ZERO-OUTAGE STRATEGY FOR MULTI-ANTENNA BROADCAST CHANNELS WITH LIMITED FEEDBACK

Fabio G. Fernandes, Renato R. Lopes*

University of Campinas - UNICAMP
Department of Communications
C.P 6101, CEP 13083-852
E-mails:{fabiogf,rlopes}@decom.fee.unicamp.br

D. Zanatta Filho†

University of York
Dept. of Electronics
York YO10 5DD, U.K.
E-mail: daniloz@decom.fee.unicamp.br

ABSTRACT

Transmission techniques for the wireless multi-antenna broadcast channel often require that the receivers feed back their channel state information (CSI) to the transmitter. In this paper, we propose a limited feedback method to approximate zero-forcing beamforming. Each user feeds back quantized information about channel direction and a deterministic lower bound on its signal-to-interference-plus-noise ratio (SINR), which require no more than an integer and a real number. With this information, the Base Station performs user scheduling, beamforming and rate adaptation. In this method, the information from both receiver and transmitter sides are taken into account to arrive at a tight lower bound on the supported rate of each user. Since a lower bound on the SINR is fed back, the proposed method avoids outage. We discuss the feedback load of the method, and show numerical results of the relationship between sum-rate and feedback load, SNR and number of users, as well as a comparison with similar methods.

1. INTRODUCTION

Extensive research has been performed on point-to-point wireless multi-antenna channels towards efficiently exploiting the spatial dimension. It has been shown [1] that adding antennas to transmitter and receiver can greatly enhance performance through techniques that exploit the degrees of freedom, in the form of spatial diversity, multiplexing, or a combination of those.

The multi-user approach to wireless multi-antenna channels allows for significant performance gains, based on differences between the channels of the users, which can be exploited by scheduling and precoding techniques. In this paper we focus on multi-antenna broadcast channels, which can model a cell with one base station and many users in a cellular network.

*Fabio Fernandes acknowledges the financial support by CAPES.
†The third author performed the work while at UNICAMP

The nonlinear technique called Dirty-Paper Coding (DPC) [2], based on interference cancellation at the transmitter, was shown to achieve the borders of the capacity region. DPC has a high computational complexity, and linear processing techniques at the transmitter, along with adequate user scheduling, are a suitable alternative. They present lower complexity than DPC, and deal differently with interference. One such technique, zero-forcing beamforming (ZFBF), multiplies the transmitted signals by weight vectors orthogonal to the channels of the other users, thus canceling interference.

A key characteristic of the multi-antenna broadcast channel is the need for channel state information at the transmitter (CSIT). However, only in systems using time-division duplex (TDD) the Base Station (BS) can estimate the downlink channels. In most practical systems, the BS must rely on channel state information (CSI) fed back by the users through a limited-rate feedback channel.

There have been several different approaches to the limited feedback problem for the multi-antenna broadcast channel recently. In [3] and [4], random beamforming is proposed, along with signal-to-interference-plus noise ratio (SINR) feedback from the users. In [5], the authors use transmit correlation, when angle spread is small enough, to add information and reduce feedback load and scheduling complexity. In [6], an approximation to ZFBF is proposed: the users feed back an estimate of their SINRs, as well as the quantized direction of their channels.

In this paper, the Deterministic SINR Lower Bound Feedback (DLBF) is proposed, its performance is analyzed and compared to methods previously proposed. Our scheme, developed independently from [6], is also based on ZFBF. However, there are significant differences between our scheme and the one proposed in [6], which will be highlighted along this paper. The most important is the fact that in DLBF the users feed back a lower bound on the SINR, and thus our method avoids outage. In [6], on the other hand, the estimate of the SINR depends on the expected value of the interference, being consequently subject to outage.

This paper is organized as follows. In Section 2, the system model is detailed. In Section 3 we discuss the problem of quantizing the channel vector. Section 4 is devoted to the development and analysis of the DLBF. Section 5 shows the results obtained with DLBF and compares it with the technique in [6]. Section 6 contains our concluding remarks.

2. SYSTEM MODEL

A multi-user downlink channel is considered in this work, where the BS has N_t antennas and each of the K users has one antenna. We consider that, the BS transmits to at most N_t different users at each block. Transmit beamforming is used to direct the signal power to each served user as well as to avoid interference. The received signal of the k-th user is

$$y_k = \mathbf{w}_k^{\mathrm{H}}\,\mathbf{h}_k\,x_k + \sum_{j=1,j\neq k}^{N_t} \mathbf{w}_j^{\mathrm{H}}\,\mathbf{h}_k\,x_j + n_k, \qquad (1)$$

where $\mathbf{h}_k$ is the flat-fading Rayleigh channel of the k-th user, with each of its components being the gain between a transmit antenna and the receiver, n_k is the white Gaussian noise component at the k-th receive antenna with a variance of σ^2, $\mathbf{w}_k$ is the beamforming weight vector and x_k is the symbol transmitted to the k-th user. It is assumed here that at each block the scheduled users are assigned the same power. We assume block fading, in which the channel is constant for a block, and varies independently from one block to another.

The transmitter has access only to the partial CSI fed back by the users, and must use only this information to perform transmit beamforming, user scheduling (choosing N_t users out of the total K to transmit to at each block) and rate adaptation (choosing at what rate to transmit to each user at each block). In this paper we focus on the limited feedback by the users. As in [6], this feedback is divided into channel direction information (CDI) and channel quality information (CQI). Neither channel estimation errors nor errors due to feedback transmission are treated in this paper. Thus, each receiver is assumed to have full knowledge of its own CSI and the feedback channel is assumed to be error-free.

3. CHANNEL DIRECTION QUANTIZATION

In order to efficiently choose the beamforming vectors, the BS must know the direction of the channel vector of each of the scheduled users. In practice, this information needs to be quantized. In this section, we review the design of a unit-norm-vector codebook to quantize the direction of the channel vectors.

The quantizer used in this paper is inspired by the Grassmann manifold, and was proposed in [7]. In Grassmannian beamforming, the codebook vectors $\{\mathbf{c}_1, \mathbf{c}_2, \mathbf{c}_3, \dots \mathbf{c}_N\}$, where $\|\mathbf{c}_i\| = 1$, for $i = 1 \dots N$, are chosen as the solution to the minimization of the maximum inner product between any two of its vectors

$$\min_{\{\mathbf{c}_1,\mathbf{c}_2,\dots,\mathbf{c}_N\}} \max_{i\neq j} |\mathbf{c}_i^{\mathrm{H}}\mathbf{c}_j|. \qquad (2)$$

Solving (2) is equivalent to picking a set of vectors by maximizing the smallest distance between them. That is, the codebook vectors are as spread as possible, thus providing good covering of the space containing the channel vector. For a given channel vector $\mathbf{h}_k \sim \mathcal{C}N(0, I_{N_t})$, the vector that best approximates its direction in the codebook is chosen as

$$\max_j |\mathbf{h}_k^{\mathrm{H}}\mathbf{c}_j|. \qquad (3)$$

The Grassmanian codebook is shown [7] to maximize (3).

In [7], it is proposed that the chosen vector (the solution to (3)) be the beamforming vector for the k-th user. This can be regarded as an approximation of matched-filtering. Nevertheless, it will be shown in Section 4 that such codebook is appropriate to be used as the approximation of the channel direction. The cardinality N of the codebook is determined by the number of bits B_d dedicated to quantize the direction: $N = 2^{B_d}$.

There is no closed solution to find a Grassmanian codebook, and the optimization problem in (2) is not convex. Therefore, on all the results in Section 5 the simulations were performed using codebooks obtained by solving (2) with MATLAB's genetic algorithm.

4. DETERMINISTIC LOWER BOUND FEEDBACK

In ZFBF with full CSIT, up to N_t users are scheduled at each block, and the beamforming vectors are chosen so that

$$|\mathbf{w}_i^{\mathrm{H}}\mathbf{h}_k| = 0 \qquad \forall \quad i \neq k. \qquad (4)$$

In a system with limited feedback, the BS does not know $\mathbf{h}_k$, but only its quantized direction $\mathbf{c}_k$, computed in (3). In this case, the beamforming vector is calculated so that

$$|\mathbf{w}_i^{\mathrm{H}}\mathbf{c}_k| = 0 \qquad \forall \quad i \neq k. \qquad (5)$$

Assuming that all users are assigned the same power, we can write the supported rate for the k-th user as

$$R_k = \log_2\left(1 + \frac{\frac{P}{N_t}\,|\mathbf{w}_k^{\mathrm{H}}\,\mathbf{h}_k|^2}{\sigma^2 + \frac{P}{N_t}\sum_{j=1,j\neq k}^{N_t} |\mathbf{w}_j^{\mathrm{H}}\,\mathbf{h}_k|^2}\right). \qquad (6)$$

However, the users cannot compute the rate as in (6), since the beamforming vectors are not available to them. Thus, we manipulate the expression to find a computable value for the user. To this end, we rewrite the channel vector $\mathbf{h}_k$ as

$$\mathbf{h}_k = ||\mathbf{h}_k||\,\tilde{\mathbf{h}}_k = ||\mathbf{h}_k||\,(a_k\,\mathbf{c}_k + \bar{a}_k\bar{\mathbf{c}}_k), \qquad (7)$$

where $\tilde{\mathbf{h}}_k$ is the normalized channel vector, a_k is the component of $\tilde{\mathbf{h}}_k$ in the direction of $\mathbf{c}_k$, $\bar{\mathbf{c}}_k$ is the unit vector orthogonal to $\mathbf{c}_k$ in the plane formed by $\mathbf{c}_k$ and $\tilde{\mathbf{h}}_k$, and $\bar{a}_k$ is the component of $\tilde{\mathbf{h}}_k$ in the direction of $\bar{\mathbf{c}}_k$. Applying (7) to (6) and using (5), we can write R_k as

$$\log_2\left(1+\frac{\frac{P}{N_t}\,\|\mathbf{h}_k\|^2\,|a_k\,\mathbf{w}_k^{\mathrm{H}}\,\mathbf{c}_k+\bar{a}_k\,\mathbf{w}_k^{\mathrm{H}}\,\bar{\mathbf{c}}_k|^2}{\sigma^2+\frac{P}{N_t}\,\|\mathbf{h}_k\|^2\,\sum_{j=1,j\neq k}^{N_t}\,|\bar{a}_k\,\mathbf{w}_j^{\mathrm{H}}\,\bar{\mathbf{c}}_k|^2}\right) \quad (8)$$

The BS uses the supported rate of each user to schedule the best set of users at a given block and to adapt their transmission rates. However, (8) cannot be calculated exactly either by the BS, since $\mathbf{h}_k$ is unknown, or by the users, since they do not know the beamforming vectors. Thus, the BS must rely on some estimate, based on the information received by the users' feedback. In this paper, we propose that the users feed back a lower bound on the SINR.

To calculate a lower bound at each receiver we impose the restriction that $|\mathbf{w}_k^{\mathrm{H}}\,\mathbf{c}_k|\geq B$, where B is a pre-defined threshold value, known to the BS and to all users. This is assumed to be true by all users, who take this into consideration to compute the lower bound. Note that the actual value of $|\mathbf{w}_k^{\mathrm{H}}\,\mathbf{c}_k|$ depends on the set of scheduled users, since each beamforming vector depends only on the quantized channels of the other scheduled users. Therefore, the BS should only consider the sets of users that meet this requirement. Note also that no outage can be caused by this requirement. If no set of N_t users meets this condition, the base station may search for a subset of N_t-1 users, and so on. In the worst case, any subset of one user can be used, since in this case, the beamforming vector can be chosen to have $|\mathbf{w}_k^{\mathrm{H}}\,\mathbf{c}_k|=1$.

In fact, as shown in [6], this restriction is equivalent to requiring that the chosen codebook vectors of the scheduled users be ϵ-orthogonal, that is:

$$|\mathbf{c}_k^{\mathrm{H}}\mathbf{c}_j|\leq\epsilon, \quad (9)$$

where ϵ and the threshold B have the following relation:

$$B=\frac{(1+\epsilon)\,(1-(N_t-1)\,\epsilon)}{1-(N_t-2)\,\epsilon} \quad (10)$$

Hence, it is possible to implement DLBF using the semi-orthogonal user selection algorithm (SUS) [8].

In the following theorem, we propose a deterministic lower bound on the SINR, which can be calculated by using local information on both receiver and transmitter sides.

Theorem 1. *Let* $\alpha_k=|a_k|=|\tilde{\mathbf{h}}_k^{\mathrm{H}}\,\mathbf{c}_k|$, $\beta_k=|\mathbf{w}_k^{\mathrm{H}}\,\mathbf{c}_k|$ *and* $\bar{\alpha}_k=|\bar{a}_k|=|\tilde{\mathbf{h}}_k^{\mathrm{H}}\,\bar{\mathbf{c}}_k|$. *Let* $\bar{B}=\sqrt{1-B^2}$. *Then, if* $\beta_k\geq B$ *for all of the* N_t *scheduled users,*

$$\hat{R}_k = \log_2\left(1+\frac{\beta_k^2}{B^2}\,\frac{\frac{P}{N_t}\,\|\mathbf{h}_k\|^2\,(\alpha_k^2\,B^2-2\,\alpha_k\,\bar{\alpha}_k\,B\,\bar{B})}{\sigma^2+\frac{P}{N_t}\,\|\mathbf{h}_k\|^2\,\gamma\,\bar{\alpha}_k^2}\right) \leq R_k. \quad (11)$$

The constant γ measures the residual interference caused by the quantization of the CDI, and should be computable by the users. It replaces the value $\sum_{j=1,j\neq k}^{N_t}\,|\mathbf{w}_j^{\mathrm{H}}\,\bar{\mathbf{c}}_k|^2$ and is defined differently depending on the limited feedback technique. In our lower bound, we consider the worst-case estimate of the interference, yielding $\gamma=1$ for $N_t=2$ and $\gamma=1-\bar{B}$ for $N_t=3$.

Proof. See Appendix A.

□

Here we highlight a significant difference between DLBF and the method proposed in [6]. In [6], the users feed back an estimate of the SINR. This estimate is similar in form to (11), and in the denominator the users compute the expected value of the interference. In this case, the constant γ is given by $\gamma=\mathrm{E}\left\{\sum_{j=1,j\neq k}^{N_t}\,|\mathbf{w}_j^{\mathrm{H}}\,\bar{\mathbf{c}}_k|^2\right\}=1$ for any N_t. With this, the actual interference may be greater than the estimated and thus the supported rate lower than the estimated. If the BS schedules an user to receive at a rate greater than its supported rate, an outage occurs. Thus, the estimate proposed in [6] is not a lower bound, and there is a non-null probability of outage.

From (11), we see that the supported rates grow as $\alpha_k\rightarrow 1$, not only because the power of the desired signal increases, but also because the interference decreases (since $\bar{\alpha}_k\rightarrow 0$). This is intuitive, since α_k measures how well the CDI is aligned with the actual channel. Further, we see that Grassmanian codebooks are optimal in the sense that they maximize (3) and hence maximize α_k.

It should be noted that in the best possible case, where there is no vector quantization error ($\alpha_k=1$) and the beamforming vector is aligned to the channel vector ($\beta_k=1$), this lower bound tends to the actual power of the desired signal. Thus, it is tighter for the users that are more likely to be scheduled.

Note that the lower bound in (11) cannot be calculated by the users, since the values of β_k are not known to them. Thus, they feed back the following SINR lower bound to the BS:

$$\frac{\frac{P}{N_t}\,\|\mathbf{h}_k\|^2\,(\alpha_k^2\,B^2-2\,\alpha_k\,\bar{\alpha}_k\,B\,\bar{B})}{\sigma^2+\frac{P}{N_t}\,\|\mathbf{h}_k\|^2\,\gamma\,\bar{\alpha}_k^2}. \quad (12)$$

The BS, on the other hand, has access to the actual values of β_k, and can easily calculate (11) by multiplying (12) by $\frac{\beta_k^2}{B^2}$ to obtain

$$\hat{R}_k=\log_2(1+\frac{\frac{P}{N_t}\,||\mathbf{h}_k||^2\,(\alpha_k^2\,\beta_k^2-2\,\alpha_k\,\bar{\alpha}_k\,\frac{\beta_k^2}{B^2}\,\bar{B})}{\sigma^2+\frac{P}{N_t}\,||\mathbf{h}_k||^2\,\gamma\,\bar{\alpha}_k^2})\leq R_k. \quad (13)$$

Thus, the BS introduces additional information into the lower bound by taking into account the alignment of the channel and the beamforming vectors. This tightens the lower bounds and makes them less dependent on the choice of the value of the threshold B. Since the optimal value of B depends on the number of users, the SNR and the fading of each user, the above characteristic is very desirable in practice.

We summarize DLBF in Table 1, in a step-by-step description of the technique.

Table 1. Description of the DLBF

1. Each user (perfectly) estimates his channel
2. The codebook vector $\mathbf{c}_k$ is chosen as in eq. (3)
3. The SINR lower bound is calculated as in eq. (12)
4. The index of the chosen codebook vector (an integer) and the SINR lower bound (a real number) are sent to the BS through an error-free channel
5. The BS tightens the lower bounds using the information regarding the alignment of the channel and the beamforming vectors
6. The BS schedules a subset of up to N_t users to transmit to during this block according to the sum-rate maximization criterion
7. The BS calculates the beamforming vectors and the rates allocated to each scheduled user
8. The BS sends the signals to the users

5. RESULTS

We have simulated a multi-antenna broadcast channel in two different scenarios: two and three transmit antennas. In all simulations, we take into account the information fed back by the users not only for the purposes of beamforming and user scheduling, but also for rate adaptation. This means that the actual rate transmitted by the BS to the k-th user is $\hat{R}_k$ as in (11), whereas R_k in (6) is the supported rate of the channel for a given set of scheduled users and beamforming vectors. The latter, nonetheless, is unknown to the BS in practice, thus we assume $\hat{R}_k$ to be the actual rate.

In the first scenario, we use a transmitter with two antennas and schedule up to two users per block by maximizing the sum-rate through exhaustive search among the possible sets of semi-orthogonal users. The total SNR is 10dB. We consider the quantization of both CDI and CQI and obtain results regarding the feedback load. In Figure 1, we plot the actual sum rate for several different feedback loads. Each curve corresponds to the best distribution of the feedback bits between CDI and CQI, considering a given allowance of feedback bits. The feedback of the CQI, in the form of the SINR lower bound as in (12), was quantized by maximizing the expected value of the rate through MATLAB's genetic algorithm, since the cost function of the maximization is not concave. Note from Figure 1 that it is usually better to increase the number of CDI bits than that of the CQI. In fact, when the total feedback load is up to 7 bits per user, the best strategy is to assign at most 2 bits for CQI feedback. This follows from the fact that quantizing a vector requires more levels than quantizing a scalar.

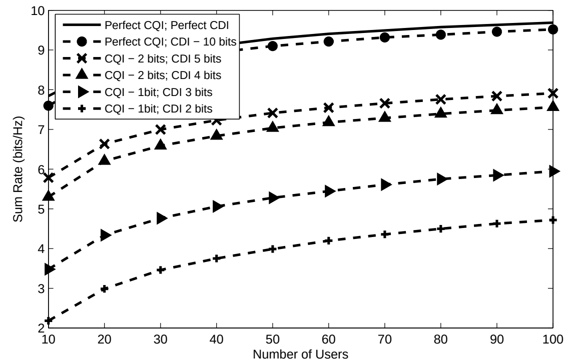

Fig. 1. Sum rates for various feedback loads for $N_t = 2$.

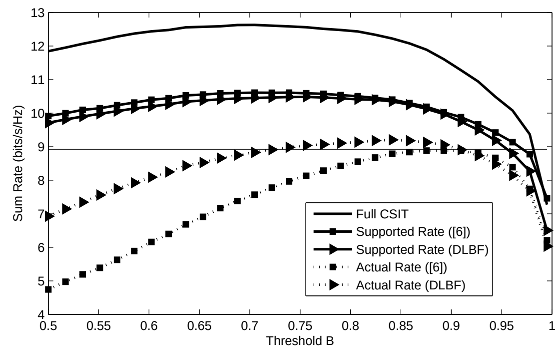

Fig. 2. Sum rate as a function of the threshold B.

In the second scenario, we assume the CQI feedback to be unquantized, and take into consideration the variations of the sum rate with the SNR, number of users, number of CDI feedback bits and the threshold B, with three antennas in the BS. The user scheduling algorithm used is the semi-orthogonal user scheduling (SUS) proposed in [6], due to its lower computational complexity.

In this scenario we plot curves for both the supported sum-rate R_k in (6) and the actual sum-rate $\hat{R}_k$ in (11), in order to show the distance between the SINR estimates and their real values. Unless otherwise specified, the total SNR is fixed at 10dB, the threshold B is fixed at 0.824, 9 bits are used to quantize CDI and the number of users is 100.

The sensitivity to the threshold B is shown in Figure 2. Due to the addition of information to the SINR lower bound at the BS, DLBF is less sensitive to changes in the threshold. The horizontal line in the figure marks the maximum sum rate achieved by the method in [6]. Clearly, DLBF achieves larger values than [6] for any threshold B between 0.73 and 0.91.

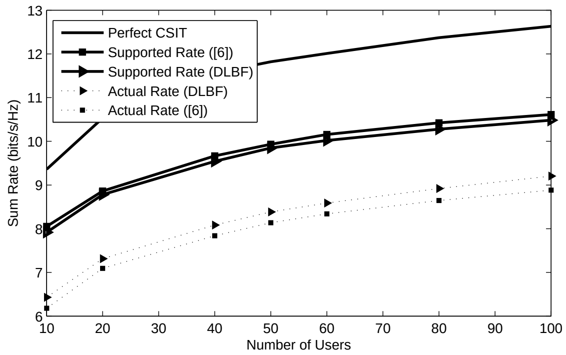

Fig. 3. Sum rate as a function of the number of users.

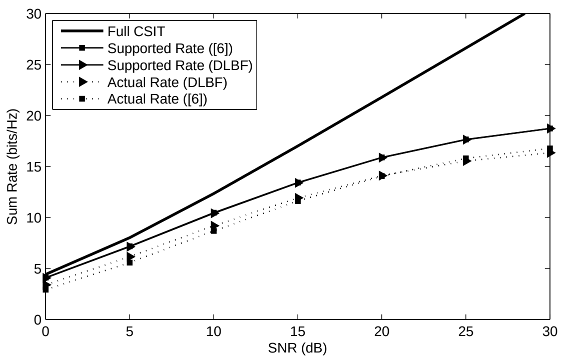

Fig. 4. Sum rate as a function of the total SNR.

We also see that the difference between the maximum rate achieved by DLBF and the maximum achieved by [6] is 0.285 bits/s/Hz.

In Figure 3, we show the sum-rate as a function of the number of users in the system. In this figure, we choose the threshold values that achieve the best rates for each number of users. Here we again compare the performance between DLBF and the method proposed in [6]. We see that DLBF outperforms [6] by around 3%. We can also observe that the rate curve for DLBF is approximately parallel to the one for Perfect CSIT, which indicates that DLBF successfully exploits multi-user diversity.

Figure 4 shows the sum rate as a function of the SNR. Both DLBF and the method in [6] start to saturate at high SNR, since residual interference plays an important role in limiting the sum rate growth. Nevertheless, DLBF starts to saturate at a lower SNR than the method in [6], since the interference estimate to compute the rate is larger in the former.

In Figure 5, the sum rate is plotted as a function of the number of CDI bits. We note that the curve for DLBF grows faster than the one for [6]. We also observe that DLBF achieves a sum rate of 8.6 bits/s/Hz with 8 CDI bits, whereas 9 bits are needed for the method in [6] to achieve the same rate.

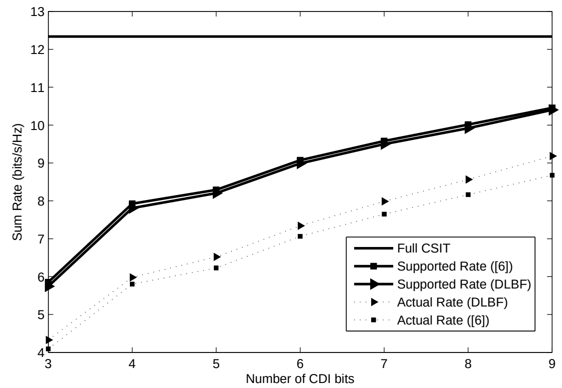

Fig. 5. Sum rate as a function of the number of CDI bits.

6. CONCLUSION

In this paper we proposed a method of limited CSI feedback for the broadcast channel, the DLBF, based on a deterministic lower bound of the instantaneous SINR of each user. We derived the lower bound and showed that it can be computed in practice by using information from both user and BS sides.

Since we use the SINR lower bound to calculate the rates to be transmitted at each block, it is guaranteed that these rates are always below the supported rates, and thus we completely avoid outage. Nonetheless, the results of our numerical simulations show that DLBF achieves fairly good performance, outperforming the method proposed in [6] for SNR less than 20 dB in the simulated scenarios.

7. REFERENCES

[1] D. Tse and P. Viswanath, *Fundamentals of Wireless Communications*, Cambridge University Press, first edition, 2005.

[2] M. Costa, "Writing on dirty paper (corresp.)," *Information Theory, IEEE Transactions on*, vol. 29, no. 3, pp. 439–441, May 1983.

[3] P. Viswanath, D. N. C. Tse, and R. Laroia, "Opportunistic beamforming using dumb antennas," *Information Theory, IEEE Transactions on*, vol. 48, no. 6, pp. 1277–1294, Jun 2002.

[4] M. Sharif and B. Hassibi, "On the capacity of mimo broadcast channels with partial side information," *Information Theory, IEEE Transactions on*, vol. 51, no. 2, pp. 506–522, Feb. 2005.

[5] M. Kountouris, D. Gesbert, and L. Pittman, "Transmit correlation-aided opportunistic beamforming and

scheduling using reciprocal channel information in multiuser mimo scheduling," in *"Proc. EUSIPCO Conference, Florence, Italy"*, 2006.

[6] N. Jindal T. Yoo and A. Goldsmith, "Multi-antenna broadcast channels with limited feedback and user selection," *IEEE Journal on Selected Areas in Communications*, vol. 25, pp. 1478–1491, September 2007.

[7] D. J. Love, , R. W. Heath Jr., and T. Strohmer, "Grassmannian beamforming for multiple-input multiple-output wireless systems," *Information Theory, IEEE Transactions on*, vol. 49, no. 10, pp. 2735–2747, Oct. 2003.

[8] T. Yoo and A. Goldsmith, "On the optimality of multiantenna broadcast scheduling using zero-forcing beamforming," *Selected Areas in Communications, IEEE Journal on*, vol. 24, no. 3, pp. 528–541, March 2006.

A. PROOF OF THEOREM 1

The SINR lower bound obtained in (11) can be divided into a lower bound on the power of the desired signal and an upper bound on interference. We will start with the former. We can rewrite the power of the desired signal in (8) as

$$\begin{aligned}\frac{P}{N_t}\,|\mathbf{w}_k^{\mathrm{H}}\mathbf{h}_k|^2 &= \frac{P}{N_t}\,\|\mathbf{h}_k\|^2\,|\mathbf{w}_k^{\mathrm{H}}\tilde{\mathbf{h}}_k|^2 = \\ &= \frac{P}{N_t}\,\|\mathbf{h}_k\|^2\,|a_k\,\mathbf{w}_k^{\mathrm{H}}\mathbf{c}_k + \bar{a}_k\,\mathbf{w}_k^{\mathrm{H}}\bar{\mathbf{c}}_k|^2\end{aligned} \tag{14}$$

Using the inequality $|x+y| \geq \big||x|-|y|\big|$, and the definitions $\alpha_k = |a_k|$, $\bar{\alpha}_k = |\bar{a}_k|$, $\beta_k = |\mathbf{w}_k^{\mathrm{H}}\mathbf{c}_k|$, and $\bar{\beta}_k = |\mathbf{w}_k^{\mathrm{H}}\bar{\mathbf{c}}_k|$, we state that

$$\frac{P}{N_t}\,\|\mathbf{h}_k\|^2\,|a_k\,\mathbf{w}_k^{\mathrm{H}}\mathbf{c}_k + \bar{a}_k\,\mathbf{w}_k^{\mathrm{H}}\bar{\mathbf{c}}_k|^2 \geq \tag{15a}$$

$$\frac{P}{N_t}\,\|\mathbf{h}_k\|^2\,|\alpha_k\,\beta_k - \bar{\alpha}_k\,\bar{\beta}_k|^2 = \tag{15b}$$

$$\frac{P}{N_t}\,\|\mathbf{h}_k\|^2\left(\alpha_k^2\,\beta_k^2 - 2\,\alpha_k\,\bar{\alpha}_k\,\beta_k\,\bar{\beta}_k + \bar{\alpha}_k^2\,\bar{\beta}_k^2\right) \geq \tag{15c}$$

$$\frac{P}{N_t}\,\|\mathbf{h}_k\|^2\left(\alpha_k^2\,\beta_k^2 - 2\,\alpha_k\,\bar{\alpha}_k\,\beta_k\,\bar{\beta}_k\right) \geq \tag{15d}$$

$$\frac{P}{N_t}\,\|\mathbf{h}_k\|^2\left(\alpha_k^2\,B_k^2 - 2\,\alpha_k\,\bar{\alpha}_k\,B\,\bar{B}\right) \tag{15e}$$

In the second inequality, we removed the last term, $\bar{\alpha}_k^2\ \bar{\beta}_k^2$, which is strictly nonnegative. Although this removal causes the lower bound to loosen, it permits that a tightening be performed at the BS, as will be shown next. This procedure is worthy when the quantization error $\bar{\alpha}$ is much smaller than α, which usually happens. The last inequality comes from the requirement that $\beta_k \geq B$ which in turn implies that $\bar{\beta}_k \leq \bar{B}$. It is necessary to use this inequality because the users have no access to the actual value of β_k.

This lower bound can be tightened, however, by multiplying (15e) by $(\frac{\beta_k}{B})^2$ at the BS:

$$\left(\frac{\beta_k}{B}\right)^2\frac{P}{N_t}\,\|\mathbf{h}_k\|^2\left(\alpha_k^2\,B_k^2 - 2\,\alpha\,\bar{\alpha}\,B\,\bar{B}\right) = \tag{16a}$$

$$\frac{P}{N_t}\,\|\mathbf{h}_k\|^2\left(\alpha_k^2\,\beta_k^2 - 2\,\alpha_k\,\bar{\alpha}_k\left(\frac{\beta_k^2}{B}\right)\bar{B}\right) = \tag{16b}$$

$$\frac{P}{N_t}\,\|\mathbf{h}_k\|^2\left(\alpha_k^2\,\beta_k^2 - 2\,\alpha_k\,\bar{\alpha}_k\,\beta_k\,\bar{B}\left(\frac{\beta_k}{B}\right)\right) \leq \tag{16c}$$

$$\frac{P}{N_t}\,\|\mathbf{h}_k\|^2\left(\alpha_k^2\,\beta_k^2 - 2\,\alpha_k\,\bar{\alpha}_k\,\beta_k\,\bar{\beta}\right) \tag{16d}$$

Note that (16a) is a tighter lower bound than (15e), since it is multiplied by a value greater than 1. In the last inequality, nonetheless, it is shown that (16a) is still a lower bound, since (15d) and (16d) are the same expression.

Now we focus on the upper bound of the interference. We define the constant $\gamma = \sum_{j=1,j\neq k}^{N_t} |\mathbf{w}_j^{\mathrm{H}}\,\bar{\mathbf{c}}_k|^2$, which cannot be calculated by the users since the beamforming vectors are unknown to them. Therefore, we use the worst-case estimate in our SINR lower bound:

$$\max_{\{\mathbf{w}_1,\ldots,\mathbf{w}_{k-1},\mathbf{w}_{k+1},\ldots,\mathbf{w}_{N_t}\}} \sum_{j=1,j\neq k}^{N_t} |\mathbf{w}_j^{\mathrm{H}}\,\bar{\mathbf{c}}_k|^2 \tag{17}$$

$$\text{subject to } \beta_i \geq B \quad \forall\; i = 1\ldots N_t.$$

Let $\mathbf{W} = [\mathbf{w}_1,\,\mathbf{w}_2,\ldots,\mathbf{w}_{k-1},\,\mathbf{w}_{k-1},\ldots,\mathbf{w}_{N_t}]$. Then, the maximization problem in (17) can be written as

$$\max_{\mathbf{W}}\ \bar{\mathbf{c}}_k^{\mathrm{H}}\left(\mathbf{W}\,\mathbf{W}^{\mathrm{H}}\right)\bar{\mathbf{c}}_k \tag{18}$$

$$\text{subject to } \beta_i \geq B \quad \forall\; i = 1\ldots N_t.$$

We suppose, without loss of generality, that $k = 1$. For the case where $N_t = 2$, $\mathbf{W} = \mathbf{w}_2$, and $\mathbf{w}_2 = e^{j\,\theta}\bar{\mathbf{c}}_1$ maximizes (18), for any θ. Therefore the upper bound is $\gamma = 1$. Note that for $N_t = 2$, $\gamma = 1$ is the actual value of γ for any realization. Since $\mathbf{c}_1$ and $\bar{\mathbf{c}}_1$ are an orthonormal base of $\mathcal{C}^2$, $|\mathbf{w}_2^{\mathrm{H}}\,\mathbf{c}_1| = 0$, and $\|\mathbf{w}_2\| = 1$, then $|\mathbf{w}_2^{\mathrm{H}}\,\bar{\mathbf{c}}_1| = 1$.

For the general case, (18) is equivalent to finding the largest eigenvalue of the matrix $\mathbf{W}\,\mathbf{W}^{\mathrm{H}}$. It is equivalent, in turn, to work with the matrix $\mathbf{W}^{\mathrm{H}}\,\mathbf{W}$, since the non-zero eigenvalues of both matrices are equal. Therefore, we can find an upper bound on this value by solving the characteristic polynomial equation of this matrix, which is composed of terms in the form $\mathbf{w}_i^{\mathrm{H}}\mathbf{w}_j$, where

$$|\mathbf{w}_i^{\mathrm{H}}\mathbf{w}_j| \;=\; 1 \quad \text{for} \quad i = j \tag{19a}$$

$$|\mathbf{w}_i^{\mathrm{H}}\mathbf{w}_j| \;=\; |b_{ij}| \leq \bar{B} \quad \text{for} \quad i \neq j. \tag{19b}$$

The inequality in (19b) is proved through the decomposition of $\mathbf{w}_i$ in the orthogonal basis $\{\mathbf{c}_i,\,\mathbf{w}_j,\check{\mathbf{w}}_i\}$:

$$\mathbf{w}_i = \beta_i \, \mathbf{c}_i + b_{ij} \, \mathbf{w}_j + \check{\mathbf{w}}_i \tag{20}$$

where $j \neq i$ and $\check{\mathbf{w}}_i$ is a vector orthogonal to c_i and w_j in the direction of w_i. Since $\|\mathbf{w}_i\| = 1$ and $|\beta_i| \geq B$, we have

$$\|\mathbf{w}_i\|^2 = \beta_i^2 + |b_{ij}|^2 + \|\check{\mathbf{w}}_i\|^2 \quad = \quad 1 \tag{21}$$
$$|b_{ij}|^2 = 1 - \beta^2 - |\check{\mathbf{w}}_i\|^2 \leq 1 - B^2 \quad = \quad \bar{B}^2 \tag{22}$$

Thus, we can solve the characteristic equation of $\mathbf{W}^{\mathrm{H}} \, \mathbf{W}$ to find the largest possible eigenvalue. For $N_t = 3$, we have

$$\det\left(\mathbf{W}^{\mathrm{H}} \, \mathbf{W} - \lambda \, I\right) \quad = \quad 0 \tag{23a}$$
$$(1-\lambda)^2 - (\mathbf{w}_2^{\mathrm{H}} \, \mathbf{w}_3)\,(\mathbf{w}_3^{\mathrm{H}} \, \mathbf{w}_2) \quad = \quad 0 \tag{23b}$$
$$1 - \lambda \quad = \quad \pm|\mathbf{w}_2^{\mathrm{H}} \, \mathbf{w}_3| \tag{23c}$$
$$\lambda_{max} \quad = \quad 1 + |\mathbf{w}_2^{\mathrm{H}} \, \mathbf{w}_3| \tag{23d}$$

Thus, for $N_t = 3$, $\lambda_{max} \leq 1 + \bar{B}$.