Efficient Power Minimization for MIMO Broadcast Channels with BD-GMD

Winston W. L. Ho
Institute for Infocomm Research
21 Heng Mui Keng Terrace
Singapore 119613
Email: stuwlh@i2r.a-star.edu.sg

Ying-Chang Liang
Institute for Infocomm Research
21 Heng Mui Keng Terrace
Singapore 119613
Email: ycliang@i2r.a-star.edu.sg

***Abstract*—The problem of downlink power minimization given user rate requirements has been solved optimally in [12], [13]. However, due to the non-linear nature of the problem, convex optimization techniques have to be used, resulting in a high computational complexity. In this paper, the power minimization problem using dirty paper coding (DPC) is investigated. The SNR is distributed equally among the subchannels of each user to reduce each user's transceiver complexity, through the use of equal-rate modulation. The zero-forcing (ZF)-DPC problem is considered, therefore facilitating a closed form solution and resulting in a simple implementation. The optimum encoding order can be found with limited computations. To further reduce the complexity, a simple suboptimal method for finding the encoding order is given. This method is shown to have a sum power very close to the ZF-optimal power. The advantages of the methods proposed are their non-iterative nature and much reduced computational complexity.**

I. Introduction

Recently, the geometric mean decomposition (GMD) [1] has been proposed for point-to-point communications, assuming channel state information at the transmitter (CSIT). Combined with non-linear techniques such as decision-feedback equalization (DFE) or dirty paper coding (DPC), a multiple-input multiple-output (MIMO) channel is decomposed into multiple identical subchannels. Therefore the same constellation can be used on different subchannels, greatly reducing the transceiver complexity. For a MIMO broadcast channel (BC), with CSIT, GMD has been generalized to block-diagonal (BD)-GMD [6],[7]. When combined with DPC, a MIMO broadcast channel creates subchannels with identical SNRs for each user. The rates for different users may be different, or may even be constrained to be equal. Rate-maximizing transceiver designs based on the BD-GMD are described in [7].

In practical scenarios, users may be placed at different distances from the base station (BS), resulting in different variances for each independent user's channel matrix. Furthermore, users may have subscribed to plans of different data rates. Therefore, practical precoding schemes have to take that into consideration. In a cellular system, users experience interference from the BS of neighbouring cells. Consequently, an important question to answer is how to minimize the transmit power of each individual BS, while maintaining the rate requirements for the group of users currently served. This would help to reduce the interference that each BS produces to neighbouring cells, and as a result improve the whole cellular system's performance.

The power minimization problem has been solved in [10] for the case of users with a single antenna each. However, choosing the optimal ordering becomes complicated for more than three users. Convex optimization [8],[9] offers iterative methods to solve several non-linear communications problems. Using the uplink-downlink duality [2],[3],[4],[5], as well as convex optimization techniques, [12], [13] and [14] are key papers that address the power minimization problem, for the case of users with multiple antennas. [11] develops broadcast schemes to satisfy each user's minimum data rate and maximum BER requirements. This is done by considering virtual rate requirements that account for the SNR gap when using QAM and Tomlinson-Harashima precoding (THP).

For the methods mentioned above, while convex optimization may provide the optimal solution, the complexity is still very high, compared to a closed form solution. (The optimal method is referred to as interference-balancing (IB), as opposed to zero-forcing (ZF), since noise is taken into account, and interference is allowed between the subchannels. Generally, IB techniques are more difficult to implement than ZF ones, but have a better performance for the low SNR region.) Solutions have to be found, that are simple to implement in terms of computational complexity, and yet have a reasonable performance compared to the optimal solution. The difficulty with convex programming is that a substantial number of iterations have to be performed before the optimal solution is found. Each iteration itself may contain a large amount of computation that may not be visible from a simple complexity order expression. Additionally, the number of iterations required for handling each channel realization is random and not easily predictable.

Since iterative solutions suffer from the weaknesses mentioned earlier, the challenge is to find simple solutions that approach the optimal. Although suboptimal, these solutions help in reducing the complexity of the hardware. In this paper, efficient non-iterative precoding methods are designed to minimize the total transmit power for the MIMO BC, subject to individual rate constraints. Furthermore, the solution to generating subchannels with identical SNRs for each user is provided.

Firstly, for a fixed encoding order, the DPC problem is considered for the case when interference between all the subchannels is completely presubtracted. This ZF scenario is considered as is permits a closed form solution for the power minimization. Next, the optimal encoding order is derived. Although this may be suboptimal compared to the ordering found for the IB-optimal solution, the main advantage is that this ordering can be computed with a finite and predictable complexity, and has been shown to be computed much faster than that for the IB solution.

In order to decrease the complexity even further, three more simplified methods are proposed to find the user ordering that approach the performance of this ZF-optimal ordering. It is seen that by combining the three simplified methods, the power for the ZF-optimal solution can be reached very closely.

Notations:

Let $\mathbf{I}_N$ denote the $N \times N$ identity matrix. Let $\mathrm{diag}(\mathbf{L})$ denote the diagonal matrix with elements from the main diagonal of $\mathbf{L}$. Let $\mathbf{A} = \mathrm{blkd}(\mathbf{A}_1, \mathbf{A}_2, \ldots, \mathbf{A}_K)$ represent the block-diagonal matrix of the form

$$\mathbf{A} = \begin{bmatrix} \mathbf{A}_1 & 0 & \ldots & 0 \\ 0 & \mathbf{A}_2 & \ldots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \ldots & \mathbf{A}_K \end{bmatrix} . \tag{1}$$

II. Channel Model

Given a cellular-type system with one 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 $N_T \geq N_R$. The input-output relation can be represented as

$$\mathbf{y} = \mathbf{H}\mathbf{x} + \mathbf{u} , \tag{2}$$

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. Multiplexing is considered, where user i has n_i data-bearing subchannels. The SNR for every subchannel of user i is set equal to γ_i. $\mathbf{H} = [\mathbf{H}_1^T, \ldots, \mathbf{H}_K^T]^T$, where each $\mathbf{H}_i$ is the channel of user i. Assume that the noise vector $\mathbf{u}$ is zero-mean circularly symmetric complex Gaussian (CSCG) with $\mathbb{E}[\mathbf{u}\mathbf{u}^H] = N_0\mathbf{I}$, and $\mathbf{u}$ is independent of $\mathbf{x}$. Assume also that $\mathbb{E}[\|\mathbf{x}\|^2] = E_s$ and $\mathbf{H}$ is full rank. Denote this downlink model by $N_T \times [n_1, \ldots, n_K]$.

III. Power Minimization for a Fixed Arbitrary Ordering

In this section, a ZF-based block-equal-rate transceiver scheme that applies DPC at the transmitter and allocates power according to SNR requirements is presented. Since there is a simple relationship between the rate and the SNR for each subchannel,

$$R_i = \log_2(1 + \rho_i) , \tag{3}$$

rate requirements can easily be translated into SNR requirements. First assume that the encoding order of the users has been determined. This scheme minimizes the transmit power with the constraint of zero inter-user interference (IUI). Linear receive equalization is performed by a block diagonal matrix $\mathbf{A}$, where $\mathbf{A} = \mathrm{blkd}(\mathbf{A}_1, \mathbf{A}_2, \ldots, \mathbf{A}_K)$, and each block $\mathbf{A}_i$ is the receive equalization matrix of user i. It has been shown [6] that a BD-GMD can be done on a matrix $\mathbf{H}$ such that

$$\mathbf{H} = \mathbf{P}\mathbf{L}\mathbf{Q}^H , \tag{4}$$

where $\mathbf{P} = \mathrm{blkd}(\mathbf{P}_1, \mathbf{P}_2, \ldots, \mathbf{P}_K)$, each $\mathbf{P}_i$ is $n_i \times n_i$ unitary, $\mathbf{Q}^H\mathbf{Q} = \mathbf{I}_{N_R}$, and $\mathbf{L}$ is a square lower triangular matrix with elements equal in blocks of $n_1, \ldots, n_K$ elements (termed "block-equal-diagonal").

The problem of power minimization can be formulated as

$$\begin{array}{ll} \texttt{minimize} & \mathrm{Tr}(\mathbf{F}^H\mathbf{F}) \\ \texttt{subject to} & \mathbf{B} \in \mathbb{L} , \ \mathbf{A} \in \mathbb{B} \\ & \mathbf{A}\mathbf{H}\mathbf{F} = \sqrt{N_0}\mathbf{\Gamma}^{1/2}\mathbf{B} \\ & \|\mathbf{A}(i,:)\| = 1 \qquad \text{for } 1 \leq i \leq N_R . \end{array} \tag{5}$$

where $\mathbb{L}$ is the set of all lower triangular matrices with unit diagonal, $\mathbb{B}$ is the set of all block diagonal matrices of the form in (1), $\mathbf{F}$ is the precoder and $\mathbf{\Gamma}$ is the diagonal matrix of SNR requirements. $\mathbf{\Gamma} = \mathrm{blkd}(\mathbf{\Gamma}_1, \ldots, \mathbf{\Gamma}_K)$, where $\mathbf{\Gamma}_i = \gamma_i\mathbf{I}_{n_i}$.

Theorem 1: Let $\mathbf{H} = \mathbf{P}\mathbf{L}\mathbf{Q}^H$ be the BD-GMD of $\mathbf{H}$, and let $\mathbf{\Lambda} = \mathrm{diag}(\mathbf{L})$. $\mathbf{\Lambda} = \mathrm{blkd}(\mathbf{\Lambda}_1, \ldots, \mathbf{\Lambda}_K)$, where $\mathbf{\Lambda}_i = r_i\mathbf{I}_{n_i}$ for some r_i. Then, (5) is solved by

$$\begin{array}{ll} \mathbf{\Omega} = \sqrt{N_0}\mathbf{\Gamma}^{1/2}\mathbf{\Lambda}^{-1} , & \mathbf{F} = \mathbf{Q}\mathbf{\Omega} , \\ \mathbf{B} = \mathbf{\Omega}^{-1}\mathbf{\Lambda}^{-1}\mathbf{L}\mathbf{\Omega} , & \mathbf{A} = \mathbf{P}^H . \end{array} \tag{6}$$

Proof: See Appendix 1. ■

Here, $\mathbf{\Omega}$ is the diagonal power allocation matrix, $\mathbf{\Omega} = \mathrm{blkd}(\mathbf{\Omega}_1, \ldots, \mathbf{\Omega}_K)$, where $\mathbf{\Omega}_i = \omega_i\mathbf{I}_{n_i}$. The minimum power required is thus

$$E_s = \mathrm{Tr}(\mathbf{F}^H\mathbf{F}) = \mathrm{Tr}(\mathbf{\Omega}^2) . \tag{7}$$

IV. User Ordering

The encoding order of the users affects the total transmission power. Let $\{\pi_1, \pi_2, \ldots, \pi_K\}$ be the optimum encoding order, where the previous π_1-th user is now the first user, and so on. As the ordering of the users results in ordering the rows of $\mathbf{H}$, this can be represented by a multiplication by a permutation matrix, $\mathbf{D}$, such that $\mathbf{D}\mathbf{H} = \mathbf{P}\mathbf{L}\mathbf{Q}^H$. Here, the i-th block $\mathbf{P}_i$ of $\mathbf{P}$ has dimensions $n_{\pi_i} \times n_{\pi_i}$.

To find the optimum user ordering that minimizes the transmit power, an exhaustive search over all ordering permutations can be applied. An "exhaustive search" may seem like a large number, but the advantage of this method over other iterative methods is that the computations involved are much less. The best (ZF) ordering can be found over a hundred times faster than by using the optimum iterative method [12].

Yet, to reduce the complexity even further, three simple algorithms to find the near-optimal encoding order will be proposed. These can be done even before performing the BD-GMD or DPC. These methods are non-iterative, and do not involve convex optimization procedures. They proceed in a successive "top-down" manner, from user 1 to user K.

A. Method 1: Successive Closest Match

From (6), $\det(\mathbf{\Omega})$ is a constant determined by $\mathbf{\Gamma}$ and $\mathbf{H}$. Seeing from (7) that E_s is minimized when the diagonal values of $\mathbf{\Omega}$ are equal, $\mathbf{\Lambda}$ is designed such that it is close to a scalar multiple of $\mathbf{\Gamma}^{1/2}$. Let this desired matrix be $\mathbf{M}$, where $\mathbf{M} = \text{blkd}(\mathbf{M}_1, \ldots, \mathbf{M}_K)$, and $\mathbf{M}_i = m_i \mathbf{I}_{n_i}$. To ensure that the determinants of $\mathbf{H}\mathbf{H}^H$ and $\mathbf{M}^2$ match, define

$$\mathbf{M} = \mathbf{\Gamma}^{1/2} \cdot \sqrt[2N_R]{\frac{\det(\mathbf{H}\mathbf{H}^H)}{\det(\mathbf{\Gamma})}} \,. \tag{8}$$

Since $\mathbf{P}$ is block-diagonal unitary, $\mathbf{Q}$ is unitary and $\mathbf{L}$ is lower triangular, the diagonal elements in $\mathbf{\Lambda}_i$ is given by

$$r_i = \sqrt[2n_i]{\det(\mathbf{\Lambda}_i^2)} = \sqrt[2n_i]{\frac{\det(\widehat{\mathbf{H}}_i \widehat{\mathbf{H}}_i^H)}{\det(\widehat{\mathbf{H}}_{i-1} \widehat{\mathbf{H}}_{i-1}^H)}} \,, \tag{9}$$

where $\widehat{\mathbf{H}}_i = [\mathbf{H}_1^T, \ldots, \mathbf{H}_i^T]^T$. It is preferred that r_i be 'close' to m_i. From (9), it is seen that r_1 is independent of the ordering of the last $K-1$ users. Thus, $\mathbf{H}_1$ can first be chosen such that r_1 is close to m_1. Following this, $\mathbf{H}_2$ is chosen such that r_2 is close to m_2 and so on. A more precise method to determine 'closeness' will be given next. From (8), the sum power required is

$$E_s = \text{Tr}(\mathbf{\Omega}^2) = N_0 \sqrt[N_R]{\frac{\det(\mathbf{\Gamma})}{\det(\mathbf{H}\mathbf{H}^H)}} \text{Tr}\left(\frac{\mathbf{M}^2}{\mathbf{\Lambda}^2}\right) \,, \tag{10}$$

where a matrix can be placed in the denominator for convenience because it is diagonal. Minimizing E_s is the same as minimizing

$$\text{Tr}\left(\frac{\mathbf{M}^2}{\mathbf{\Lambda}^2}\right) = \text{Tr}\left(\frac{\mathbf{M}_1^2}{\mathbf{\Lambda}_1^2}\right) + \text{Tr}\left(\frac{\check{\mathbf{M}}_2^2}{\check{\mathbf{\Lambda}}_2^2}\right), \tag{11}$$

where $\check{\mathbf{M}}_i = \text{blkd}(\mathbf{M}_i, \ldots, \mathbf{M}_K)$ and $\check{\mathbf{\Lambda}}_i = \text{blkd}(\mathbf{\Lambda}_i, \ldots, \mathbf{\Lambda}_K)$. We have

$$\det\left(\frac{\mathbf{M}^2}{\mathbf{\Lambda}^2}\right) = 1 = \det\left(\frac{\mathbf{M}_1^2}{\mathbf{\Lambda}_1^2}\right) \cdot \det\left(\frac{\check{\mathbf{M}}_2^2}{\check{\mathbf{\Lambda}}_2^2}\right), \tag{12}$$

In general, for this "top-down" approach, the effect of choosing a particular $\mathbf{\Lambda}_1$ on the following $\mathbf{\Lambda}_i$'s is not known. Let the best-case $\check{\mathbf{\Lambda}}_2$ to minimize (11) given $\mathbf{\Lambda}_1$ be $\widetilde{\mathbf{\Lambda}}_2$.

$$\frac{\check{\mathbf{M}}_2^2}{\widetilde{\mathbf{\Lambda}}_2^2} = \mathbf{I}_{\check{n}_2} \cdot \sqrt[\check{n}_2]{\det\left(\frac{\mathbf{\Lambda}_1^2}{\mathbf{M}_1^2}\right)} \tag{13}$$

where $\check{n}_i = \sum_{j=i}^K n_j$. Therefore (11) is equivalent to

$$n_1 \sqrt[n_1]{\det\left(\frac{\mathbf{M}_1^2}{\mathbf{\Lambda}_1^2}\right)} + \check{n}_2 \sqrt[\check{n}_2]{\det\left(\frac{\mathbf{\Lambda}_1^2}{\mathbf{M}_1^2}\right)}. \tag{14}$$

Since $\det(\mathbf{\Lambda}_1^2)$ can be found from $\mathbf{H}_1$ using (9), $\mathbf{H}_1$ is chosen to minimize (14). Next, the selection of users 2 to K will be described.

Define $\hat{\mathbf{M}}_i = \text{blkd}(\mathbf{M}_1, \ldots, \mathbf{M}_i)$ and $\hat{\mathbf{\Lambda}}_i = \text{blkd}(\mathbf{\Lambda}_1, \ldots, \mathbf{\Lambda}_i)$. For the i-th user, since $\hat{\mathbf{\Lambda}}_{i-1}$ has been determined, minimizing E_s is equivalent to minimizing

$$\text{Tr}\left(\frac{\mathbf{M}_i^2}{\mathbf{\Lambda}_i^2}\right) + \text{Tr}\left(\frac{\check{\mathbf{M}}_{i+1}^2}{\check{\mathbf{\Lambda}}_{i+1}^2}\right). \tag{15}$$

Again, we have

$$\begin{aligned} 1 &= \det\left(\frac{\mathbf{M}^2}{\mathbf{\Lambda}^2}\right) \\ &= \det\left(\frac{\hat{\mathbf{M}}_{i-1}^2}{\hat{\mathbf{\Lambda}}_{i-1}^2}\right) \cdot \det\left(\frac{\mathbf{M}_i^2}{\mathbf{\Lambda}_i^2}\right) \cdot \det\left(\frac{\check{\mathbf{M}}_{i+1}^2}{\check{\mathbf{\Lambda}}_{i+1}^2}\right), \end{aligned} \tag{16}$$

so the best-case $\check{\mathbf{\Lambda}}_{i+1}$ to minimize (15) is $\widetilde{\mathbf{\Lambda}}_{i+1}$, where

$$\frac{\check{\mathbf{M}}_{i+1}^2}{\widetilde{\mathbf{\Lambda}}_{i+1}^2} = \mathbf{I}_{\check{n}_{i+1}} \cdot \sqrt[\check{n}_{i+1}]{\det\left(\frac{\hat{\mathbf{\Lambda}}_{i-1}^2}{\hat{\mathbf{M}}_{i-1}^2}\right) \det\left(\frac{\mathbf{\Lambda}_i^2}{\mathbf{M}_i^2}\right)} \tag{17}$$

Therefore (15) is equivalent to

$$n_i \sqrt[n_i]{\det\left(\frac{\mathbf{M}_i^2}{\mathbf{\Lambda}_i^2}\right)} + \check{n}_{i+1} \sqrt[\check{n}_{i+1}]{\det\left(\frac{\hat{\mathbf{\Lambda}}_{i-1}^2}{\hat{\mathbf{M}}_{i-1}^2}\right) \det\left(\frac{\mathbf{\Lambda}_i^2}{\mathbf{M}_i^2}\right)} \tag{18}$$

Also, $\det(\mathbf{\Lambda}_i^2)$ can be calculated from $\mathbf{H}_i$ using (9), where the value $\det(\widehat{\mathbf{H}}_{i-1}\widehat{\mathbf{H}}_{i-1}^H)$ has already been found from the earlier step. $\mathbf{H}_i$ is chosen to minimize (18), and so on until user K, where there is only 1 choice. Thus, let this method be called *successive closest match* (SCM).

B. Method 2: Minimize r_i

When users have equal channel strengths, the unordered BD-GMD, which is basically a QR decomposition, $\mathbf{P}^H\mathbf{H} = \mathbf{L}\mathbf{Q}^H$, usually has the first diagonal element of $\mathbf{L}$ much larger than the last element. If equal SNRs are desired for each user, which is usually the case, minimizing the first diagonal element tends to decrease the spread in the diagonal values of $\mathbf{L}$.

Therefore, this method can be stated simply. Starting from user 1, using (9), $\mathbf{H}_i$ is chosen to minimize r_i, and so on for users 2 to K.

C. Method 3: Minimize Channel Strength

Consider the case where users are at different distances from the base station, resulting in different channel strengths. Suppose equal SNRs are desired for each user. In the dual uplink channel, it is expected that user with the weakest channel should be decoded last, in order to improve his achievable rate. In the downlink, this corresponds to encoding the user with the weakest channel first.

Thus, again starting from user 1, $\mathbf{H}_i$ is chosen to minimize $\text{Tr}(\mathbf{H}_i\mathbf{H}_i^H)/n_i$, and so on until user K.

D. 'Best Choice' Method

Simulations show that for different settings of user channel strengths, user antenna numbers and user SNR requirements, different methods are best for minimizing the total transmit power. Usually, method 1 (SCM) gives the best performance. Due to the reasons mentioned in sections IV-B and IV-C, methods 2 or 3 may perform the best. In fact, there is a slight possibility that a particular original ordering is already optimal.

Therefore, it makes sense to select the best of methods 1 to 3 as well as the original ordering.

V. Computational Complexity

To find the optimum user ordering that minimizes the total transmit power, an exhaustive search across all the user permutations can be done. For K users, there are $K!$ permutations. For each permutation, K determinants has to be calculated based on (9), before the transmit power can be evaluated using (10), resulting in a total of $KK!$ determinant calculations. Since the value of r_i is independent of the ordering of the first $i-1$ users, the number of determinants to be calculated can be reduced to

$$N_d = K! \sum_{i=0}^{K-1} \frac{1}{i!} \ . \tag{19}$$

On the other hand, the number of determinants to calculate for the proposed SCM method is

$$N_s = \sum_{i=1}^{K} i = \frac{K(K+1)}{2} \ . \tag{20}$$

Note that the calculation of the determinants of $\mathbf{M}_i^2$, $1 \leq i \leq K$, have been omitted as they are diagonal matrices. Also, in (18), $\det(\hat{\mathbf{\Lambda}}_{i-1}^2)$ can be found from

$$\det(\hat{\mathbf{\Lambda}}_{i-1}^2) = \det(\hat{\mathbf{\Lambda}}_{i-2}^2)\det(\mathbf{\Lambda}_{i-1}^2) \tag{21}$$

where the two terms on the right have already been calculated in the previous step.

Method 2 (min r_i) also requires N_s number of determinant calculations. Method 3 (min chan) is the simplest, without requiring any determinant calculations.

The 'best choice' method is interesting. Since it is a composition of methods 1 to 3, the number of determinants to be computed is $2N_s$. An additional $4K$ determinants have to be calculated to find r_i using (9). Following that, (10) can be evaluated to find the minimum power of all these 4 orderings. Finally the BD-GMD is applied to the best ordering. The complexity of the BD-GMD is only K times as high as the GMD [1].

VI. Simulation Results

Consider the $N_T \times [n_1, \ldots, n_K]$ downlink scenario. Let $\mathbf{n} = [n_1, \ldots, n_K]$ be the antenna numbers of the users. Let $\boldsymbol{\gamma} = [\gamma_1, \ldots, \gamma_K]$ be the vector of SNR requirements for each user. Let $\mathbf{c} = [c_1, \ldots, c_K]$ be the channel strengths of each user. The elements of the channel matrix of user i are modelled as i.i.d. zero-mean CSCG with variance c_i.

Figures 1 to 6 show the simulation results. 300 Monte Carlo trials are performed for each value of γ. In general, method 1 (SCM) performs the best, followed by method 2 (min r_i), then method 3 (min chan), although settings have been found in which method 2 or method 3 performs best. The figures are chosen to give the most general settings encountered.

For figures 1 to 4, the channel strengths for each user are set equal, i.e. $\mathbf{c} = [1,1,1,1,1,1]$. Fig. 1 represents the case of 2 antennas per user, and equal SNR requirements. Fig. 2 shows the case where different users have different SNR requirements, for example if they have subscribed to plans of different data rates. Fig. 3 is more general, when users can have 4, 2, or 1 antennas. In Fig. 4, the user with more antennas is assigned a lower SNR requirement. This is reasonable because more data streams are permitted for this user, so each stream is allowed to have a lower data rate, if the rates for the different users are comparable. Fig. 5 presents the case similar to Fig. 1, but this time with varying channel strengths. This represents the practical scenario where users are positioned at varying distances from the BS. Fig. 6 is a generalization where users have different antenna numbers and different SNR requirements.

VII. Conclusion

The optimal solution to the broadcast power minimization problem given user SINR requirements has been solved optimally using iterative methods and convex optimization. However, these methods are computationally expensive, as mentioned in the introduction. A major hurdle for MIMO systems is the high complexity involved.

In this paper, the problem of the ZF power minimization using DPC is formulated and solved in a closed form expression, using the BD-GMD. The optimal ordering can be found much faster than for the optimal IB method. To speed up the process of obtaining the best ordering, sub-optimal methods have been proposed. The methods have been shown to reach the ZF-optimal power very closely.

Acknowledgment

For this work, the authors are grateful for the funding provided by the Institute for Infocomm Research (I^2R) and the Agency for Science, Technology and Research (A*STAR) of Singapore. Also, the authors would like to thank Shaowei Lin for assistance in deriving the proof in appendix A, on the optimality of the BD-GMD.

Appendix A
Proof of Theorem 1

Proof: The Lagrangian $\mathscr{L}(\mathbf{F}, \mathbf{A}, \boldsymbol{\alpha}, \widetilde{\boldsymbol{\rho}}, \boldsymbol{\mu})$ for problem (5) is

$$\mathrm{Tr}(\mathbf{F}^H\mathbf{F} + \mathrm{Re}(\widetilde{\boldsymbol{\rho}}^H(\mathbf{AHF} - \sqrt{N_0}\mathbf{\Gamma}^{1/2})) + \boldsymbol{\mu}(\mathbf{AA}^H - \mathbf{I})) \ , \tag{22}$$

where $\widetilde{\boldsymbol{\rho}}$, $\boldsymbol{\mu}$ are Lagrange multipliers, $\widetilde{\boldsymbol{\rho}}$ an upper triangular complex matrix, $\boldsymbol{\mu}$ a real-valued diagonal matrix, and $\mathrm{Re}(\mathbf{X})$

the real-part of a complex matrix $\mathbf{X}$. If $\mathbf{F}$ and $\mathbf{A}$ are optimal, then they satisfy

$$\nabla_{\mathbf{F}}\mathscr{L} = 2\mathbf{F} + (\mathbf{A}\mathbf{H})^H \widetilde{\boldsymbol{\rho}} = \mathbf{0} \tag{23}$$

$$\nabla_{\mathbf{A}_i}\mathscr{L} = [\widetilde{\boldsymbol{\rho}}(\mathbf{H}\mathbf{F})^H]_i + 2\boldsymbol{\mu}_i \mathbf{A}_i = \mathbf{0} \quad \text{for } 1 \le i \le K \tag{24}$$

where $\mathbf{A}_i$, $\boldsymbol{\mu}_i$ and $[(\widetilde{\boldsymbol{\rho}}\mathbf{H}\mathbf{F})^H]_i$ are the i-th diagonal block of each matrix respectively. Begin by letting $\boldsymbol{\rho} = -\frac{1}{2}\widetilde{\boldsymbol{\rho}}$, also upper triangular. From (23),

$$\mathbf{F} = \mathbf{H}^H \mathbf{A}^H \boldsymbol{\rho} \ . \tag{25}$$

Define $\bar{\mathbf{J}} = \sqrt{N_0}\boldsymbol{\Gamma}^{1/2}\mathbf{B}$, a lower triangular matrix. From (24),

$$\begin{aligned} \boldsymbol{\mu}_i \boldsymbol{A}_i \boldsymbol{A}_i^H &= \left[-\frac{1}{2}\widetilde{\boldsymbol{\rho}}\mathbf{F}^H\mathbf{H}^H\right]_i \boldsymbol{A}_i^H \\ &= [\boldsymbol{\rho}\mathbf{F}^H\mathbf{H}^H]_i \boldsymbol{A}_i^H \\ &= [\boldsymbol{\rho}\mathbf{F}^H\mathbf{H}^H \boldsymbol{A}^H]_i \\ &= [\boldsymbol{\rho}\bar{\mathbf{J}}^H]_i \ . \end{aligned} \tag{26}$$

Since $\boldsymbol{\mu}_i$ is diagonal, $\boldsymbol{A}_i\boldsymbol{A}_i^H$ is upper triangular. As $\boldsymbol{A}_i\boldsymbol{A}_i^H$ is also hermitian, it has to be diagonal. Together with the constraint of unit row norm of $\mathbf{A}$, it follows that $\mathbf{A}$ is unitary. Likewise,

$$\begin{aligned} (\bar{\mathbf{J}}^H)\boldsymbol{\rho} &= (\mathbf{F}^H\mathbf{H}^H\mathbf{A}^H)\boldsymbol{\rho} \\ &= \mathbf{F}^H\mathbf{F} \ . \end{aligned} \tag{27}$$

Since $\mathbf{F}^H\mathbf{F}$ is upper triangular and hermitian, it is diagonal.

$$\mathbf{F}^H\mathbf{F} = \mathrm{diag}(\boldsymbol{\rho})\sqrt{N_0}\boldsymbol{\Gamma}^{1/2} \ . \tag{28}$$

As the diagonal elements of $\mathbf{F}^H\mathbf{F}$ are positive real, the diagonal elements of $\boldsymbol{\rho}$ are also positive real. Define

$$\bar{\boldsymbol{\Lambda}} = (\mathbf{F}^H\mathbf{F})^{-1/2} \ , \tag{29}$$

where $\bar{\boldsymbol{\Lambda}}$ is a diagonal matrix of positive real entries. Therefore

$$(\mathbf{F}\bar{\boldsymbol{\Lambda}})^H(\mathbf{F}\bar{\boldsymbol{\Lambda}}) = \mathbf{I} \ , \tag{30}$$

Let the unitary matrix $\mathbf{F}\bar{\boldsymbol{\Lambda}}$ be denoted by $\bar{\mathbf{Q}}$. Then by (5),

$$\mathbf{A}\mathbf{H}\mathbf{F}\bar{\boldsymbol{\Lambda}} = \mathbf{A}\mathbf{H}\bar{\mathbf{Q}} = \bar{\mathbf{J}}\bar{\boldsymbol{\Lambda}} \tag{31}$$

$$\mathbf{H} = \mathbf{A}^H(\bar{\mathbf{J}}\bar{\boldsymbol{\Lambda}})\bar{\mathbf{Q}}^H \ , \tag{32}$$

where $\bar{\mathbf{J}}\bar{\boldsymbol{\Lambda}}$ is lower triangular. Let $\bar{\mathbf{L}} = \bar{\mathbf{J}}\bar{\boldsymbol{\Lambda}}$. So $\mathrm{diag}(\bar{\mathbf{L}}) = \sqrt{N_0}\boldsymbol{\Gamma}^{1/2}\bar{\boldsymbol{\Lambda}}$. Denote each diagonal block of $\bar{\mathbf{L}}$ corresponding to user i as $[\bar{\mathbf{L}}]_i$. It follows that

$$\begin{aligned} \det([\bar{\mathbf{L}}]_i) &= \det([\bar{\mathbf{J}}]_i)\det([\bar{\boldsymbol{\Lambda}}]_i) \\ &= (\sqrt{N_0\gamma_i})^{n_i}\det([\bar{\boldsymbol{\Lambda}}]_i) \ . \end{aligned} \tag{33}$$

Define $\widehat{\mathbf{H}}_i = [\mathbf{H}_1^T, \ldots, \mathbf{H}_i^T]^T$. Since $\mathbf{A}$ is block diagonal unitary and $\bar{\mathbf{Q}}$ is unitary, it can be seen that

$$\det([\bar{\mathbf{L}}]_i) = \sqrt{\frac{\det(\widehat{\mathbf{H}}_i\widehat{\mathbf{H}}_i^H)}{\det(\widehat{\mathbf{H}}_{i-1}\widehat{\mathbf{H}}_{i-1}^H)}} \ . \tag{34}$$

Thus $\det([\bar{\boldsymbol{\Lambda}}]_i)$ is a constant determined by the $\mathbf{H}$, γ_i and n_i. Recall from (7) and (30) that the power needed is

$$E_s = \mathrm{Tr}(\mathbf{F}^H\mathbf{F}) = \mathrm{Tr}(\bar{\boldsymbol{\Lambda}}^{-2}) \ . \tag{35}$$

Therefore, E_s will be minimized when the diagonal elements of $[\bar{\boldsymbol{\Lambda}}]_i$ are equal. Since the diagonal elements of $[\bar{\mathbf{J}}]_i$ are equal, the same is true for the diagonal values of $[\bar{\mathbf{L}}]_i$. Therefore, from (32), and the BD-GMD decomposition, $\mathbf{H} = \mathbf{P}\mathbf{L}\mathbf{Q}^H$,

$$\bar{\mathbf{L}} = \bar{\mathbf{J}}\bar{\boldsymbol{\Lambda}} = \mathbf{L} \ , \quad \bar{\mathbf{Q}} = \mathbf{Q} \ , \quad \mathbf{A} = \mathbf{P}^H \ , \tag{36}$$

where

$$\bar{\boldsymbol{\Lambda}} = \mathrm{diag}(\bar{\mathbf{J}})^{-1}\mathrm{diag}(\mathbf{L}) = (\sqrt{N_0}\boldsymbol{\Gamma}^{1/2})^{-1}\boldsymbol{\Lambda} \ . \tag{37}$$

Define

$$\boldsymbol{\Omega} = \bar{\boldsymbol{\Lambda}}^{-1} = \sqrt{N_0}\boldsymbol{\Gamma}^{1/2}\boldsymbol{\Lambda}^{-1} \ . \tag{38}$$

Finally,

$$\begin{aligned} \mathbf{F} &= \mathbf{Q}\boldsymbol{\Omega} \ , \\ \mathbf{B} &= (\sqrt{N_0}\boldsymbol{\Gamma}^{1/2})^{-1}\mathbf{L}\bar{\boldsymbol{\Lambda}}^{-1} = \boldsymbol{\Omega}^{-1}\boldsymbol{\Lambda}^{-1}\mathbf{L}\boldsymbol{\Omega} \end{aligned} \tag{39}$$

completes the solution to (5). ■

REFERENCES

[1] 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.

[2] 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.

[3] 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.

[4] W. Yu and J. Cioffi, "Sum Capacity of Gaussian Vector Broadcast Channels," *IEEE Trans. Inform. Theory*, vol. 50, no. 9, pp. 1875–1892, Sep. 2004.

[5] 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, Jul. 2003.

[6] S. Lin, W. W. L. Ho, and Y.-C. Liang, "Block-diagonal Geometric Mean Decomposition (BD-GMD) for Multiuser MIMO Broadcast Channels," *Int. Symp. Personal, Indoor and Mobile Radio Commun.*, Helsinki, 11–14 Sep. 2006.

[7] S. Lin, W. W. L. Ho, and Y.-C. Liang, "MIMO Broadcast Communications with Block-Diagonal Geometric Mean Decomposition," —

[8] S. Boyd and L. Vandenberghe, *Convex Optimization.* Cambridge, U.K.: Cambridge Univ. Press, 2004.

[9] Z. Q. Luo and W. Yu, "An Introduction to Convex Optimization for Communications and Signal Processing," *IEEE J. Select. Areas Commun.*, vol. 24, pp. 1426–1438, Aug. 2006.

[10] 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, Jul. 2005.

[11] C.-H. F. Fung, W. Yu, and T. J. Lim, "Precoding for the Multiantenna Downlink: Multiuser SNR Gap and Optimal User Ordering," *IEEE Trans. Commun.*, vol. 55, pp. 188–197, Jan. 2007.

[12] J. Lee and N. Jindal "Symmetric Capacity of MIMO Downlink Channels," *IEEE Int. Symp. Inform. Theory*, pp. 1031–1035, Jul. 2006.

[13] M. Mohseni, R. Zhang, and J. M. Cioffi "Optimized Transmission for Fading Multiple-Access and Broadcast Channels With Multiple Antennas," *IEEE J. Select. Areas Commun.*, vol. 24, pp. 1627–1639, Aug. 2006.

[14] C.-H. F. Fung, W. Yu, and T. J. Lim, "Multi-antenna downlink precoding with individual rate constraints: power minimization and user ordering," *Proc. Int. Conf. Commun. Systems*, pp. 45–49, Sep. 2004.

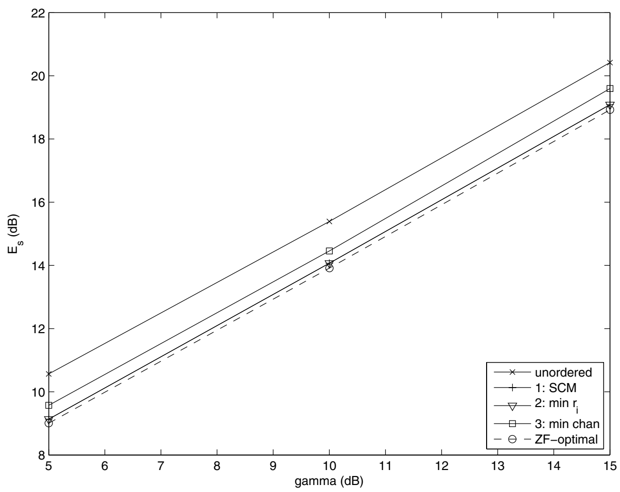

Fig. 1. $12 \times [2, 2, 2, 2, 2, 2]$, $\boldsymbol{\gamma} = [\gamma, \gamma, \gamma, \gamma, \gamma, \gamma]$.

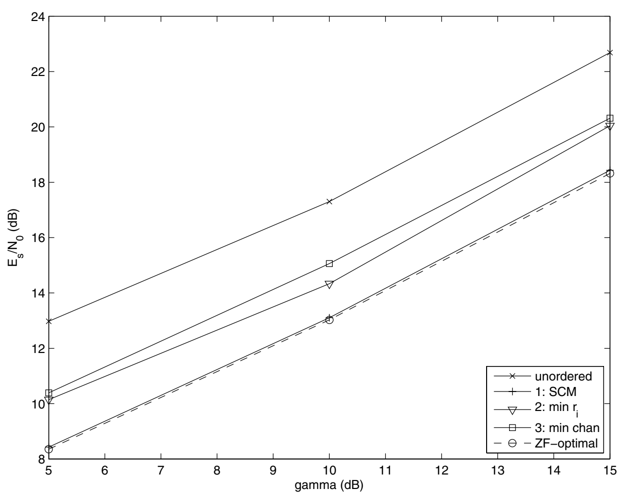

Fig. 2. $12 \times [2, 2, 2, 2, 2, 2]$, $\boldsymbol{\gamma} = [\gamma/2, \gamma/2, \gamma, \gamma, 2\gamma, 2\gamma]$.

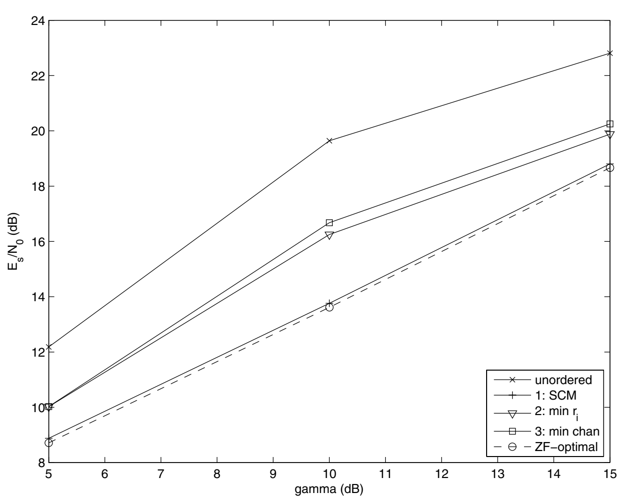

Fig. 3. $12 \times [4, 2, 2, 2, 1, 1]$, $\boldsymbol{\gamma} = [\gamma, \gamma, \gamma, \gamma, \gamma, \gamma]$.

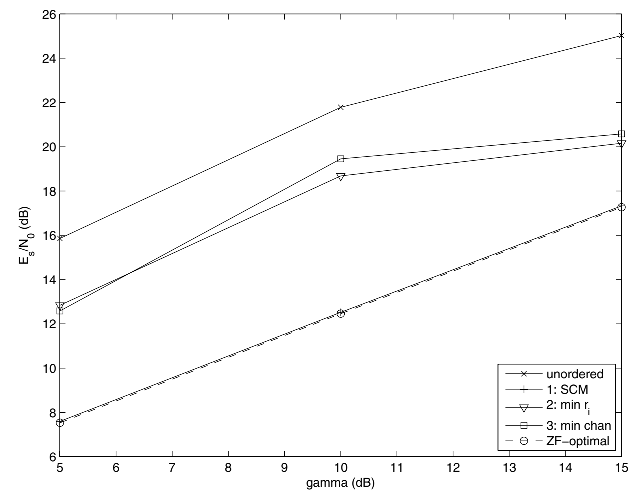

Fig. 4. $12 \times [4, 2, 2, 2, 1, 1]$, $\boldsymbol{\gamma} = [\gamma/2, \gamma, \gamma, \gamma, 2\gamma, 2\gamma]$.

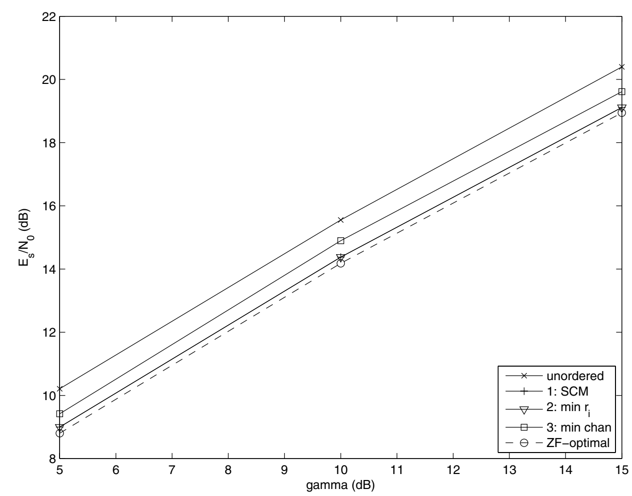

Fig. 5. $\mathbf{c} = [1.5, 1.5, 1, 1, 0.5, 0.5]$, $12 \times [2, 2, 2, 2, 2, 2]$, $\boldsymbol{\gamma} = [\gamma, \gamma, \gamma, \gamma, \gamma, \gamma]$.

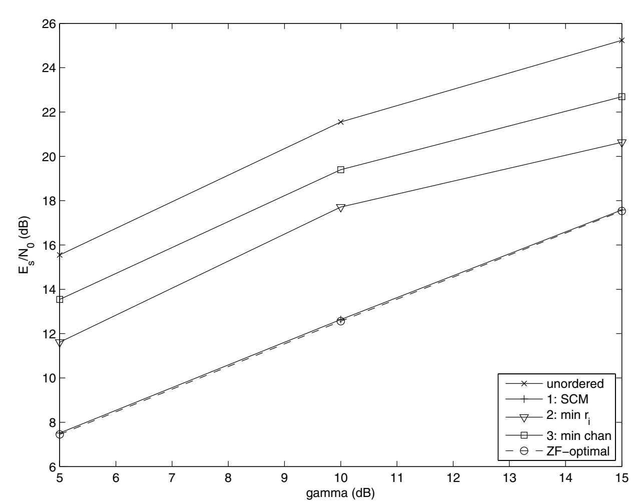

Fig. 6. $\mathbf{c} = [1, 1.5, 1, 0.5, 1.5, 0.5]$, $12 \times [4, 2, 2, 2, 1, 1]$, $\boldsymbol{\gamma} = [\gamma/2, \gamma, \gamma, \gamma, 2\gamma, 2\gamma]$.