{SINR}'_{g_k,m}$, we have the following result.

Theorem 3: For M fixed and K sufficiently large, the throughput of GBF-SUP scales like

$$M \log \log K + M \log \frac{P}{M} + \sum_{g=1}^G \log |\mathbf{\Lambda}_g| + o(1). \quad (10)$$

Proof (Sketch): We first investigate the asymptotic behavior of the maximum of $\text{SINR}'_{g_k,m}, k = 1, \dots, K_g$ by using the result in [10] on extreme value theory. In a way to apply the known result, we verify the two conditions in [2, Corollary A.1]. So, we evaluate the following growth function $g(x) = \frac{1-F(x)}{f(x)}$, where $f(x)$ and $F(x)$ are the pdf and cdf of $\text{SINR}'_{g_k,m}$, to see if it is a positive constant, which is the first condition. It turns out that the first condition is met such that $\lim_{x \rightarrow \infty} g(x) = \gamma_{g,m}^{-1} > 0$, where $\gamma_{g,m} = \frac{M}{\alpha \lambda_{g,m} P}$. As to the second condition, we need to find u to satisfy $1 - F(u) = e^{-\gamma_{g,m} u} \sum_{i=1}^{M-1} \eta_i \left(\frac{1}{\sigma_i^2} + \frac{u}{\lambda_{g,m}} \right)^{-1} = \frac{1}{K'_g}$. By rewriting this equation as $\gamma_{g,m} u = \log K'_g + \log \sum_{i=1}^{M-1} \eta_i \left(\frac{1}{\sigma_i^2} + \frac{u}{\lambda_{g,m}} \right)^{-1}$ and if we suppose $u = O(\log K'_g)$, it can be easily seen that

$$u = \gamma_{g,m}^{-1} \log K - \gamma_{g,m}^{-1} \log \log K + O(1). \quad (11)$$

where we used the fact that $K'_g = O(K)$.

It immediately follows from [2, Corollary A.1] and (11) that $\max_{1 \leq k \leq K_g} \text{SINR}'_{g_k,m}$ behaves like $\gamma_{g,m}^{-1} \log K + O(\log \log K)$ for sufficiently large K . Applying this to a throughput resulting from $\text{SINR}'_{g_k,m}$, we have (10). ■

Combining the results in Theorem 2 and 3, we have that GBF-SUP can achieve the optimal sum-rate scaling law, despite the very limited CSI at the BS.

VI. NUMERICAL RESULTS

We consider here a BS with $M = 4$ and $\frac{2\pi}{3}$ directional uniform linear array, and UTs uniformly distributed over the range $[-60^\circ, 60^\circ]$ with Δ_i uniformly distributed in the range $[5^\circ, 15^\circ]$. The users' transmit correlation matrices are generated by the one-ring channel model. We let $T = 2$, $G = 4$, $\Delta_g = 10^\circ$, and select angle of departure sets for each class as $\{-57.5, -23, 7.5, 41.5\}^\circ$ and $\{-41.5, -7.5, 23, 57.5\}^\circ$.

Fig. 1 compares the sum rates of DPC with greedy user selection (sum capacity), ZFBF-SUS [12], RBF, and GBF-SUP for $M = 4$. When $K = 100$, GBF-SUP already achieves a large portion of the throughput of ZFBF-SUS and even noticeably outperforms as K increases. Note that DPC

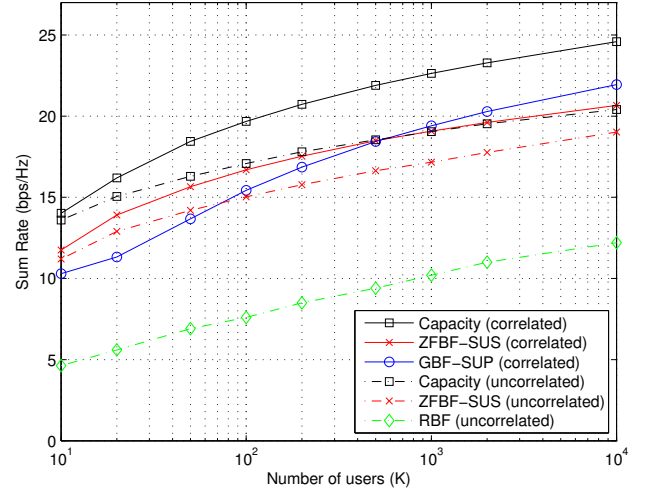


Fig. 1. Sum rates of different MU-MIMO schemes versus the number of users with $M = 4$, SNR = 10 dB, and $\Delta_i \in [5^\circ, 15^\circ]$.

and ZFBF-SUS assume the full CSI, while GBF-SUP needs only 3-bit *long-term* CSI feedback, requiring *no instantaneous* feedback on beam index except for the SINR feedback. As M increases, however, GBF-SUP is shown to require gradually much larger K to approach the performance of ZFBF-SUS. We can also see that the sum capacity of this realistic correlated Rayleigh fading channel is approximately 24% larger than that of the i.i.d. Rayleigh fading channel when $K = 10,000$.

REFERENCES

- [1] B. Hochwald and S. Vishwanath, "Space-time multiple access: Linear growth in the sum rate," in *Proc. Allerton*, pp. 387–396, 2002.
- [2] M. Sharif and B. Hassibi, "On the capacity of mimo broadcast channel with partial side information," *IEEE Trans. on Inform. Theory*, vol. 51, no. 2, pp. 506–522, 2005.
- [3] T. Y. Al-Naffouri, M. Sharif, and B. Hassibi, "How much does transmit correlation affect the sum-rate scaling of MIMO Gaussian broadcast channels?" *IEEE Trans. on Commun.*, vol. 57, no. 2, pp. 562–572, 2009.
- [4] M. Kountouris, D. Gesbert, and L. Pittman, "Transmit correlation-aided opportunistic beamforming and scheduling," 2006, pp. 2409–2414.
- [5] D. Hammarwall, M. Bengtsson, and B. E. Ottersten, "Utilizing the spatial information provided by channel norm feedback in sdma systems," *IEEE Trans. on Sig. Proc.*, vol. 56, no. 7, pp. 3278–3293, 2008.
- [6] J. Nam, J.-Y. Ahn, A. Adhikary, and G. Caire, "Joint spatial division and multiplexing: Realizing massive MIMO gains with limited channel state information," *CISS*, pp. 1–6, 2012.
- [7] A. Adhikary, J. Nam, J.-Y. Ahn, and G. Caire, "Joint spatial division and multiplexing – the large-scale array regime," *submitted to IEEE Trans. on Inform. Theory*, 2012. [Online]. Available: <http://arxiv.org/abs/1209.1402>
- [8] P. Viswanath and a. R. D.N.C.Tse, "Opportunistic beamforming using dumb antennas," *IEEE Trans. on Inform. Theory*, vol. 48, no. 6, pp. 1277–1294, 2002.
- [9] 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.
- [10] N. T. Uzgoren, "The asymptotic development of the distribution of the extreme values of a sample," in *Studies in Mathematics and Mechanics Presented to Richard von Mises*, pp. 346–353, 1954.
- [11] 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.
- [12] T. Yoo and A. Goldsmith, "On the optimality of multi-antenna broadcast scheduling using zero-forcing beamforming," *IEEE J. Select. Areas Commun.*, vol. 24, no. 3, pp. 528–541, 2006.