Block Diagonal Geometric Mean Decomposition (BD-GMD) for MIMO Broadcast Channels

Shaowei Lin, Winston W. L. Ho, and Ying-Chang Liang, *Senior Member, IEEE*

***Abstract*—In recent years, the research on multiple-input multiple-output (MIMO) broadcast channels has attracted much interest, especially since the discovery of the broadcast channel capacity achievable through the use of dirty paper coding (DPC). In this paper, we propose a new matrix decomposition, called the block diagonal geometric mean decomposition (BD-GMD), and develop transceiver designs that combine DPC with BD-GMD for MIMO broadcast channels. We also extend the BD-GMD to the block diagonal uniform channel decomposition (BD-UCD) with which the MIMO broadcast channel capacity can be achieved. Our proposed schemes decompose each user's MIMO channel into parallel subchannels with *identical* SNRs/SINRs, thus equal-rate coding can be applied across the subchannels of each user. Numerical simulations show that the proposed schemes demonstrate superior performance over conventional schemes.**

***Index Terms*—MIMO systems, broadcast channel (BC), uplink-downlink duality, equal-rate coding, geometric mean decomposition (GMD), block diagonal geometric mean decomposition (BD-GMD), block diagonal uniform channel decomposition (BD-UCD), decision feedback equalization (DFE), dirty paper coding (DPC), channel capacity.**

I. INTRODUCTION

MULTIPLE-INPUT multiple-output (MIMO) is a promising technology for next generation wireless communications systems. Relative to its single-input single-output (SISO) counterpart, MIMO techniques exhibit considerable improvements in reliability and throughput for communications systems [1], [2], [17].

For the single-user scenario, the singular value decomposition (SVD) can be used to generate multiple subchannels. However, the usually large condition number of the channel matrix results in subchannels with vastly different signal-to-noise-ratios (SNRs). From the information theoretical viewpoint, in order to achieve the channel capacity, variable-rate coding can be used across these decoupled subchannels. However, this may require special code design due to the variation of the supported rate of the subchannels. In practice, bit loading is designed to match the supported rates of each subchannel, i.e., each subchannel is allocated with one modulation/coding scheme (MCS). By doing so, however, the good subchannels may require very high order QAMs, which may be difficult to implement for wireless systems due to the existence of phase noises and synchronization errors. On the other hand, if the same MCS is used on all the subchannels, the BER performance will be dominated by the weakest subchannel. Recently, geometric mean decomposition (GMD) [5] has been proposed to resolve this problem by decomposing the MIMO channel into multiple subchannels with identical SNRs, with the help of dirty paper coding (DPC) [25] or interference cancellation. With that, the same MCS can be applied across all the subchannels, thus the order of the required QAM can be relatively small due to the sharing among all subchannels. Therefore, the GMD-based scheme is useful when there is a limitation on the order of QAM being used in practical system design.

Multiuser communications is an area of intense research, especially since the recent discovery of the MIMO broadcast channel capacity [7], [8], [9], [10]. It has been shown that when the channel state information (CSI) is available at the transmitter, the capacity region of the MIMO broadcast channel is exactly equal to the capacity region of its dual uplink multiple access channel. While the uplink capacity is achievable by the minimum mean squared error (MMSE) decision feedback equalizer (DFE) [16], the broadcast channel capacity is likewise achievable via DPC. In multiuser broadcast channels, unlike the single-user MIMO, cooperation is not permitted between different users. However, receive equalization is possible for each user. There have been several schemes proposed for low complexity MIMO broadcast communications. These include schemes based on TDMA [19], linear block diagonalization [20], [21], [22], and random / opportunistic beamforming [23], [24]. However, these schemes would normally fail to achieve the broadcast channel capacity.

In this paper, the single-user GMD is extended to the multiuser MIMO broadcast scenario. Our design criteria is that for each user, subchannels with identical SNRs are to be created. To achieve this, a new matrix decomposition, called the block diagonal geometric mean decomposition (BD-GMD), is first proposed. Four applications of the BD-GMD are then designed. The first two applications are low-complexity zero-forcing (ZF)-based schemes, while the later two are MMSE-based schemes of higher complexity. All four of them use DPC at the transmitter. Tomlinson-Harashima precoding (THP) [26], [27], a simple suboptimal implementation of DPC, can also be used in conjunction with these new schemes. In the first application, a BD-GMD-based DPC scheme is designed, that generates subchannels with identical SNRs for each user, while different rates may be experienced between different users. Such a scheme is said to be *block-equal-rate*. In the second application, a scheme that allows equal-rate coding among the subchannels of *all* users is constructed. Such a scheme is said to be *equal-rate*. Power control via channel inversion is shown to optimize the achievable sum-rate. The

Manuscript received February 10, 2007; revised June 21, 2007; accepted September 18, 2007. The associate editor coordinating the review of this paper and approving it for publication was S. Aissa.

S. Lin is with the University of California, Berkeley, USA.

W. W. L. Ho and Y.-C. Liang are with the Institute for Infocomm Research, 21 Heng Mui Keng Terrace, Singapore 119613 (e-mail: ycliang@i2r.a-star.edu.sg).

Digital Object Identifier 10.1109/TWC.2008.070171.

resulting scheme is called the Equal-Rate BD-GMD. User ordering is proposed to improve the fairness of the BD-GMD-based DPC scheme or to increase the achievable sum-rate for the Equal-Rate BD-GMD scheme.

ZF-based schemes are inherently capacity lossy. In the third application, using the uplink-downlink duality [7], [8], the BD-GMD is extended to the block diagonal uniform channel decomposition (BD-UCD) which achieves the sum-rate capacity for the MIMO broadcast channel. Specifically, Jindal *et al.*'s sum-power iterative water-filling algorithm [12] is used to obtain the optimal transmit power allocation and precoder for the scheme. In the fourth application, an equal-rate scheme called the Equal-Rate BD-UCD is proposed. This is done by first considering the problem of uplink beamforming under signal-to-interference-plus-noise-ratio (SINR) constraints [13] for mobile users with multiple antennas. An efficient algorithm which finds a near-optimal solution is then given, and duality is applied to produce the desired scheme.

The first two applications are ZF extensions to the GMD scheme, while the later two applications can be considered as MMSE extensions to the uniform channel decomposition (UCD) scheme [6]. For each of the ZF and MMSE cases, the first scheme proposed is a block-equal-rate scheme, while the second is an equal-rate scheme. For the equal-rate schemes, besides the fairness achieved among the different users, by providing each user with the same rate and if THP is used as a suboptimal DPC technique, the same modulus operator can be used for *all* subchannels across *all* the users. This obviates the need to implement multiple modulus operators to match the different modulation schemes used by the various users. Simulation results show that our proposed schemes exhibit superior performance over existing schemes.

The paper is organized as follows. The MIMO broadcast channel model and capacity results are presented in Section II. A review of conventional transceiver schemes, MMSE-DFE, MMSE-DPC, and ZF-DPC is also given in this section. In Section III, the block diagonal geometric mean decomposition is proposed. In Sections IV-A and IV-B, the ZF-based schemes, BD-GMD and Equal-Rate BD-GMD, are designed. Section V-A details the application of the uplink-downlink duality result to the MIMO broadcast situation. Sections V-B and V-C lay out the design of the MMSE-based schemes, BD-UCD and Equal-Rate BD-UCD. Simulation results are presented and discussed in Section VI, and a conclusion is given in Section VII.

The following notations are used in the paper. The boldface is used to denote matrices and vectors, and $\mathrm{E}[\cdot]$ for expectation. Let $\mathrm{Tr}(\mathbf{X})$, $\mathbf{X}^T$, $\mathbf{X}^H$ and $\mathbf{X}^{-1}$ denote the matrix trace, transpose, conjugate transpose and inverse, respectively, for a matrix $\mathbf{X}$. $[\mathbf{X}]_{i,j}$ will usually denote the matrix element at the i-th row and j-th column, unless otherwise stated. $\mathrm{diag}(\mathbf{X}_1, \ldots, \mathbf{X}_n)$ denotes the block diagonal matrix with diagonal elements $\mathbf{X}_1, \ldots, \mathbf{X}_n$, while $\mathrm{diag}(\mathbf{X})$ represents the diagonal matrix with the same diagonal as a matrix $\mathbf{X}$. $\mathcal{U}(\mathbf{X})$ and $\mathcal{L}(\mathbf{X})$ denote the upper and lower triangular matrices, respectively, formed using the matrix $\mathbf{X}$. $\|\cdot\|$ denotes the vector Euclidean norm, and $|x|$ the absolute value of a complex number x.

II. Channel Model and Preliminaries

Given an infrastructure based system with one base station (BS) and K mobile users, consider the *broadcast channel* from the BS to the mobile users. The BS is equipped with N_T antennas, and the i-th mobile user has n_i antennas. Let $N_R = \sum_{i=1}^{K} n_i$ be the total number of receive antennas, where $N_R \leq N_T$. Denote this channel by $N_T \times \{n_1, \ldots, n_K\}$. The input-output relation can be represented as

$$\mathbf{y} = \mathbf{H}\mathbf{x} + \mathbf{u} \ , \tag{1}$$

where $\mathbf{x}$ is the $N_T \times 1$ transmit signal vector at the BS, $\mathbf{y}$ the $N_R \times 1$ receive signal vector with $\mathbf{y} = [\mathbf{y}_1^T, \cdots, \mathbf{y}_K^T]^T$, and each $\mathbf{y}_i$ the $n_i \times 1$ receive signal vector of user i. Assume that the noise vector $\mathbf{u}$ is a zero-mean, circularly symmetric complex Gaussian (CSCG) vector with $\mathrm{E}[\mathbf{u}\mathbf{u}^H] = N_0\mathbf{I}$, and $\mathbf{u}$ is independent of $\mathbf{x}$. Assume also that $\mathrm{E}[\|\mathbf{x}\|^2] = E_s$ and $\det(\mathbf{H}\mathbf{H}^H) \neq 0$. Let $\rho = E_s/N_0$ be the SNR. It will also be useful to write $\mathbf{H} = [\mathbf{H}_1^T, \mathbf{H}_2^T, \ldots, \mathbf{H}_n^T]^T$, where $\mathbf{H}_i$ is the $n_i \times N_T$ channel matrix of user i.

If $\mathbf{x}$ is a Gaussian random vector, the sum-capacity of this broadcast channel [7], [8], [9], [10] is given by

$$\sup_{\mathrm{Tr}(\mathbf{F}\mathbf{F}^H) \leq E_s} \log\det\left(\mathbf{I} + \frac{1}{N_0}\mathbf{H}^H\mathbf{F}\mathbf{F}^H\mathbf{H}\right) \ , \tag{2}$$

where $\mathbf{F}$ is a block diagonal matrix: $\mathbf{F} = \mathrm{diag}(\mathbf{F}_1, \cdots, \mathbf{F}_K)$, where each block $\mathbf{F}_i$ is a $n_i \times n_i$ matrix. This sum-capacity can be achieved by a scheme that combines DPC with linear pre-equalization. A fundamental step in proving this theorem is showing that there is a dual relationship between uplink DFE and downlink DPC schemes. This uplink-downlink duality result will be discussed in greater detail in Section V-A. In [11], the authors extended the capacity theorem by verifying that the capacity *region* of the Gaussian MIMO broadcast channel is precisely the DPC rate region.

We now review some of the basic transceivers for point-to-point MIMO and broadcast channels, the results of which will be used in our transceiver design for MIMO broadcast channels.

A. MMSE-DFE

Consider the $N_T \times N_R$ point-to-point channel $\mathbf{y} = \mathbf{H}\mathbf{x} + \mathbf{u}$ where $\mathrm{E}[\mathbf{x}\mathbf{x}^H] = (E_s/N_T)\mathbf{I}$ and $\mathrm{E}[\mathbf{u}\mathbf{u}^H] = N_0\mathbf{I}$. The MMSE-based DFE can be represented by the block diagram in Figure 1. The columns $\mathbf{w}_i$ of its nulling matrix $\mathbf{W}$ are given by

$$\mathbf{w}_i = \left(\sum_{j=1}^{i} \mathbf{h}_j\mathbf{h}_j^H + \eta\mathbf{I}\right)^{-1} \mathbf{h}_i \ , \quad 1 \leq i \leq N_T \tag{3}$$

where $\eta = N_0(N_T/E_s)$ and $\mathbf{H} = [\mathbf{h}_1, \ldots, \mathbf{h}_{N_T}]$. It applies successive interference cancelation (SIC) via the feedback matrix $\mathbf{B} - \mathbf{I}$, where $\mathbf{B}$ is referred to as the interference matrix and is given by

$$[\mathbf{B}]_{i,j} = \begin{cases} [\mathbf{W}^H\mathbf{H}]_{i,j} & \text{if } i < j, \\ [\mathbf{I}]_{i,j} & \text{otherwise.} \end{cases} \tag{4}$$

Denote this monic upper triangular matrix by $\mathbf{B} = \mathcal{U}(\mathbf{W}^H\mathbf{H})$. For convenience, also define $\mathcal{L}(\mathbf{X}) = \mathcal{U}(\mathbf{X}^H)^H$ for any square

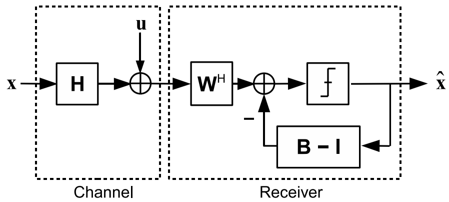

Fig. 1. Block diagram of the MMSE-DFE scheme.

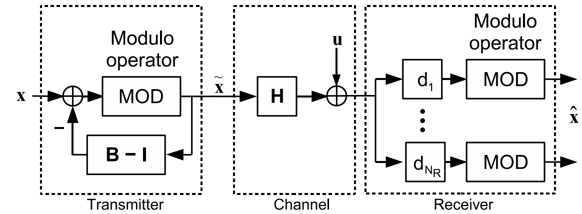

Fig. 2. Block diagram of the MMSE-DPC scheme using THP.

matrix $\mathbf{X}$. Now, alternatively, the nulling and interference matrices can be found via the QR-decomposition [3]

$$\begin{bmatrix} \mathbf{H} \\ \sqrt{\eta}\mathbf{I} \end{bmatrix} = \mathbf{Q}_1\mathbf{R}_1 \ , \quad \mathbf{Q}_1 = \begin{bmatrix} \mathbf{Q}_{1u} \\ \mathbf{Q}_{1d} \end{bmatrix} \ , \quad \mathbf{R}_1 = \mathbf{\Lambda}_1\mathbf{B}_1 \ , \tag{5}$$

where $\mathbf{Q}_1$ has orthonormal columns, $\mathbf{Q}_{1u}$ is $N_R \times N_T$, $\mathbf{Q}_{1d}$ is $N_T \times N_T$, $\mathbf{R}_1$ is $N_T \times N_T$ and upper triangular, $\mathbf{\Lambda}_1 = \mathrm{diag}(\mathbf{R}_1)$, and $\mathbf{B}_1$ is upper triangular with unit diagonal. (Note that $\mathbf{Q}_{1u}$ and $\mathbf{Q}_{1d}$ are not unitary.) Then, the nulling and interference matrices satisfy

$$\mathbf{W}^H = \mathbf{\Lambda}_1^{-1}\mathbf{Q}_{1u}^H \quad \text{and} \quad \mathbf{B} = \mathbf{B}_1 \ . \tag{6}$$

The symbols are detected from $\hat{x}_{N_T}$ to $\hat{x}_1$ as follows:

$\texttt{for}\ i = N_T : -1 : 1$
$\quad \hat{x}_i = \mathcal{C}\left[[\mathbf{W}^H\mathbf{y}]_i - \sum_{j=i+1}^{N_T}[\mathbf{B}]_{i,j}\hat{x}_j\right]$
$\texttt{end}$

where $\mathcal{C}$ denotes the mapping to the nearest signal point in the constellation. Ignoring the effect of error propagation, the MMSE-DFE scheme produces decoupled subchannels of the form $y_i = r_i x_i + u_i$ where r_i is the i-th diagonal element of $\mathbf{\Lambda}_1$. In [6], it was shown that

$$\eta(1+\rho_i) = r_i^2 \ , \tag{7}$$

where ρ_i is the SINR of the i-th subchannel. Thus, the capacity of the scheme can be written as

$$\sum_{i=1}^{N_T}\log(1+\rho_i) = \sum_{i=1}^{N_T}\log\left(\frac{r_i^2}{\eta}\right) = \log\det\left(\mathbf{I} + \frac{1}{\eta}\mathbf{H}^H\mathbf{H}\right) \ . \tag{8}$$

This gives another proof that the MMSE-DFE receiver is information lossless [16], [6].

B. MMSE-based DPC

One major problem with DFEs is error propagation. If CSI is known at the transmitter, interference between subchannels can be cancelled completely before transmission via DPC. Here, a general view of MMSE-based DPC via successive interference pre-subtraction is developed. Consider once again the $N_T \times N_R$ point-to-point channel $\mathbf{y} = \mathbf{H}\mathbf{x} + \mathbf{u}$ from Section II-A. However, it will not be required that $\mathrm{E}[\mathbf{u}\mathbf{u}^H] = N_0\mathbf{I}$ but only that $\mathrm{E}[|u_i|^2] = N_0$ for each i. Assume that there is no collaboration between the receive antennas. Writing $h_{ij} = [\mathbf{H}]_{i,j}$, the i-th subchannel is

$$y_i = (\sum_{j<i} h_{ij}x_j) + h_{ii}x_i + (\sum_{j>i} h_{ij}x_j) + u_i \ . \tag{9}$$

The hope is to treat $(\sum_{j<i} h_{ij}x_j)$ as interference terms to be cancelled at the transmitter. If these interference terms are cancelled perfectly, then a single input single output (SISO) MMSE receiver that sees $(\sum_{j>i} h_{ij}x_j) + u_i$ as noise terms can be used on each subchannel. The corresponding MMSE coefficient for the i-th subchannel is

$$d_i = \frac{h_{ii}^*}{\eta + \sum_{j\geq i}|h_{ij}|^2} \ , \tag{10}$$

where x^* is the conjugate of a complex number x. Denoting $\mathbf{D}_d = \mathrm{diag}(d_1, \ldots, d_{N_R})$, the equivalent channel is now $\mathbf{D}_d\mathbf{H}$. Thus, the interference terms can be represented by the lower triangular unit-diagonal matrix $\mathbf{B} = \mathcal{L}(\mathbf{D}_d\mathbf{H})$, called the interference matrix. Meanwhile, the SINR of the i-th subchannel is given by

$$\rho_i = \frac{|h_{ii}|^2}{\eta + \sum_{j>i}|h_{ij}|^2} \ . \tag{11}$$

A simple and useful relation between (10) and (11) can be noted at this point. Let $\Sigma_0 = \eta + \sum_{j>i}|h_{ij}|^2$. Then, $\rho_i = |h_{ii}|^2/\Sigma_0$ and $d_i = h_{ii}^*/(\Sigma_0 + |h_{ii}|^2)$. Eliminating Σ_0 gives

$$d_i = \frac{\rho_i}{h_{ii}(1+\rho_i)} \ . \tag{12}$$

One low-complexity suboptimal implementation of DPC is Tomlinson-Harashima precoding (THP) [14]. The block diagram of a MMSE-based DPC scheme using THP is shown in Figure 2. The vector $\tilde{\mathbf{x}}$ to be transmitted can be evaluated from $\tilde{x}_1$ to $\tilde{x}_{N_T}$ using

$\tilde{x}_1 = x_1$
$\texttt{for}\ i = 2 : 1 : N_T$
$\quad \tilde{x}_i = \mathrm{mod}\left[x_i - \sum_{j=1}^{i-1}[\mathbf{B}]_{i,j}\tilde{x}_j\right]$
$\texttt{end}$

A downside of THP is the slight increase in the average transmit power by a factor of $M/(M-1)$ for M-QAM symbols, called the *precoding loss*. For large constellations, this loss is negligible.

C. ZF-DPC

Consider the MIMO broadcast channel described in (1). If CSI is available at the transmitter, interference cancellation via dirty paper precoding can be performed. THP can be used as a suboptimal implementation of DPC. Conventional precoding schemes often treat multiple antennas of different users as different virtual users. One example is the zero-forcing THP (ZF-THP) scheme [14]. It is based on the QR decomposition $\mathbf{H}^H = \mathbf{Q}\mathbf{R}$, or $\mathbf{H} = \mathbf{R}^H\mathbf{Q}^H$. The linear precoder $\mathbf{Q}$ is applied before transmission so that $\mathbf{x} = \mathbf{Q}\mathbf{s}$, where $\mathbf{s}$ is the vector of information symbols to be sent. This transforms the channel to $\mathbf{y} = \mathbf{R}^H\mathbf{s} + \mathbf{u}$, on which THP is applied to pre-subtract the interference represented by the lower triangular matrix $\mathbf{R}^H$. Thus, N_R decoupled subchannels $y_i = r_i s_i + u_i$, where r_i is the ith diagonal element of $\mathbf{R}$, are obtained.

Another example comes from [15]. The authors considered pre-equalization matrices $\mathbf{F}$ such that the resulting channel matrix $\mathbf{HF}$ is lower triangular and has diagonal elements all equal to a certain value, say r. Here, $\mathbf{F}$ need not be unitary but only has to satisfy the power constraint $\mathrm{Tr}(\mathbf{FF}^H) \leq E_s$. The precoder $\mathbf{F}$ that maximizes r in $\mathbf{HF}$ can be found algorithmically. The scheme now only needs to perform THP to cancel the interference represented by the lower triangular matrix $\mathbf{HF}$ before pre-equalizing with $\mathbf{F}$ and transmitting the signal. This scheme generates N_R decoupled subchannels $y_i = rs_i + u_i$ on which equal-rate coding can be applied. Hence, their scheme will be referred to as the Equal-Rate ZF-THP scheme.

III. Block Diagonal Geometric Mean Decomposition

In this section, the block diagonal geometric mean decomposition (BD-GMD) is developed. We begin by describing the motivation for developing such a decomposition. In the case where some of the mobile users have multiple antennas, performance gain can be expected from doing equalization on the receiver side. This equalization can only be done for the data streams of the same user, and not between users. It can be represented as a premultiplication of the receive signal $\mathbf{y}(n)$ by a block diagonal matrix $\mathbf{A}$: $\mathbf{A} = \mathrm{diag}(\mathbf{A}_1, \cdots, \mathbf{A}_K)$, where each block $\mathbf{A}_i$ is the $n_i \times n_i$ equalization matrix of user i. Each row of $\mathbf{A}_k$ is assumed to be of unit norm so that the noise vector $\mathbf{u}$ is not amplified by $\mathbf{A}$. Note that the situation where the mobile users have single antennas is represented by the case $\mathbf{A} = \mathbf{I}$. Equalization by $\mathbf{A}$ gives the equivalent channel matrix $\mathbf{AH}$ which can then be cancelled by transmitter techniques such as ZF-THP or MMSE-THP.

Suppose ZF-THP is used at the transmitter, and the QR-decomposition $\mathbf{AH} = \mathbf{R}^H\mathbf{Q}^H$ is considered. Since the aim is to construct a block-equal-rate scheme, it is natural to ask if it can be accomplished by choosing an appropriate equalization matrix $\mathbf{A}$. To simplify the problem, assume further that $\mathbf{A}$ is unitary. Our problem can now be stated as follows. Let $\mathbf{H}$ be an $N_R \times N_T$ matrix, and $n_1, ..., n_K$ a sequence of integers such that $N_R = \sum_{i=1}^K n_i$. Consider matrix decompositions of the following form [5]: $\mathbf{H} = \mathbf{PLQ}^H$, where $\mathbf{Q}$ is a $N_T \times N_R$ matrix with orthonormal columns, $\mathbf{L}$ is a $N_R \times N_R$ lower triangular matrix, and $\mathbf{P}$ is a $N_R \times N_R$ block diagonal matrix of the form $\mathrm{diag}(\mathbf{P}_1, \mathbf{P}_2, \ldots, \mathbf{P}_K)$ where each block $\mathbf{P}_i$ is a unitary $n_i \times n_i$ matrix. The task is to find a matrix decomposition such that the diagonal elements of $\mathbf{L}$ are equal in blocks of $n_1, \ldots, n_K$ elements respectively.

A. Proposed Algorithm

The algorithm to solve the above problem is as follows. Write the product $\mathbf{H} = \mathbf{PLQ}^H$ as

$$\begin{bmatrix} \mathbf{H}_1 \\ \boldsymbol{\mathcal{H}} \end{bmatrix} = \begin{bmatrix} \mathbf{P}_1 & \mathbf{0} \\ \mathbf{0} & \boldsymbol{\mathcal{P}} \end{bmatrix} \begin{bmatrix} \mathbf{L}_1 & \mathbf{0} \\ \boldsymbol{\Xi} & \boldsymbol{\mathcal{L}} \end{bmatrix} \begin{bmatrix} \mathbf{Q}_1^H \\ \boldsymbol{\mathcal{Q}}^H \end{bmatrix}, \quad (13)$$

where $\mathbf{H}_1$ and $\mathbf{Q}_1^H$ are $n_1 \times N_T$ submatrices, and $\mathbf{L}_1$ and $\mathbf{P}_1$ are $n_1 \times n_1$ square matrices. $\boldsymbol{\mathcal{H}}$ denotes the combined channel matrix of all the remaining users. Expanding (13) gives the following two equations

$$\mathbf{H}_1 = \mathbf{P}_1\mathbf{L}_1\mathbf{Q}_1^H \ , \quad (14)$$

$$\boldsymbol{\mathcal{H}} = \boldsymbol{\mathcal{P}}\boldsymbol{\Xi}\mathbf{Q}_1^H + \boldsymbol{\mathcal{P}}\boldsymbol{\mathcal{L}}\boldsymbol{\mathcal{Q}}^H \ . \quad (15)$$

From equation (14), it can be seen that by using the GMD, the diagonal elements of $\mathbf{L}_1$ can be made equal. Now, since $\mathbf{Q}$ has orthonormal columns, the submatrices $\mathbf{Q}_1$ and $\boldsymbol{\mathcal{Q}}$ are orthonormal to each other. Thus, from equation (15), multiplication by the projection matrix $\mathbf{I} - \mathbf{Q}_1\mathbf{Q}_1^H$ gives

$$\boldsymbol{\mathcal{H}}(\mathbf{I} - \mathbf{Q}_1\mathbf{Q}_1^H) = \boldsymbol{\mathcal{P}}\boldsymbol{\mathcal{L}}\boldsymbol{\mathcal{Q}}^H \ . \quad (16)$$

Here, the right side of (16) has the same form as (13), so the algorithm proceeds recursively. Finally, to solve for $\boldsymbol{\Xi}$, equation (15) is multiplied by $\boldsymbol{\mathcal{P}}^H$ and $\mathbf{Q}_1$, giving

$$\boldsymbol{\Xi} = \boldsymbol{\mathcal{P}}^H\boldsymbol{\mathcal{H}}\mathbf{Q}_1 \ . \quad (17)$$

The decomposition that achieves equal diagonal elements in each block of $\mathbf{L}$ will be referred to as the block diagonal geometric mean decomposition (BD-GMD).

B. Diagonal Elements

Consider a BD-GMD decomposition $\mathbf{H} = \mathbf{PLQ}^H$. Let the diagonal element of the i-th block of $\mathbf{L}$ be r_i. To calculate each r_i, equations (13) and (14) are generalized to get

$$\begin{bmatrix} \widehat{\mathbf{H}}_i \\ \boldsymbol{\mathcal{H}} \end{bmatrix} = \begin{bmatrix} \widehat{\mathbf{P}}_i & \mathbf{0} \\ \mathbf{0} & \boldsymbol{\mathcal{P}} \end{bmatrix} \begin{bmatrix} \widehat{\mathbf{L}}_i & \mathbf{0} \\ \boldsymbol{\Xi} & \boldsymbol{\mathcal{L}} \end{bmatrix} \begin{bmatrix} \widehat{\mathbf{Q}}_i^H \\ \boldsymbol{\mathcal{Q}}^H \end{bmatrix} \quad (18)$$

$$\widehat{\mathbf{H}}_i = \widehat{\mathbf{P}}_i\widehat{\mathbf{L}}_i\widehat{\mathbf{Q}}_i^H \ , \quad (19)$$

where $\widehat{\mathbf{H}}_i$ represents the combined channel matrix of users 1 to i. The submatrices $\widehat{\mathbf{H}}_i$, $\widehat{\mathbf{P}}_i$, $\widehat{\mathbf{L}}_i$ and $\widehat{\mathbf{Q}}_i^H$ each have $\sum_{j=1}^i n_j$ rows. Because $\widehat{\mathbf{P}}_i$ and $\widehat{\mathbf{Q}}_i$ have orthonormal columns, equation (19) shows that the singular values of $\widehat{\mathbf{H}}_i$ and $\widehat{\mathbf{L}}_i$ must be the same. Thus,

$$\det(\widehat{\mathbf{H}}_i\widehat{\mathbf{H}}_i^H) = \det(\widehat{\mathbf{L}}_i\widehat{\mathbf{L}}_i^H) = \prod_{j=1}^{i} r_j^{2n_j} \ . \quad (20)$$

Therefore, the i-th diagonal element is given by

$$r_i = \sqrt[2n_i]{\frac{\det(\widehat{\mathbf{H}}_i\widehat{\mathbf{H}}_i^H)}{\det(\widehat{\mathbf{H}}_{i-1}\widehat{\mathbf{H}}_{i-1}^H)}} \ . \quad (21)$$

IV. ZF-based Schemes

The previous section introduced the new decomposition for multiuser communications. In this section, transceiver designs based on this decomposition would be described. The following two schemes are ZF-based schemes. This means that they are of a relatively low complexity, and offer a reasonable tradeoff between performance and efficiency.

A. BD-GMD-based DPC Scheme

Having more than one user means that different decompositions can be obtained by changing the encoding order of the users. As a result, different sets of values for the diagonal elements of $\mathbf{L}$ can be obtained. Let $\{\pi_1, \pi_2, \ldots, \pi_K\}$ be the ordering of the users where the previous π_1-th user is now the first user, and so on. Since the ordering of the users result in the ordering of the rows of $\mathbf{H}$, this ordering may also be represented by a permutation matrix $\mathbf{D}$ such that $\mathbf{DH} = \mathbf{PLQ}^H$. Here, the i-th block $\mathbf{P}_i$ of $\mathbf{P}$ has dimensions $n_{\pi_i} \times n_{\pi_i}$. Suppose $\mathbf{s} = [\mathbf{s}_1^T, \cdots, \mathbf{s}_K^T]^T$ is the vector of information symbols to be sent, where $\mathbf{s}_i$ is the corresponding $n_i \times 1$ vector of user i. Write $\mathbf{L} = \mathbf{\Lambda B}$ with $\mathbf{\Lambda} = \mathrm{diag}(\mathbf{L})$, and $\mathbf{B}$ a lower triangular matrix with unit diagonal. Multiplying (1) by $\mathbf{D}$ gives

$$\widetilde{\mathbf{y}} = \mathbf{PLQ}^H \mathbf{x} + \widetilde{\mathbf{u}} \ , \tag{22}$$

where $\widetilde{\mathbf{y}}$ is the reordered received signal vector. Let $\widetilde{\mathbf{s}} = \mathbf{Ds}$ be the reordered information symbol vector. Using $\mathbf{x} = \mathbf{Q}\widetilde{\mathbf{s}}$ and $\widetilde{\mathbf{z}} = \mathbf{P}^H \widetilde{\mathbf{y}}$ for the transmit and receive equalization respectively transforms the channel to

$$\widetilde{\mathbf{z}} = \mathbf{L}\widetilde{\mathbf{s}} + \widetilde{\mathbf{u}}' \ . \tag{23}$$

Now, DPC is performed at the transmitter to pre-subtract the interference represented by $\mathbf{L}$. As a result, user π_i enjoys n_{π_i} independent and equivalent subchannels of the form: $z_i = r_i s_i + u_i$, where r_i is the diagonal element of the i-th block of $\mathbf{L}$. The available transmit power E_s is distributed equally among the N_R subchannels. Then, the achievable sum-rate for the scheme is given by

$$C = \sum_{i=1}^{K} n_{\pi_i} \log_2 \left(1 + \frac{E_s}{N_0 N_R} r_i^2\right) . \tag{24}$$

Figure 3 shows the block diagram of the BD-GMD-based scheme that performs user ordering and uses THP instead of DPC. Here, $\mathbf{P}_i^H$ is the i-th sub-block of the block diagonal unitary matrix $\mathbf{P}^H$. To improve the performance of the scheme, different constellations can be applied to the users. The base station and mobile users decide *a priori* on a fixed set of constellations to use. Before data transmission over a block period of time, the BS informs each user which constellation to apply, depending on the performance of their channels. The user then uses the same constellation for all his subchannels since the subchannels have identical SNRs. In certain scenarios, it may be better to use a modified form of the BD-GMD that involves subchannel selection. The reason is similar to that of the subchannel selection principle for the single-user GMD [5]. Additionally, in the case of multiuser communications, subchannel selection for one user may actually provide a better performance for later encoded users, due to the provision of more spatial degrees of freedom. For conciseness and clarity, subchannel selection will not be further developed in this paper. However, this is an interesting concept that could be considered in future works.

1) Ordering for the BD-GMD-based Scheme: From Section II-C, it is seen that the i-th diagonal element corresponds to the channel gain of the i-th subchannel. Here, assume that $\mathbf{H}$ contains i.i.d. Gaussian entries. Now, the diagonal elements of $\mathbf{L}$ will usually be in decreasing order, because $\mathbf{LQ}^H$ is the QR-decomposition of $\mathbf{P}^H\mathbf{H}$. The first diagonal element can often be many times that of the last one. Thus, the first subchannel usually enjoys much better performance than the last subchannel. Equation (21) tells us that the diagonal elements r_i depend on the ordering of the rows of $\mathbf{H}$ or, in other words, the ordering of the users. The hope is to improve the fairness of the scheme by ordering the users to increase the size of the last few diagonal elements.

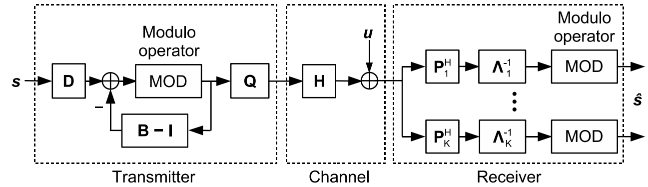

Fig. 3. Block diagram of BD-GMD-based scheme with user ordering and THP.

A method inspired by the BLAST [17] ordering is used. First, note that $\det(\widehat{\mathbf{H}}_{i-1}\widehat{\mathbf{H}}_{i-1}^H)$ does not change with the order of the rows of $\widehat{\mathbf{H}}_{i-1}$. Thus, from (21), it is seen that the value of r_K depends only on the choice of $\mathbf{H}_K$ and not the order of the first $K-1$ users. Therefore, to maximize r_K, $\mathbf{H}_K$ is chosen to minimize (21). Following that, $\mathbf{H}_{K-1}$ is chosen to minimize (21), and so on. The decomposition that optimizes the diagonal elements in such a way is called the *ordered* BD-GMD. For the BD-GMD-based scheme, this ordering tends to increase the user fairness.

B. Equal-Rate BD-GMD Scheme

While BD-GMD achieves equal rates for the sub-channels of each user, the achievable rates between different users vary greatly. In this section, equal rates for all subchannels across all the users is achieved by relying on optimal transmit power control. This is the Equal-Rate BD-GMD scheme. It maximizes the sum-rate while ensuring overall equal rates.

Following the example of Section IV-A, we will now construct a ZF-based scheme which maximizes the achievable sum-rate given overall equal rates. Let $\mathbf{F}$, $\mathbf{J}$ and $\mathbf{A}$ be the pre-equalization, interference and receive equalization matrices respectively of a general equal-rate ZF-based scheme. Let α be the channel gain which is identical for all the subchannels. To optimize the achievable sum-rate, we only need to maximize the channel gain:

$$\begin{aligned} \texttt{maximize} \quad & \alpha \\ \texttt{subject to} \quad & \mathbf{J} \in \mathbb{L} \ , \ \mathbf{A} \in \mathbb{B} \ , \\ & \mathbf{AHF} = \alpha \mathbf{J} \ , \quad \mathrm{Tr}(\mathbf{F}^H\mathbf{F}) \leq E_s \ , \\ & \|\mathbf{A}(i,:)\| = 1 \qquad \text{for } 1 \leq i \leq N_R \ . \end{aligned} \tag{25}$$

where $\mathbb{L}$ is the set of all monic lower triangular matrices and $\mathbb{B}$ is the set of all block diagonal matrices in which the i-th block is a $n_i \times n_i$ submatrix. The solution of this problem is by channel inversion, as the following theorem shows.

Theorem 1: Suppose $\det(\mathbf{HH}^H) \neq 0$. Let $\mathbf{H} = \mathbf{PLQ}^H$ be the BD-GMD of $\mathbf{H}$, and let $\mathbf{\Lambda} = \mathrm{diag}(\mathbf{L})$. Then, the

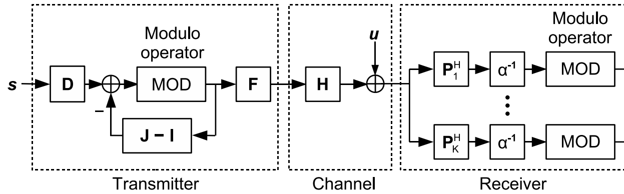

Fig. 4. Block diagram of equal-rate BD-GMD scheme with user ordering and THP.

optimization problem (25) is solved by

$$\mathbf{F} = \alpha \mathbf{Q}\mathbf{\Lambda}^{-1}, \quad \mathbf{J} = \mathbf{L}\mathbf{\Lambda}^{-1},$$
$$\mathbf{A} = \mathbf{P}^H, \quad \alpha = \sqrt{\frac{E_s}{\mathrm{Tr}(\mathbf{\Lambda}^{-2})}} \ . \tag{26}$$

Proof: See Appendix A. ■

Using the pre-equalization and receive equalization matrices given in Theorem 1, the channel becomes $\mathbf{z} = \alpha\mathbf{J}\mathbf{s} + \mathbf{u}'$. DPC is then done at the transmitter to cancel the interference. As a result, every user enjoys independent and equivalent subchannels of the form $z = \alpha s + u$. The achievable sum-rate for the scheme is given by

$$C = N_R \log_2(1 + {N_0}^{-1}\alpha^2) \ . \tag{27}$$

Given a fixed order of users, Theorem 1 tells us that the above Equal-Rate BD-GMD scheme optimizes the zero-forcing linear beamforming vectors and power allocation required to achieve maximum throughput and equal rates for every subchannel of every user.

1) Ordering for the Equal-Rate BD-GMD Scheme: Some improvement in performance can be expected by ordering the users appropriately. From (26), we have

$$\alpha^2 = \frac{E_s}{\mathrm{Tr}(\mathbf{\Lambda}^{-2})} = \frac{E_s}{\sum_{i=1}^{K} n_i r_i^{-2}} \ , \tag{28}$$

where r_i is the diagonal element of the i-th block of $\mathbf{\Lambda}$. Hence, given the constraint that $\prod_{i=1}^{K} r_i^{2n_i} = \det(\mathbf{H}\mathbf{H}^H)$ is constant, (28) is maximized when the r_i's are equal. In general, the channel gain α increases when the spread among $\{r_i\}_{i=1}^K$ decreases. This can be accomplished by using the *ordered* BD-GMD which reduces the gap between the largest and smallest diagonal elements of $\mathbf{\Lambda}$. The ordering described in Section IV-A1 is near-optimal for maximizing the sum-rate, for the case where $\mathbf{H}$ contains i.i.d. Gaussian entries. The best ordering may be found by an exhaustive search. Figure 4 shows the block diagram of an Equal-Rate BD-GMD scheme that applies the ordered BD-GMD. As before, THP is used as a suboptimal implementation of DPC in the scheme.

V. MMSE-based Schemes and Uplink-Downlink Duality

The previous section described ZF-based schemes. In this section, two MMSE-based schemes are introduced. These schemes would have improved performance over the ZF-based schemes, but at the cost of a moderately higher complexity. At high SNRs, the sum rates of the ZF-based schemes approach that of the MMSE-based schemes. Before going into the transceiver designs, a brief description of the uplink-downlink duality is crucial in explaining the design of the MMSE-based schemes.

A. Uplink-Downlink Duality

This section explains the uplink-downlink duality that would be vital in converting the uplink solution to the downlink case, for the proposed schemes in sections V-B and V-C. We return to the $N_T \times \{n_1, \ldots, n_K\}$ MIMO broadcast channel described in Section II. The uplink-downlink duality results [7], [8] will now be used to construct a DPC scheme that consumes the same power and achieves the same rates as a given MMSE-DFE scheme. This DPC scheme is *dual* to the MMSE-DFE scheme.

First, suppose the following MMSE-DFE scheme is given. Consider the $\{n_1, \ldots, n_K\} \times N_T$ uplink channel

$$\mathbf{y} = \mathbf{H}^H\mathbf{x} + \mathbf{u} \ , \tag{29}$$

which has K mobile users with $n_1, \ldots, n_K$ transmit antennas respectively, and a BS with N_T receive antennas. Let $\mathrm{E}[\mathbf{u}\mathbf{u}^H] = N_0\mathbf{I}$, and $\mathrm{E}[\|\mathbf{x}\|^2] = E_s$. This channel is dual to the broadcast channel in Section II. Meanwhile, let each user i be equipped with a pre-determined linear precoder $\mathbf{F}_i$. Combine all the precoders in a block diagonal matrix $\mathbf{F}$, and consider a MMSE-DFE receiver at the BS. Using (5), the QR decomposition for the equivalent channel $\mathbf{H}^H\mathbf{F}$ is as follows:

$$\begin{bmatrix} \mathbf{H}^H\mathbf{F} \\ \sqrt{N_0}\mathbf{I} \end{bmatrix} = \begin{bmatrix} \mathbf{Q}_u \\ \mathbf{Q}_d \end{bmatrix} \mathbf{\Lambda}\mathbf{B} \ . \tag{30}$$

It implies that the nulling matrix is $\mathbf{W}^H = \mathbf{\Lambda}^{-1}\mathbf{Q}_u^H$, and that the interference matrix is $\mathbf{B}$. Normalize the columns of $\mathbf{F}$ by writing its i-th column as $\sqrt{p_i}\mathbf{f}_i$ where $\sqrt{p_i}$ is the norm and $\mathbf{f}_i$ a unit column vector. Since

$$\mathbf{x} = \mathbf{F}\mathbf{s} = \sum_{i=1}^{N_R} \sqrt{p_i}\mathbf{f}_i s_i \ , \tag{31}$$

p_i represents the power allocated to the i-th information symbol s_i. Thus, the total power is $\sum_{i=1}^{N_R} p_i = \mathrm{Tr}[\mathbf{F}\mathbf{F}^H] = E_s$. Also, normalize the columns of $\mathbf{W}$, writing the i-th column as $c_i\mathbf{w}_i$ where c_i is the norm. Here, c_i can be thought of as an MMSE weight similar to that in (10), since it scales the signal in the i-th subchannel. Assuming that the symbols are cancelled perfectly in the SIC, the SINR of the i-th subchannel is given by

$$\rho_i = \frac{p_i|\mathbf{w}_i^H\mathbf{H}^H\mathbf{f}_i|^2}{N_0 + \sum_{j<i} p_j|\mathbf{w}_i^H\mathbf{H}^H\mathbf{f}_j|^2} \ . \tag{32}$$

Now, construct the dual DPC scheme for the broadcast channel as follows. Let $\widetilde{\mathbf{F}}$ be the linear precoder at the BS, $\widetilde{\mathbf{B}}$ the interference matrix, and $\widetilde{\mathbf{W}}^H$ the combined nulling matrix of the mobile users. The block diagram for this scheme is shown in Figure 5, where $\widetilde{\mathbf{W}}_i^H$ is the i-th sub-block of the block diagonal matrix $\widetilde{\mathbf{W}}^H$. First, define the i-th column of $\widetilde{\mathbf{F}}$ to be $\sqrt{q_i}\mathbf{w}_i$. Here, q_i is an unknown representing the power allocated to the i-th information symbol. Next, define the i-th column of $\widetilde{\mathbf{W}}$ to be $d_i\mathbf{f}_i$ where d_i is an unknown MMSE

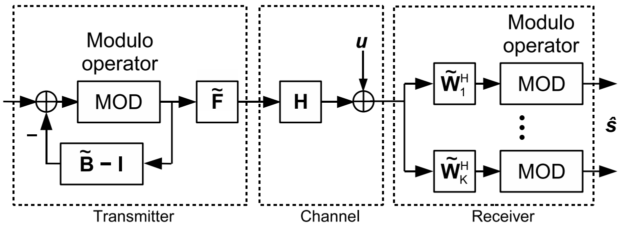

Fig. 5. Block diagram of dual DPC scheme.

weight. Since the goal is to achieve the same SINRs, by (11), the power coefficients q_i need to satisfy

$$\rho_i = \frac{q_i|\mathbf{f}_i^H\mathbf{H}\mathbf{w}_i|^2}{N_0 + \sum_{j>i} q_j|\mathbf{f}_i^H\mathbf{H}\mathbf{w}_j|^2} \quad \text{for } 1 \le i \le N_R \ . \tag{33}$$

Using the notations $\mathbf{q} = [q_1, \ldots, q_{N_R}]^T$, $\boldsymbol{\rho} = [\rho_1, \ldots, \rho_{N_R}]^T$ and $\alpha_{ij} = |\mathbf{f}_i^H\mathbf{H}\mathbf{w}_j|^2$, rewrite (33) in matrix form [6]:

$$\begin{bmatrix} \alpha_{11} & -\rho_1\alpha_{12} & \ldots & -\rho_1\alpha_{1N_R} \\ 0 & \alpha_{22} & \ldots & -\rho_2\alpha_{2N_R} \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \ldots & \alpha_{N_RN_R} \end{bmatrix} \mathbf{q} = N_0\boldsymbol{\rho} \ . \tag{34}$$

Thus, $\mathbf{q}$ can be derived. In [7], the authors showed that $\sum_{i=1}^{N_R} q_i = \sum_{i=1}^{N_R} p_i$, so both the MMSE-DFE scheme and the dual DPC scheme consume the same total power E_s.

To compute the MMSE weights d_i, the relation in (12) is exploited:

$$d_i = \frac{\rho_i}{(1+\rho_i)\sqrt{q_i}\mathbf{f}_i^H\mathbf{H}\mathbf{w}_i} \ . \tag{35}$$

By the same reasoning on c_i, the MMSE weights for the uplink channel, it follows that

$$c_i = \frac{\rho_i}{(1+\rho_i)\sqrt{p_i}\mathbf{w}_i^H\mathbf{H}^H\mathbf{f}_i} \ . \tag{36}$$

Since c_i is defined to be a vector norm, c_i is real so $\mathbf{w}_i^H\mathbf{H}^H\mathbf{f}_i$ is real. Consequently, $\mathbf{w}_i^H\mathbf{H}^H\mathbf{f}_i = \mathbf{f}_i^H\mathbf{H}\mathbf{w}_i$ and thus $d_i\sqrt{q_i} = c_i\sqrt{p_i}$. Denote $\mathbf{D}_q = \mathrm{diag}(\sqrt{q_1}, \ldots, \sqrt{q_{N_R}})$ and similarly define diagonal matrices $\mathbf{D}_p$, $\mathbf{D}_c$ and $\mathbf{D}_d$ for $\{\sqrt{p_i}\}$, $\{c_i\}$ and $\{d_i\}$. Thus, $\mathbf{D}_d\mathbf{D}_q = \mathbf{D}_c\mathbf{D}_p$. Finally, to complete the dual DPC scheme, the interference matrix $\widetilde{\mathbf{B}}$ is computed. Using the relations

$$\begin{aligned} \mathbf{B} = \mathcal{U}(\mathbf{W}^H\mathbf{H}^H\mathbf{F}) \ , \ \widetilde{\mathbf{W}} = \mathbf{F}\mathbf{D}_p^{-1}\mathbf{D}_d = \mathbf{F}\mathbf{D}_q^{-1}\mathbf{D}_c \ , \\ \widetilde{\mathbf{F}} = \mathbf{W}\mathbf{D}_c^{-1}\mathbf{D}_q \ , \ \mathcal{L}(\mathbf{X})^H = \mathcal{U}(\mathbf{X}^H) \ , \end{aligned} \tag{37}$$

the following result is derived:

$$\widetilde{\mathbf{B}} = \mathcal{L}(\widetilde{\mathbf{W}}^H\mathbf{H}\widetilde{\mathbf{F}}) = \mathbf{D}_c\mathbf{D}_q^{-1}\mathbf{B}^H\mathbf{D}_c^{-1}\mathbf{D}_q \ . \tag{38}$$

Using (7), the achievable sum-rate for both the MMSE-DFE and dual DPC scheme can be written as

$$\begin{aligned} \sum_{i=1}^{N_R}\log(1+\rho_i) &= \sum_{i=1}^{N_R}\log\left(\frac{\lambda_i^2}{N_0}\right) \\ &= \log\det(\mathbf{I} + \frac{1}{N_0}\mathbf{H}^H\mathbf{F}\mathbf{F}^H\mathbf{H}) \ . \end{aligned} \tag{39}$$

B. BD-UCD Scheme

In this section, a DPC scheme that is block-equal-rate and achieves the broadcast channel capacity is constructed. The idea is to first design a capacity-achieving MMSE-DFE scheme that generates subchannels with identical SINRs for each user by choosing an appropriate precoder $\mathbf{F}$. Then, the results from Section V-A gives us the dual DPC scheme. This BD-UCD scheme is capacity-achieving and ensures equal rates per block.

We begin with the dual uplink channel from Section V-A. Let $\bar{\mathbf{F}}$ be a linear precoder for this uplink channel that achieves its sum-capacity, i.e. $\bar{\mathbf{F}}$ solves the optimization problem in (2). Different methods of solving this problem are described in [12], including Jindal et al.'s sum-power iterative water-filling algorithm which will be used in this paper. Note that for any block diagonal unitary $\bar{\mathbf{P}}$, the precoder $\bar{\mathbf{F}}\bar{\mathbf{P}}$ gives the same sum-capacity as $\bar{\mathbf{F}}$. It remains for us to choose $\bar{\mathbf{P}}$ such that the MMSE-DFE scheme using the precoder $\mathbf{F} = \bar{\mathbf{F}}\bar{\mathbf{P}}$ is block-equal-rate. From (5), this is equivalent to finding $\bar{\mathbf{P}}$ such that the QR decomposition

$$\begin{bmatrix} \mathbf{H}^H\bar{\mathbf{F}}\bar{\mathbf{P}} \\ \sqrt{N_0}\mathbf{I} \end{bmatrix} = \begin{bmatrix} \mathbf{Q}_u \\ \mathbf{Q}_d \end{bmatrix} \boldsymbol{\Lambda}\mathbf{B} \tag{40}$$

gives a $\boldsymbol{\Lambda}$ whose diagonal elements are equal in blocks of $n_1, \ldots, n_K$ elements respectively. Following [6], rewrite the LHS of (40) as

$$\begin{bmatrix} \mathbf{I} & \mathbf{0} \\ \mathbf{0} & \bar{\mathbf{P}}^H \end{bmatrix} \begin{bmatrix} \mathbf{H}^H\bar{\mathbf{F}} \\ \sqrt{N_0}\mathbf{I} \end{bmatrix} \bar{\mathbf{P}} \ . \tag{41}$$

Consider the BD-GMD of the middle term

$$\begin{bmatrix} \mathbf{H}^H\bar{\mathbf{F}} \\ \sqrt{N_0}\mathbf{I} \end{bmatrix}^H = \mathbf{P}\mathbf{L}\mathbf{Q}^H \ , \tag{42}$$

where $\mathbf{P}$ is block diagonal, $\mathbf{L}$ is lower triangular, and both $\mathbf{P}$ and $\mathbf{Q}$ have orthonormal columns. Consequently, (40) becomes

$$\begin{bmatrix} \mathbf{I} & \mathbf{0} \\ \mathbf{0} & \bar{\mathbf{P}}^H \end{bmatrix} \mathbf{Q}\mathbf{L}^H\mathbf{P}^H\bar{\mathbf{P}} = \begin{bmatrix} \mathbf{Q}_u \\ \mathbf{Q}_d \end{bmatrix} \boldsymbol{\Lambda}\mathbf{B} \ . \tag{43}$$

Hence, we can choose $\bar{\mathbf{P}} = \mathbf{P}$, $\boldsymbol{\Lambda}\mathbf{B} = \mathbf{L}^H$ and $\mathbf{Q}_u$ to be the top N_T rows of $\mathbf{Q}$. This gives us our desired capacity-achieving block-equal-rate MMSE-DFE scheme.

The dual DPC scheme that is block-equal-rate and capacity-achieving can now be constructed. Figure 5 shows the block diagram of this scheme using THP. Using the duality technique in Section V-A, the pre-equalization, interference and nulling matrices are

$$\begin{aligned} \widetilde{\mathbf{F}} &= \mathbf{Q}_u\boldsymbol{\Lambda}^{-1}\mathbf{D}_c^{-1}\mathbf{D}_q \ , \\ \widetilde{\mathbf{B}} &= \mathbf{D}_c\mathbf{D}_q^{-1}\mathbf{L}\boldsymbol{\Lambda}^{-1}\mathbf{D}_c^{-1}\mathbf{D}_q \ , \\ \widetilde{\mathbf{W}} &= \bar{\mathbf{F}}\mathbf{P}\mathbf{D}_q^{-1}\mathbf{D}_c \ , \end{aligned} \tag{44}$$

where $\mathbf{D}_q$ is calculated from (34) and $\mathbf{D}_c$ contains the column norms of $\mathbf{Q}_u\boldsymbol{\Lambda}^{-1}$. To speed up the calculation of $\mathbf{D}_q$, (7) can be used instead of (32) to compute ρ_i. This dual DPC scheme is called the block diagonal UCD (BD-UCD), since the special single-user case of $K = 1$ is the UCD scheme in [6]. Its sum-capacity, as shown by (39), is precisely the broadcast channel sum-capacity.

C. Equal-Rate BD-UCD Scheme

In this section, a near-optimal DPC scheme for the broadcast channel that generates decoupled subchannels all with identical SINRs will be constructed. This Equal-Rate BD-UCD scheme maximizes the sum-rate given overall equal rates. This construction can be generalized to other rate constraints for the users. The crux lies in choosing the right uplink precoder $\bar{\mathbf{F}}$ so that the method in Section V-B produces the desired equal-rate scheme. The resulting scheme is called the *Equal-Rate* BD-UCD. The rest of this section will focus on finding this precoder. Let $\mathbf{H}^H = [\mathbf{H}_1^H, \mathbf{H}_2^H, \ldots, \mathbf{H}_K^H]$ be the uplink channel, where each $\mathbf{H}_i^H$ has n_i columns. Let $\bar{\mathbf{F}}_i$ be the i-th block of the block diagonal $\bar{\mathbf{F}}$. Then, the rate of user i, where user K is decoded first, is [8]

$$R_i = \log \frac{\det(\mathbf{I} + \frac{1}{N_0}\sum_{j\le i}\mathbf{H}_j^H \bar{\mathbf{F}}_j \bar{\mathbf{F}}_j^H \mathbf{H}_j)}{\det(\mathbf{I} + \frac{1}{N_0}\sum_{j< i}\mathbf{H}_j^H \bar{\mathbf{F}}_j \bar{\mathbf{F}}_j^H \mathbf{H}_j)} \ . \tag{45}$$

Ideally, for each i, $R_i = n_i \bar{R}$ for some $\bar{R}$, and $\mathrm{Tr}(\bar{\mathbf{F}}\bar{\mathbf{F}}^H) \le E_s$. The goal is to find $\bar{\mathbf{F}}$ such that $\bar{R}$ is maximized. While [18] has solved the symmetric capacity maximization, the scheme proposed here adds a further constraint that for any particular user, its subchannels have equal SINRs. The symmetric capacity maximizing solution would require time-sharing between schemes with different encoding orders. However, in order to minimize the complexity for our scheme, only a single decoding order is chosen, which may be suboptimal. In this section, the general equal-rate problem is solved with a fixed user ordering. A near-optimal algorithm of low complexity inspired by [13] and [12] is proposed below.

The basic building block of the algorithm is as follows: given a target rate $\bar{R}$, find a precoder $\bar{\mathbf{F}}$ that achieves the rate $\bar{R}$ for every subchannel with minimum power. Using a trick from [12], rewrite (45) as

$$n_i \bar{R} = \log\det(\mathbf{I} + \frac{1}{N_0}\mathbf{G}_i \bar{\mathbf{F}}_i \bar{\mathbf{F}}_i^H \mathbf{G}_i^H) \ , \tag{46}$$

where $\mathbf{G}_i = (\mathbf{I} + \frac{1}{N_0}\sum_{j<i}\mathbf{H}_j^H \bar{\mathbf{F}}_j \bar{\mathbf{F}}_j^H \mathbf{H}_j)^{-1/2}\mathbf{H}_i$ is the equivalent channel with the interference of all the other users. Thus, if $\mathbf{G}_i$ is given, then the minimum power $\bar{\mathbf{F}}_i$ satisfying (46) can be found by water-filling over it. Since $\mathbf{G}_i$ only depends on $\bar{\mathbf{F}}_j$ for $j < i$, equation (46) can be solved successively from $i = 1$ to $i = K$. Of course, to find the $\bar{\mathbf{F}}$ with minimum power satisfying the equations in (46) for all i, the $\bar{\mathbf{F}}_i$'s may need to be optimized jointly using iterative methods. However, to avoid incurring a high complexity cost, the above non-iterative algorithm will suffice for now. Let $\mathrm{P}(\bar{R})$ be the power $\mathrm{Tr}(\bar{\mathbf{F}}\bar{\mathbf{F}}^H)$ of the precoder $\bar{\mathbf{F}}$ computed by the above algorithm for a target rate $(\bar{R})$. The algorithm can now be completed by iteratively finding $\bar{R}$ such that $\mathrm{P}(\bar{R}) = E_s$. Since there is a near-linear relation between $\bar{R}$ and $\log \mathrm{P}(\bar{R})$, a simple numerical method like the secant method can be used. Simulations show that convergence is typically achieved in less than six iterations.

VI. Simulation Results

In this section, computer simulation results are presented to evaluate the performance of the four schemes proposed in this paper. In the simulations, the $12 \times \{4, 4, 4\}$ broadcast channel is considered. The elements of the channel matrix $\mathbf{H}$ are assumed to be independent and CSCG with zero mean and unit variance. The results are based on 3000 Monte Carlo realizations of $\mathbf{H}$. To compute the BER curves illustrated in Figures 6, 7 and 10, THP is applied for interference pre-subtraction at the transmitter. The transmit power is scaled down by a factor of $(M-1)/M$ for M-QAM to account for the THP precoding loss. All the schemes shown use 16-QAM unless otherwise indicated. Table I gives a summary of the schemes discussed in this paper, with italics to indicate the new proposed schemes.

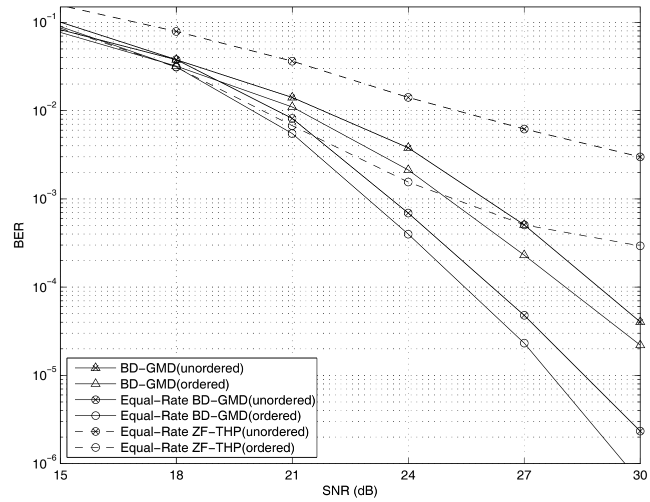

Fig. 6. BER performance comparison for ordered and unordered ZF-based schemes using THP and 16-QAM.

A. ZF-Based Schemes

The ZF schemes are the BD-GMD-based scheme and the Equal-Rate BD-GMD scheme. In the figures, the solid lines represent the new schemes while the dotted ones represent conventional schemes. Also, the squares and triangles indicate block-equal-rate schemes, while the circles indicate equal-rate ones. In Figures 7 and 8, all the schemes implement user ordering.

Figure 6 shows the gains in BER performance due to ordering users in the scheme. The unordered schemes are represented by shapes with crosses. Both the ordered BD-GMD and ordered Equal-Rate BD-GMD showed an improvement of about 1 dB at BER of 10^{-4} over their unordered counterparts respectively. The improvement is more appreciable for the Equal-Rate ZF-THP [15], with a gain of 6 dB even at BER of 10^{-3}. This gain in BER performance is due primarily to the improvement that user ordering has on the channel gain of the worst subchannel whose BER constitutes a major part of the average BER. Figure 7 demonstrates the effect of using different constellations on BER performance. 16-QAM is used for every user, unless "64,16,4-QAM" is stated, where the user with the largest channel gain is assigned 64-QAM, the next user 16-QAM, and the last user 4-QAM. (For simplicity, this assignment is fixed.) As expected, the BD-GMD(64,16,4-QAM) experiences a gain of 3 dB at BER of 10^{-5} over the BD-GMD, while the ZF-THP(64,16,4-QAM)

TABLE I
A COMPARISON BETWEEN VARIOUS SCHEMES

	ZF-based		**MMSE-based**	
	Block-equal-rate	Equal-rate	Block-equal-rate	Equal-rate
Single User	GMD		UCD	
Multiple Users				
- Single Antenna	ZF-DPC	Equal-Rate ZF-DPC	MMSE-DPC	-
- Multiple Antennas	*BD-GMD*	*Equal-Rate BD-GMD*	*BD-UCD*	*Equal-Rate BD-UCD*

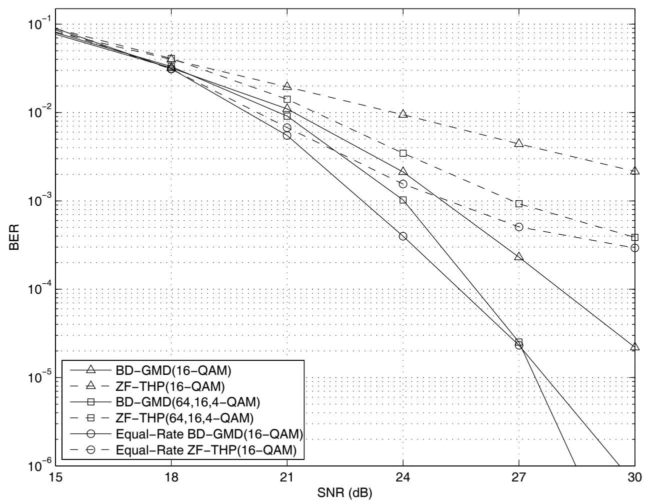

Fig. 7. Effect of receiver equalization on BER performance of ZF-based schemes using THP and user ordering.

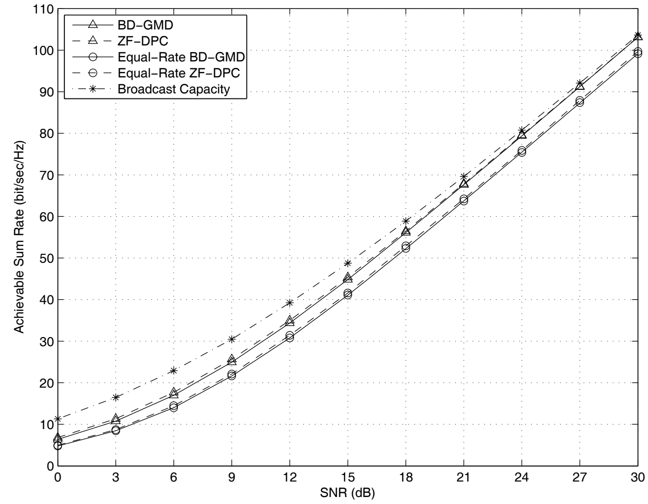

Fig. 8. Achievable sum-rate for ZF-based schemes with DPC and user ordering.

sees an improvement of more than 4 dB at BER of 10^{-3} over the ZF-THP. These gains are achieved without significant increase in receiver complexity. Also, by fitting the users with appropriate constellations to suit their channel gains, higher data rates can be achieved.

Figure 7 also highlights the significant improvement in BER performance of the new schemes over the conventional ones. The BD-GMD shows a gain of more than 6 dB at BER of 10^{-3} over the ZF-THP, while the Equal-Rate BD-GMD also gained more than 6 dB at BER of 10^{-4} over the Equal-Rate ZF-THP. At high SNRs, the steeper gradients of the BER curves for the new schemes is evidence of the diversity gain which resulted from linear equalization at the receivers. Figure 8 shows that the achievable sum-rates of the new schemes are negligibly less than that of the conventional schemes. While equalization at the mobile users' side help to increase the achievable sum-rates for the block diagonal schemes, there is more freedom for row ordering in the conventional schemes ($12! \approx 4.8 \times 10^8$ possible permutations) than for user ordering in the block diagonal schemes ($3! = 6$ possible permutations). The gain from ordering slightly overrides the advantage from receiver equalization. Finally, the performance of the equal-rate schemes is also compared with that of the block-equal-rate ones in Figure 7. The Equal-Rate BD-GMD shows a gain of 3 dB at BER of 10^{-5} over the BD-GMD. For the conventional Equal-Rate ZF-THP, the gain is even larger, with more than 6 dB improvement at BER of 10^{-3} over the ZF-THP. Note that for the equal-rate schemes, the BER performance of every user is the same as that shown in the figures, while for the block-equal-rate schemes, the BER curve is more indicative of the performance of the worst user. Thus, the gains described above represents the improvements experienced by the worst user when an equal-rate criterion is imposed. From Figure 8, the additional equal-rate criteria causes a slight loss of 1 dB in achievable sum-rate. This loss is due to the allocation of more power to the worst user in order to fulfill the equal-rate criterion.

B. MMSE-Based Schemes

The MMSE schemes are the BD-UCD and the Equal-Rate BD-UCD. In Figures 9 and 10, the solid lines represent the ZF-based schemes, while the dashed lines represent the MMSE-based ones. Again, the triangles indicate the block-equal-rate schemes and the circles the equal-rate ones.

In Figure 9, there is only a tiny improvement in achievable sum-rate of the BD-UCD over the MMSE-DPC. This can be understood by studying the effect of pre-equalization on the dual uplink channel capacity. The power loading aspect of pre-equalization has a much greater effect on capacity than the beam-forming aspect. Thus, in the dual broadcast case, although MMSE-DPC does not enjoy equalization at the receivers, it does not suffer any significant loss in capacity. However, in Figure 10, the BD-UCD shows a dramatic improvement in BER performance over the MMSE-THP, with more than 6 dB gain at BER of 10^{-3}. This is because of the diversity gain afforded by linear equalization at the receivers. Next, the advantage of using a MMSE-based scheme against its ZF-based counterpart is shown. The BD-UCD enjoys a slight 1.5 dB gain in achievable sum-rate over the BD-GMD at the low SNR region. As expected, their performance converges

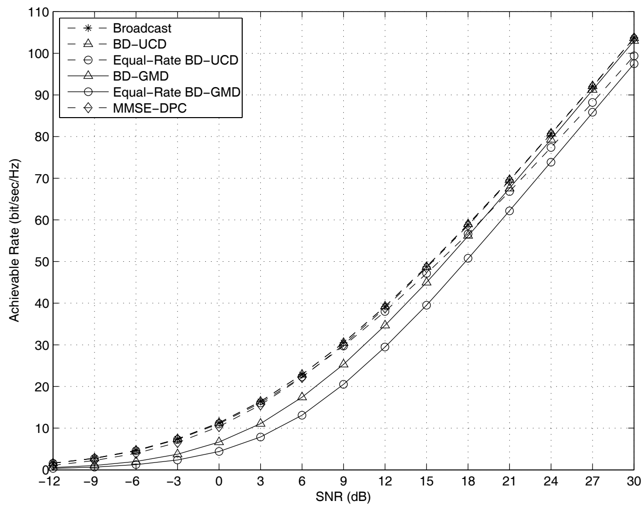

Fig. 9. Comparison of achievable sum-rate for ZF-based and MMSE-based schemes.

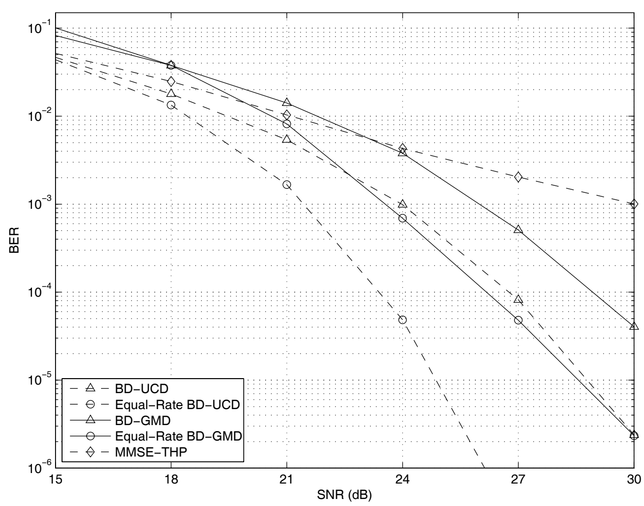

Fig. 10. BER performance comparison for ZF-based and MMSE-based schemes using THP and 16-QAM.

with increasing SNR. Similarly, the achievable sum-rates of the Equal-Rate BD-GMD and Equal-Rate BD-UCD converge at high SNR. In terms of BER performance, the effect of error minimization is seen more clearly. In Figure 10, the BD-UCD consistently shows a 2 dB gain in BER performance over the BD-GMD at all SNRs. Meanwhile, the Equal-Rate BD-UCD shows an improvement of as much as 4.5 dB over the Equal-Rate BD-GMD at BER of 10^{-6}. Finally, the feasibility of providing equal rates for all users is studied. In Figure 9, it is seen that the Equal-Rate BD-UCD achieves the same sum-rates as the BD-UCD at low SNR. Capacity in this SNR region is affected more by noise than by equal-rate constraints. At high SNR, the Equal-Rate BD-UCD only suffers a 1.5 dB loss. This also shows that the near-optimal iterative beamforming algorithm in Section V-C does not experience much performance loss. In Figure 10, the Equal-Rate BD-UCD demonstrates its superior BER performance over the BD-UCD, with about 4 dB gain at BER of 10^{-5}. This is because its worst subchannel is greatly elevated by the equal-rate constraint.

VII. Conclusion

We have proposed a novel block diagonal geometric mean decomposition (BD-GMD) for the MIMO broadcast channel. This decomposition creates spatial subchannels with *identical* SNRs/SINRs for each user. This allows the use of equal-rate coding which has benefits for the system implementation, especially in the design of modulation and coding schemes. Four applications based on the BD-GMD were proposed: BD-GMD, Equal-Rate BD-GMD, BD-UCD, and Equal-Rate BD-UCD, all of which apply DPC at the transmitter for interference pre-subtraction. The first two schemes are ZF-based schemes, while the later two are MMSE-based. Also, for each ZF and MMSE case, the first scheme is block-equal-rate while the second scheme is equal-rate. In the ZF schemes, user ordering is exploited to gain improvements in achievable sum-rates and user fairness. The optimal power allocation needed to provide equal rates for every user was also derived. As for the MMSE schemes, the BD-UCD achieves the broadcast channel capacity, while the Equal-Rate BD-UCD provides all users with equal rates without much loss of sum rate. Simulation results have shown that the new schemes exhibit excellent BER and sum rate properties.

Acknowledgment

The authors would like to thank the support provided by the Institute for Infocomm Research (I^{2}R) and the Agency for Science, Technology and Research (A*STAR) of Singapore. Furthermore, they are grateful to the anonymous reviewers for their invaluable comments that helped to improve the quality of this manuscript.

Appendix A
Proof of Theorem 1

Proof: Write $\mathbf{F} = \alpha\widetilde{\mathbf{F}}$. The condition $\mathrm{Tr}(\mathbf{F}^H\mathbf{F}) \le E_s$ can now be expressed as

$$\alpha^2 \le \frac{E_s}{\mathrm{Tr}(\widetilde{\mathbf{F}}^H\widetilde{\mathbf{F}})} \ , \tag{47}$$

so maximizing α is the same as minimizing $\mathrm{Tr}(\widetilde{\mathbf{F}}^H\widetilde{\mathbf{F}})$. Thus, (25) is equivalent to the problem

$$\begin{aligned} \texttt{minimize} \quad & \mathrm{Tr}(\widetilde{\mathbf{F}}^H\widetilde{\mathbf{F}}) \\ \texttt{subject to} \quad & \mathbf{J} \in \mathbb{L} \ , \ \mathbf{A} \in \mathbb{B} \ , \\ & \mathbf{A}\mathbf{H}\widetilde{\mathbf{F}} = \mathbf{J} \ , \\ & \|\mathbf{A}(i,:)\| = 1 \qquad \text{for } 1 \le i \le N_R \ . \end{aligned} \tag{48}$$

The Lagrangian $\mathscr{L}(\widetilde{\mathbf{F}}, \mathbf{A}, \mathbf{\Theta}, \mathbf{\Gamma})$ of this problem is

$$\mathrm{Tr}(\widetilde{\mathbf{F}}^H\widetilde{\mathbf{F}} - \mathrm{Re}(2\mathbf{\Theta}^H(\mathbf{A}\mathbf{H}\widetilde{\mathbf{F}} - \mathbf{I})) + \mathbf{\Gamma}(\mathbf{A}\mathbf{A}^H - \mathbf{I})) \ , \tag{49}$$

where $\mathbf{\Theta}$, $\mathbf{\Gamma}$ are Lagrange multipliers, $\mathbf{\Theta}$ is upper triangular, $\mathbf{\Gamma}$ is a real-valued diagonal matrix and $\mathrm{Re}(\mathbf{X})$ is the real-part of a complex matrix $\mathbf{X}$. If $\widetilde{\mathbf{F}}$ and $\mathbf{A}$ are optimal, then they satisfy

$$\nabla_{\widetilde{\mathbf{F}}}\mathscr{L} = 0 \implies \widetilde{\mathbf{F}} = (\mathbf{A}\mathbf{H})^H\mathbf{\Theta} \tag{50}$$

$$\nabla_{\mathbf{A}_i}\mathscr{L} = 0 \implies [\mathbf{\Theta}(\mathbf{H}\widetilde{\mathbf{F}})^H]_i = \mathbf{\Gamma}_i\mathbf{A}_i \quad \text{for } 1 \le i \le K \ , \tag{51}$$

where $\mathbf{A}_i$, $\mathbf{\Gamma}_i$, and $[\mathbf{\Theta}(\mathbf{H}\widetilde{\mathbf{F}})^H]_i$ are the i-th diagonal block of each matrix respectively. We begin by noting the following two important relations

$$\begin{aligned}\mathbf{J}^H\mathbf{\Theta} &= (\widetilde{\mathbf{F}}^H\mathbf{H}^H\mathbf{A}^H)\mathbf{\Theta}\\ &= \widetilde{\mathbf{F}}^H(\mathbf{H}^H\mathbf{A}^H\mathbf{\Theta})\\ &= \widetilde{\mathbf{F}}^H\widetilde{\mathbf{F}}\end{aligned} \tag{52}$$

$$\begin{aligned}\mathbf{\Gamma}_i\mathbf{A}_i\mathbf{A}_i^H &= [\mathbf{\Theta}\widetilde{\mathbf{F}}^H\mathbf{H}^H]_i\mathbf{A}_i^H\\ &= [\mathbf{\Theta}\widetilde{\mathbf{F}}^H\mathbf{H}^H\mathbf{A}^H]_i\\ &= [\mathbf{\Theta}\mathbf{J}^H]_i\\ &= \mathbf{\Theta}_i\mathbf{J}_i^H\end{aligned} \tag{53}$$

Lemma 1: There exists an optimal solution to (48) where $\mathbf{A}$ is unitary.

Proof: Pick any optimal solution to (48), and consider its matrix $\mathbf{A}$. From the third line in (48), $\det(\mathbf{J}) \neq 0$ implies that $\mathbf{A}$, $\mathbf{H}$ and $\widetilde{\mathbf{F}}$ have full rank. Similarly, by (50), $\det(\mathbf{\Theta}) \neq 0$. Then, by (53), $\det(\mathbf{\Gamma}_i) \neq 0$. Therefore, the diagonal elements of $\mathbf{\Gamma}_i$ are non-zero.

From (53), $\mathbf{A}_i\mathbf{A}_i^H = \mathbf{\Gamma}_i^{-1}\mathbf{\Theta}_i\mathbf{J}_i^H$ is upper triangular. On the other hand, $\mathbf{A}_i\mathbf{A}_i^H = [\mathbf{A}_i\mathbf{A}_i^H]^H$ is lower triangular. Thus, $\mathbf{A}_i\mathbf{A}_i^H$ must be diagonal. Since the rows of $\mathbf{A}_i$ are of unit norm, we have $\mathbf{A}_i\mathbf{A}_i^H = \mathbf{I}$ so $\mathbf{A}_i$ is unitary. Since $\mathbf{A}_i$ is unitary for all i, the lemma follows. ∎

Lemma 2: There exists an optimal solution to (48) where $\mathbf{A}$ is unitary and $\widetilde{\mathbf{F}} = \widetilde{\mathbf{Q}}\widetilde{\mathbf{\Omega}}$ where $\widetilde{\mathbf{Q}}$ has orthonormal columns and $\widetilde{\mathbf{\Omega}}$ is diagonal with non-negative real elements.

Proof: Using Lemma 1, pick any optimal solution to (48) in which $\mathbf{A}$ is unitary. From (52), we have $\widetilde{\mathbf{F}}^H\widetilde{\mathbf{F}} = \mathbf{J}^H\mathbf{\Theta}$ which is upper triangular. On the other hand, $\widetilde{\mathbf{F}}^H\widetilde{\mathbf{F}} = [\widetilde{\mathbf{F}}^H\widetilde{\mathbf{F}}]^H$ is lower triangular, so $\widetilde{\mathbf{F}}^H\widetilde{\mathbf{F}}$ must be diagonal. Furthermore, we can write $\widetilde{\mathbf{F}}^H\widetilde{\mathbf{F}} = \widetilde{\mathbf{\Omega}}^2$ where $\widetilde{\mathbf{\Omega}}$ is the diagonal matrix of the column norms of $\widetilde{\mathbf{F}}$, so $\widetilde{\mathbf{\Omega}}$ has non-negative real elements. Let $\widetilde{\mathbf{Q}}$ be the matrix of the unit column vectors of $\mathbf{F}$. Hence, $\widetilde{\mathbf{F}} = \widetilde{\mathbf{Q}}\widetilde{\mathbf{\Omega}}$. It is easy to check that $\widetilde{\mathbf{Q}}^H\widetilde{\mathbf{Q}} = \mathbf{I}$ from $\widetilde{\mathbf{F}}^H\widetilde{\mathbf{F}} = \widetilde{\mathbf{\Omega}}^2$, so $\widetilde{\mathbf{Q}}$ has orthonormal columns and this completes the lemma. ∎

As a result of these two lemmas, and the third line in (48),

$$\begin{aligned}&\mathbf{A}\mathbf{H}\widetilde{\mathbf{F}}\widetilde{\mathbf{\Omega}}^{-1} = \mathbf{A}\mathbf{H}\widetilde{\mathbf{Q}} = \mathbf{J}\widetilde{\mathbf{\Omega}}^{-1}\ , \ \text{and}\\ &\mathbf{H} = \mathbf{A}^H(\mathbf{J}\widetilde{\mathbf{\Omega}}^{-1})\widetilde{\mathbf{Q}}^H\ .\end{aligned} \tag{54}$$

Define $\widetilde{\mathbf{L}} = \mathbf{J}\widetilde{\mathbf{\Omega}}^{-1}$. Denote each diagonal block of $\widetilde{\mathbf{L}}$ corresponding to user i as $[\widetilde{\mathbf{L}}]_i$. It follows that $\det([\widetilde{\mathbf{L}}]_i) = \det([\widetilde{\mathbf{\Omega}}^{-1}]_i)$. Define $\widehat{\mathbf{H}}_i = [\mathbf{H}_1^T, \ldots, \mathbf{H}_i^T]^T$. Since $\mathbf{A}$ is block diagonal unitary and $\widetilde{\mathbf{Q}}$ has orthonormal columns, it can be seen that $\det([\widetilde{\mathbf{L}}]_i) = \sqrt{\frac{\det(\widehat{\mathbf{H}}_i\widehat{\mathbf{H}}_i^H)}{\det(\widehat{\mathbf{H}}_{i-1}\widehat{\mathbf{H}}_{i-1}^H)}}$. Thus, $\det([\widetilde{\mathbf{\Omega}}]_i)$ is a constant determined by the $\mathbf{H}$. As $\mathrm{Tr}(\widetilde{\mathbf{F}}^H\widetilde{\mathbf{F}}) = \mathrm{Tr}(\widetilde{\mathbf{\Omega}}^2)$, $\mathrm{Tr}(\widetilde{\mathbf{F}}^H\widetilde{\mathbf{F}})$ will be minimized when the diagonal elements of $[\widetilde{\mathbf{\Omega}}]_i$ are equal. Since $\widetilde{\mathbf{L}} = \mathbf{J}\widetilde{\mathbf{\Omega}}^{-1}$, the diagonal elements of $[\widetilde{\mathbf{L}}]_i$ are equal. Therefore, referring to (54), the BD-GMD ($\mathbf{H} = \mathbf{P}\mathbf{L}\mathbf{Q}^H$) provides the solution to the optimization (48), where

$$\widetilde{\mathbf{L}} = \mathbf{J}\widetilde{\mathbf{\Omega}}^{-1} = \mathbf{L}\ , \quad \widetilde{\mathbf{Q}} = \mathbf{Q}\ , \quad \text{and} \quad \mathbf{A} = \mathbf{P}^H\ . \tag{55}$$

Consequently, $\widetilde{\mathbf{\Omega}}^{-1} = \mathrm{diag}(\mathbf{L}) = \mathbf{\Lambda}$, and $\mathbf{J} = \mathbf{L}\mathbf{\Lambda}^{-1}$.

$$\mathbf{F} = \alpha\widetilde{\mathbf{F}} = \alpha\mathbf{Q}\mathbf{\Lambda}^{-1} \quad \text{and} \quad \alpha^2 = \frac{E_s}{\mathrm{Tr}(\widetilde{\mathbf{F}}\widetilde{\mathbf{F}}^H)} = \frac{E_s}{\mathrm{Tr}(\mathbf{\Lambda}^{-2})}\ . \tag{56}$$

This completes the proof for theorem 1. ∎

References

[1] I. E. Telatar, "Capacity of multi-antenna Gaussian channels," *European Trans. Telecomm.*, vol. 10, no. 6, pp. 585–596, Nov. 1999.

[2] G. J. Foschini and M. J. Gans, "On limits of wireless communications in a fading environment when using multiple antennas," *Wireless Personal Commun.*, vol. 6, no. 3, pp. 311–335, Mar. 1998.

[3] B. Hassibi, "A fast square-root implementation for BLAST," in *Proc. Asilomar Conf. Signals, Syst., Comput.*, pp. 1255–1259, Pacific Grove, Nov. 2000.

[4] A. Paulraj, R. Nabar, and D. Gore, *Introduction to Space-Time Wireless Communications*. Cambridge Univeristy Press, 2003.

[5] Y. Jiang, J. Li, and W. W. Hager, "Joint transceiver design for MIMO communications using geometric mean decomposition," *IEEE Trans. Signal Processing*, vol. 53, no. 10, pp. 3791–3803, Oct. 2005.

[6] Y. Jiang, J. Li, and W. Hager, "Uniform channel decomposition for MIMO communications," *IEEE Trans. Signal Processing*, vol. 53, no. 11, pp. 4283–4294, Nov. 2005.

[7] P. Viswanath and D. Tse, "Sum capacity of the vector Gaussian broadcast channel and uplink-downlink duality," *IEEE Trans. Inform. Theory*, vol. 49, no. 8, pp. 1912–1921, Aug. 2003.

[8] S. Vishwanath, N. Jindal, and A. Goldsmith, "Duality, achievable rates, and sum-rate capacity of Gaussian MIMO broadcast channels," *IEEE Trans. Inform. Theory*, vol. 49, no. 10, pp. 2658–2668, Oct. 2003.

[9] W. Yu and J. Cioffi, "Sum capacity of Gaussian vector broadcast channels," *IEEE Trans. Inform. Theory*, vol. 50, no. 9, pp. 1875–1892, Sept. 2004.

[10] G. Caire and S. Shamai (Shitz), "On the achievable throughput of a multiantenna Gaussian broadcast channel," *IEEE Trans. Inform. Theory*, vol. 49, no. 7, pp. 1691-1706, July 2003.

[11] H. Weingarten, Y. Steinberg, and S. Shamai, "The capacity region of the Gaussian MIMO broadcast channel," *Proc. IEEE Int. Symp. Information Theory*, pp. 174, Chicago, July 2004.

[12] N. Jindal, W. Rhee, S. Vishwanath, S. A. Jafar, and A. Goldsmith, "Sum power iterative water-filling for multi-antenna Gaussian broadcast channels," *IEEE Trans. Inform. Theory*, vol. 51, no. 4, pp. 1570–1580, Apr. 2005.

[13] M. Schubert and H. Boche, "Iterative multiuser uplink and downlink beamforming under SINR constraints," *IEEE Trans. Signal Processing*, vol. 53, no. 7, pp. 2324–2334, July 2005.

[14] R. F. H. Fischer, C. Windpassinger, A. Lampe, and J. B. Huber, "Space-time transmission using Tomlinson-Harashima precoding," in *Proc. ITG Conf. Source and Channel Coding*, pp. 139–147, Berlin, Jan. 2002.

[15] J. Liu and W. A. Krzymien, "A novel nonlinear precoding algorithm for the downlink of multiple antenna multi-user systems," in *Proc. IEEE VTC*, vol. 2, pp. 887–891, May 2005.

[16] M. K. Varanasi and T. Guess, "Optimum decision feedback multiuser equalization with successive decoding achieves the total capacity of the Gaussian multiple-access channel," in *Proc. Asilomar Conf. Signals, Syst., Comput.*, pp. 1405–1409, Monterey, Nov. 1997.

[17] P. W. Wolnainsky, G. J. Foschini, G. D. Golden, and R. A. Valenzuela, "V-BLAST: an architecture for achieving very high data rates over the rich-scattering wireless channel," in *Proc. ISSSE*, pp. 295–300, Pisa, Oct. 1998.

[18] J. Lee and N. Jindal, "Symmetric capacity of MIMO downlink channels," in *Proc. IEEE Int. Symp. Inform. Theory*, pp. 1031–1035, July 2006.

[19] N. Jindal and A. Goldsmith, "Dirty-paper coding versus TDMA for MIMO broadcast channels," *IEEE Trans. Inform. Theory*, vol. 51, no. 5, pp. 1783–1794, May 2005.

[20] Z. Pan, K. K. Wong, and T. S. Ng, "Generalized multiuser orthogonal space-division multiplexing," *IEEE Trans. Wireless Commun.*, vol. 3, no. 6, pp. 1969–1973, Nov. 2004.

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

[22] L. U. Choi and R. D. Murch, "A transmit preprocessing technique for multiuser MIMO systems using a decomposition approach," *IEEE Trans. Wireless Commun.*, vol. 3, no. 1, pp. 20–24, Jan. 2004.

[23] P. Viswanath, D. N. C. Tse, and R. Laroia, "Opportunistic beamforming using dumb antennas," *IEEE Trans. Inform. Theory*, vol. 48, no. 6, pp. 1277–1294, June 2002.

[24] Y.-C. Liang and R. Zhang, "Random beamforming for MIMO systems with multiuser diversity," in *Proc. Int. Symp. Personal, Indoor and Mobile Radio Commun.*, vol. 1, pp. 290–294, Sept. 2004.

[25] M. H. M. Costa, "Writing on dirty paper," *IEEE Trans. Inform. Theory*, vol. 29, no. 3, pp. 439–441, May 1983.

[26] G. Ginis and J. M. Cioffi, "A multi-user precoding scheme achieving crosstalk cancellation with application to DSL systems," in *Proc. Asilomar Conf. Signals, Systems, Computers* vol. 2, pp. 1627–1631, Oct. 2000.

[27] C. Windpassinger, R. F. H. Fischer, T. Vencel, and J. B. Huber, "Precoding in multiantenna and multiuser communications," *IEEE Trans. Wireless Commun.*, vol. 3, no. 4, pp. 1305–1316, July 2004.

Shaowei Lin received the B.Sc. degree (with honors) in mathematics from Stanford University, Stanford, CA in 2005. From 2005 to 2006, he was a Research Officer with the Institute for Infocomm Research, Singapore. Currently, he is pursuing his Ph.D. degree in mathematics at the University of California, Berkeley, CA. His research interests are in applied mathematics.

Winston W. L. Ho (S'06) received the B.Eng. (Hons.) degree in electrical engineering from the National University of Singapore (NUS) in 2004. He is currently pursuing a Ph.D. degree at NUS, under a scholarship from the Agency for Science, Technology and Research, Singapore. In 2003, during his half year industrial attachment, he worked at the Institute for Communications Research, now known as the Institute for Infocomm Research. His research interests include multiple antenna systems, cooperative communications, multiuser systems, and communication theory.

Ying-Chang Liang (SM'00) received PhD degree in Electrical Engineering in 1993. He is now Senior Scientist in the Institute for Infocomm Research (I2R), Singapore, where he has been leading the research activities in the area of cognitive radio and cooperative communications and the standardization activities in IEEE 802.22 wireless regional networks (WRAN) for which his team has made fundamental contributions in physical layer, MAC layer and spectrum sensing solutions. He also holds adjunct associate professorship positions in Nanyang Technological University (NTU) and National University of Singapore (NUS), both in Singapore, and adjunct professorship position with University of Electronic Science & Technology of China (UESTC). He has been teaching graduate courses in NUS since 2004. From Dec 2002 to Dec 2003, Dr Liang was a visiting scholar with the Department of Electrical Engineering, Stanford University. His research interest includes cognitive radio, dynamic spectrum access, reconfigurable signal processing for broadband communications, space-time wireless communications, wireless networking, information theory and statistical signal processing.

Dr Liang is now an Associate Editor of *IEEE Transactions on Vehicular Technology*. He was an Associate Editor of *IEEE Transactions on Wireless Communications* from 2002 to 2005, Lead Guest-Editor of *IEEE Journal on Selected Areas in Communications*, Special Issue on Cognitive Radio: Theory and Applications, and Guest-Editor of *Computer Networks Journal* (Elsevier) Special Issue on Cognitive Wireless Networks. He received the Best Paper Awards from IEEE VTC-Fall'1999 and IEEE PIMRC'2005, and 2007 Institute of Engineers Singapore (IES) Prestigious Engineering Achievement Award. Dr Liang has served for various IEEE conferences as technical program committee (TPC) member. He was Publication Chair of 2001 IEEE Workshop on Statistical Signal Processing, TPC Co-Chair of 2006 IEEE International Conference on Communication Systems (ICCS'2006), Panel Co-Chair of 2008 IEEE Vehicular Technology Conference Spring (VTC'2008-Spring), TPC Co-Chair of 3rd International Conference on Cognitive Radio Oriented Wireless Networks and Communications (CrownCom'2008), Deputy Chair of 2008 IEEE Symposium on New Frontiers in Dynamic Spectrum Access Networks (DySPAN'2008), and Co-Chair, Thematic Program on Random matrix theory and its applications in statistics and wireless communications, Institute for Mathematical Sciences, National University of Singapore, 2006. Dr Liang is a Senior Member of IEEE. He holds six granted patents and more than 15 filed patents.