MIMO Broadcast Scheduling with Limited Channel State Information

Charles Swannack†, Elif Uysal-Biyikoglu⋆, and Gregory W. Wornell†

†*Signals, Information and Algorithms Laboratory, MIT, Cambridge, MA 02139, USA*
⋆*Dept. of Elect. and Comp. Engineering, The Ohio State University, Columbus, OH 43210, USA*

Abstract

We consider the multiple-input multiple-output (MIMO) broadcast channel in which there are m transmit antennas and n uncoordinated users with a single receive antenna. We examine the maximum throughput in such a system in the scenario when the number of users is much greater than the number of transmit antennas. We derive a lower bound the probability that there exists a set of users such that each user receives a specified rate. We show that through the use of a simple norm-threshold feedback protocol the maximal scaling of the sum rate is achievable and requires channel state information (CSI) about only $O(1)$ users.

1 Introduction

The wireless downlink has been a source of interesting problems in recent years. The question of how to most efficiently schedule independent data streams that are intended for multiple receivers is of substantial interest due to its practical relevance[1]. As noted in [1–3], the problem is especially rich when there are multiple antennas at the transmitter, that is, in a *multiuser* MIMO broadcast channel. Making full use of the capacity region of this channel model requires *multiplexing.* When there are more users than the channel dimension, the sum-rate maximizing set of users will be a subset of the user pool, and this subset needs to be selected depending on the channel state. Hence, the scheduling and multiplexing need to be done jointly. Of course, searching the complete set of users at every scheduling instant is not scalable. The motivation for the work in this paper is finding a scalable search algorithm, both in terms of computational complexity, and in terms of channel state information at the transmitter, that guarantees a performance close to optimal.

With perfect channel state information (CSI) at the transmitter, a throughput maximizing scheduling/multiplexing scheme is to employ Dirty-Paper coding on the set of users which at that time can achieve the highest sum rate [3]. Insights on throughput optimization under *complexity-reducing restrictions* have appeared in [1, 3–7]. It has been recognized that using sets for which there are guarantees on the channel norms and the magnitudes of inner products such that users' channels are near-orthogonal can provide close to optimal performance [1, 3, 5–7]. Such approaches have recently led to bounding throughput and developing low complexity section algorithms using graph based approaches [3, 5, 8–10].

[1] This setup could describe a network of sensors which receive data from a centralized location, for example, or a group of subscribers that download data from a fixed wireless transmitter.

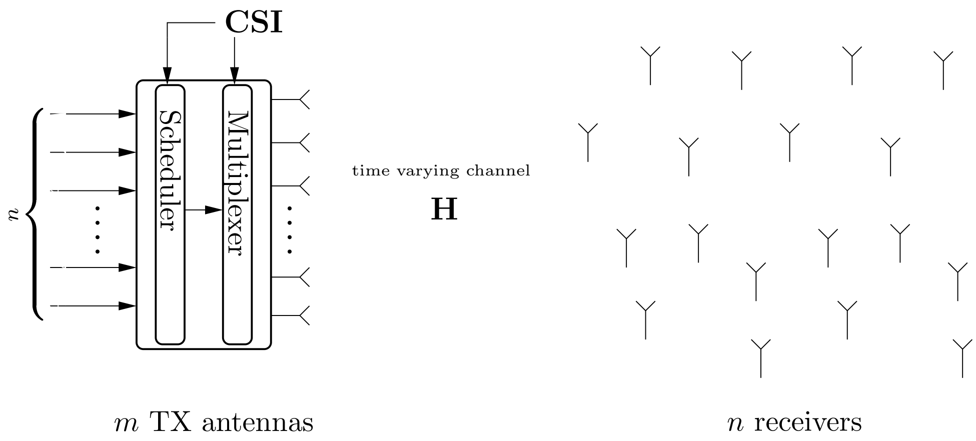

Figure 1: The MIMO downlink system overview

In [3], we showed that the search complexity may be reduced by limiting attention to a small subset of users. More explicitly we found a lower bound to the probability that a set with a given SINR can be found by a search among a random choice of users, and showed that the probability exhibits a phase transition at a small multiple of the transmit dimension after which it is close to 1. We showed that consequently, even with a sub-optimal multiplexing scheme such as zero-forcing beamforming, the asymptotic scaling of throughput with the number of users n is on the order of $\log \log n$, which is known to be the optimal scaling rate. However, $\log \log n$ increases so slowly that in practice not only the scaling law but constant factors in the rate expression will matter. Hence, *the asymptotic scaling is not the primary issue of interest.*

In this paper, we argue that the central question of interest is *how many users need to be considered before a certain sum rate is guaranteed*. This is also intimately tied to the question of *how much CSI* is needed at the transmitter to guarantee a certain sum rate since if only a small subsets need to be examined limiting the CSI at the transmitter to a small number of channels is sufficient.

To analyze the behavior of the maximal sum rate with CSI, we consider the probability distribution function of the maximal sum rate, and bound it under varying degrees of CSI. Specifically, the outline of the paper is as follows. In Section 3, we provide a tight bound on the probability distribution function of sum rate with zero-forcing beamforming, with complete CSI at the transmitter. In Section 4, we limit the number of users whose channel states are to be known by the transmitter. We analyze what that number should be to guarantee exceeding a certain rate with zero-forcing multiplexing. We conclude by suggesting an algorithm for multiuser scheduling/multiplexing.

2 Preliminaries

The problem addressed in this paper is obtaining the probability distribution of sum rate as a function the amount of information at the scheduler about the users' channels. In order to state the problem more formally, let us first make some definitions about the system model, scheduling method and multiplexing.

2.1 Channel model

We consider a broadcast channel in which n independent data streams arrive at a transmit base equipped with m-antennas, and are intended for n uncoordinated users each equipped with a single receive antenna (see Figure 1). The data streams for different users are separately queued at the transmitter, to be delivered to respective receivers. We will assume the queues are under infinite backlog[2]. Further, we will assume that the transmitter has a peak power constraint P.

Let $\mathbf{h}_i$ be the vector such that the jth coordinate of $\mathbf{h}_i$ contains the fading coefficient between the jth transmit antenna and the ith receive antenna. We assume that the $\mathbf{h}_i$'s are independent, and complex zero mean Gaussian with variance $1/2m$. We make the block-fading assumption where the fading coefficients are constant over blocks of length T and independent from block to block. Hence, every T time units, the channel vector for each user is drawn $i.i.d$ with $\mathbf{h}_i \sim \mathcal{N}_{\mathbb{C}}\left(0, \frac{1}{2m}\mathbf{I}\right)$. We assume that T is sufficiently long that error probabilities for any user can be made arbitrarily small. Each receiver has perfect knowledge of their channel state. In this paper we will consider the effects of varying degrees of CSI at the transmitter on the distribution function of the rate.

2.2 Scheduling and Multiplexing

We now discuss those aspects of the user selection and multiplexing that will be needed for the rest of the paper. A more detailed discussion of the multiplexing scheme used is provided in the Appendix.

As channel state is constant during blocks of length T, we can restrict attention without loss of generality to scheduling policies that select a group of users to precode every T time units. Henceforth, T will be the duration of a *scheduling interval*. The scheduler is located at the transmitter. At the beginning of each interval, the scheduler will choose a subset of users to precode among users for which it has channel state information. The case of perfect transmitter CSI is the special case where this subset is the whole user pool, which we will denote by $\mathcal{U}$. In the perfect CSI case we have the scheduler choose the set $\mathcal{A}$ such that

$$\mathcal{A} = \arg\max_{\mathcal{A}\subset\mathcal{U}} |\mathcal{A}|\ \hat{f}_{\text{zf}}(\mathcal{A}) \quad \text{where} \quad \hat{f}_{\text{zf}}(\mathcal{A}) = \log\left(1 + \frac{P}{\text{Tr}(W_{\mathcal{A}}^{-1})}\right) \tag{1}$$

In (1), the function $\hat{f}_{\text{zf}}$ is the sum rate, $\sum_{i\in\mathcal{A}} R_i$ achievable by zero-forcing beamforming[3] under *constant power allocation*. We derive this function in the Appendix and argue that the flat power allocation is not only a convenient choice but one that does not result in significant loss when there is selection from multiple users, as in our case.

We will consider efficient, achievable methods for selecting a set of users that is close to the maximum rate. We thus, in the next section, begin by finding a lower bound on the throughput using zero forcing multiplexing and constant power allocation, under perfect CSI. After that, we use this bound to develop a protocol for limiting the amount of CSI at the transmitter.

[2]This simplifying assumption allows us to focus on user selection and multiplexing. Incorporating the queuing processes is not difficult, while being essentially independent of the issues we address in this paper. The reader can refer to [3] for a treatment.

[3]Please see the Appendix

3 Sum rate under perfect transmitter CSI

In this section we obtain an upper bound on the distribution function of the sum rate under zero-forcing multiplexing with flat power allocation and *perfect CSI at the transmitter*. The behavior of this bound yields important insights into the question of how *much* information is necessary at the transmitter to approximate the maximal obtainable sum rate for multiuser MIMO scheduling. In particular, we show that there is a sharp phase transition in the distribution of bound on the maximal rate whose location is $O(\log\log n)$.

We begin by considering the case of selecting sets of a fixed size l where $0 < l \leq m$. Let $R^l_{\text{sum}}(n)$ be the random variable that takes the value of the maximal achievable rate at a given scheduling interval in a system with n users by selecting a set of size l. That is,

$$R^l_{\text{sum}}(n) \triangleq \max_{\mathcal{A}\subset\mathcal{U}:|\mathcal{A}|=l} |\mathcal{A}|\ \hat{f}_{\text{zf}}(\mathcal{A})$$

Note that when only considering a single set of users we can, by lower bounding $1/\text{Tr}\left(\mathbf{W}^{-1}_{\mathcal{A}}\right)$ by the minimal eigenvalue, bound the pdf of $R^l_{\text{sum}}(l)$ as

$$\begin{aligned} p_l(r) \triangleq \Pr(R^l_{\text{sum}}(l) > r) &= \Pr\left(\frac{1}{\text{Tr}\left(\mathbf{W}^{-1}_{\mathcal{A}}\right)} > \frac{1}{P}\left(\exp\left(\frac{r}{l}\right) - 1\right)\right) \\ &\geq \Pr\left(\lambda_{\min}(\mathbf{W}^{-1}_{\mathcal{A}}) > \frac{l}{P}\left(\exp\left(\frac{r}{l}\right) - 1\right)\right) \end{aligned} \tag{2}$$

We consider the case in which the scheduler chooses a set of size $l = m$, *i.e.* a set with size equal to the number of transmit antennas. These results can be easily extended to the general case of $0 < l \leq m$. In the case that $|\mathcal{A}| = m$ the smallest eigenvalue of $\mathbf{W}^{-1}_{\mathcal{A}}$ is exponentially distributed[4] with parameter $m/2$ [12]. Thus, using the pdf on the smallest eigenvalue of a complex Wishart matrix we have in the case that $|\mathcal{A}| = m$,

$$p_m(r) \geq \exp\left(\frac{-m^2}{2P}\left(\exp\left(\frac{r}{m}\right) - 1\right)\right) \tag{3}$$

The main result of this section is the following bound on the distribution of the maximum rate under zero forcing multiplexing with subset selection.

Theorem 3.1. *For a scheduler with perfect knowledge of all n users' channel vectors, under zero-forcing with flat power allocation, the probability that sum rate is greater than r is lower bounded as:*

$$\Pr(R^m_{\text{sum}} \geq r) \geq 1 - \exp\left(-\left\lfloor\frac{n}{m}\right\rfloor \max\left\{\frac{p_m(r)}{2(1-p_m(r))}, \log\left(1 + \frac{p_m(r)}{1-p_m(r)}\right)\right\}\right) \tag{4}$$

To interpret the behavior of this bound as a function of both the target rate r and the number of users n we substitute (3) in to the right hand side of (4) and treat the inequality in the resulting bound as

$$\Pr(R^m_{\text{sum}} \geq r) \approx 1 - \exp\left(\frac{-n}{\exp\exp\frac{r}{m}}\right). \tag{5}$$

[4] Recall that the smallest eigenvalue of a complex Wishart matrix is in general a hypergeometric function of matrix argument (see [11] for the distribution function and efficient ways to compute this distribution).

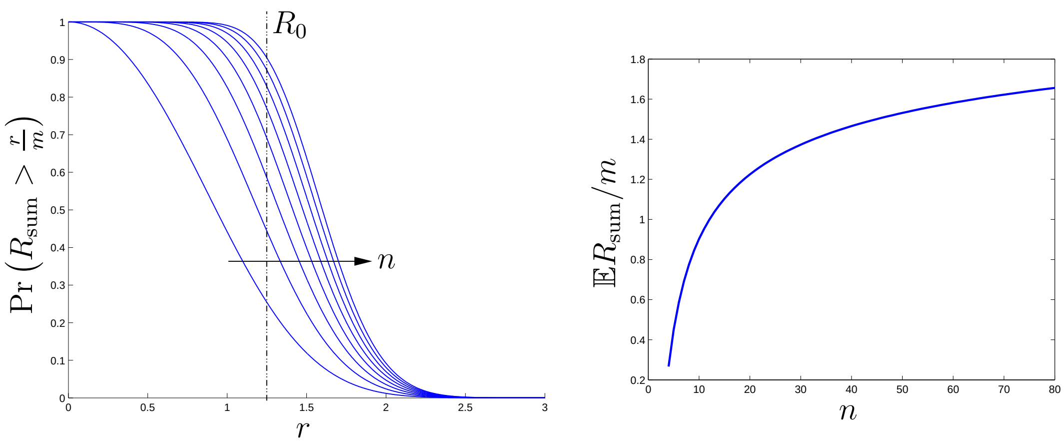

Figure 2: (a) The lower bound on the complementary distribution function of the expected sum rate and (b) the resulting bound on the expected rate

Examining (5) note that for a fixed rate this bound decays exponentially n. Thus, if we wish to schedule users to achieve some constant rate the probability of finding a set approaches one exponentially in the number of users known at the transmitter. On the other hand, if we are in a system in which n is fixed, the probability of finding a set with sum rate r varies extremely rapidly in r. Due to this triple exponential the complementary distribution function transitions rapidly from 0 to 1 over a very narrow rate interval. For this reason we may think of this triple exponential essentially as the step function at some rate R_0, *i.e.* $1 - u(r - R_0)$. Considering the location of the transition R_0 we can, using the approximation (5), see that R_0 is of order $\log\log n$. This behavior can be seen in Figure 2 (a).

Before proceeding to the proof of Theorem 3.1, we make the following remarks.

Remarks:

1. Note that if r increases slower than $\log\log n$ in (5) the right hand side is bounded away from zero. Thus, for each scheduling interval we can, with non-zero probability, find a set whose sum rate is of order $\log\log n$ and thus the value of the expected sum rate is of order $\log\log n$.

2. Note that the expected sum rate is essentially R_0 due to the fact that the expected value is simply the integral over the complementary distribution function (for the distribution $1 - u(r - R_0)$ the expected value is R_0). The expected sum rate can be seen in Figure 2 (b). Notice the slow increase of $\log\log n$.

3. By assuming all users are orthogonal, we have that the sum rate under any multiplexing scheme is $\Theta(\log\log n)$. Further, using low-complexity schemes that search among a small subset of users one can achieve a expected sum rate of order $\log\log n$ as observed in [8]. Thus, it appears that this scaling behavior will be achieved with many multiplexing schemes.

4. Since every user in the selected set is allocated equal rate, the complementary distribution function of rate, $\Pr(R^l_{\text{sum}} \geq r)$, may be interpreted as the probability that there exists a set such that every user gets a rate at least r/l. Incidentally, $\Pr(R^l_{\text{sum}} \leq r)$ can be interpreted as the *outage probability*. That is, the probability that there exists no set such that each user can be allocated a rate at least r/l.

We now turn to our proof of Theorem 3.1.

Proof of Theorem 3.1. Define the indicator random variable $1_{\mathcal{A}}(r)$ to be if 1 if $\hat{f}_{\text{zf}}(\mathcal{A}) > r$ for the set $\mathcal{A}$ and zero otherwise. Further, let $X_{\text{zf}}^{l}(r)$ count the number of sets of size l that obtain a rate above r. More explicitly, let

$$1_{\mathcal{A}} \triangleq 1_{\mathcal{A}}(r) = \begin{cases} 1 & \text{if} \quad f_{\text{zf}}(\mathcal{A}) > r \\ 0 & \text{otherwise} \end{cases} \qquad \text{and} \qquad X_{\text{zf}}^{l}(r,n) = \sum_{\substack{\mathcal{A} \subset \mathcal{U}_n \\ |\mathcal{A}|=l}} 1_{\mathcal{A}}(r)$$

Thus, $\Pr(R_{\text{sum}}^{l} \leq r)$ is equal to the probability there does not exist a set that can achieve a rate less than r. That is, the probability of the event $X_{\text{zf}}^{l}(r) = 0$. Standard approaches at bounding the above sum will not be useful since the indicator random variables in the summation are dependent. We will use the following Lemma that may be obtained by specializing the bounds of [13, Thrm. 2.1].

Lemma 3.2. *Let $\mathcal{P}_l(\mathcal{U})$ be the collection of all unordered sets of size l on n items and let*

$$X = \sum_{\mathcal{A} \in \mathcal{P}_l(\mathcal{U})} 1_{\mathcal{A}} \tag{6}$$

where $\{1_{\mathcal{A}}\}$ is a family of Bernoulli random variables with $\Pr(1_{\mathcal{A}} = 1) = p$, which are independent if $\mathcal{A} \cap \mathcal{B} = \emptyset$. Then,

$$\Pr(X = 0) \;\; \leq \;\; \exp\left(-\left\lfloor \frac{n}{l} \right\rfloor \max\left\{\frac{p}{2(1-p)}, \log\left(1 + \frac{p}{1-p}\right)\right\}\right) \tag{7}$$

In order to use this bound we need good estimates on the probability that $1_{\mathcal{A}} = 1$. We have by (2) the probability that $1_{\mathcal{A}} = 1$ as,

$$\Pr(1_{\mathcal{A}} = 1) = \Pr(\hat{f}_{\text{zf}}(\mathcal{A}) > r) \geq \Pr\left(\lambda_{\min}(\mathbf{W}_{\mathcal{A}}^{-1}) > \frac{|\mathcal{A}|}{P}\left(\exp\left(\frac{r}{|\mathcal{A}|}\right) - 1\right)\right) \tag{8}$$

Thus, combining (8) with Lemma 3.2 and specializing to the case that $l = m$, we have the desired result.

□

4 Sum rate under limited CSI

In this section, we discuss the implications of Theorem 3.1 on the amount of CSI needed at the transmitter to achieve a close to optimal scaling in the expected rate. We have shown that for the MIMO broadcast channel with multiuser scheduling the distribution of the rate has a rapid transition from 1 to 0 at a rate R_0 that is of order $\log\log n$. Thus, we incur a small penalty in expected rate if the transmitter is only informed of a small subset of users. In this section we will examine which users should report back their channel vectors to the transmitter in order for the system to approach the maximal sum rate under full CSI.

First, consider the effects of randomly selecting a subset of constant size, say N_{CSI}, at each scheduling interval. It should be clear that using this approach we will need N_{CSI} to be a fixed fraction of the total user population to obtain a rate close to maximal. More precisely, we will need the size of the randomly selected set to scale on the order of n to achieve a $\log\log n$ scaling in rate. In order to obtain the same rate, with a smaller N_{CSI} we may only consider the set of N_{CSI} users with the largest norm. To do this in a distributed fashion we may have users report their channels if the norm is greater that some threshold, say ρ. This will typically yield a smaller, randomly sized, set of users. Further, using a threshold that is $O(\log n)$ allows us to achieve the $\log\log n$ scaling in expected rate while reducing the expected number of users known at the transmitter to a number independent of the total number of users n, *i.e.* $O(1)$.

To see that selection of users for feedback based on a norm threshold has these properties consider the scheduler at the transmitter that only selects a subset of users that meet the norm constraint if the inner product between every pair of users is less than some constant ϵ [8]. Then, we can with non-zero probability find a set of users that meets this inner product constraint if $\epsilon/\rho > 0$. Further, using a simple bound on the minimum eigenvalue we may bound the zero forcing rate expression as

$$\hat{f}_{\text{zf}}(\mathcal{A}) \geq \log\left(1 + \frac{P\rho}{m}\frac{\left(1 - m\frac{\epsilon}{\rho}\right)^m}{\left(1 + (m-1)\frac{\epsilon}{\rho}\right)^{m-1}}\right) \tag{9}$$

Thus, for $\rho(n) = \log(n)$ we achieve a $\log\log n$ scaling while having the expected number of users that feedback their channels independent of n. The choice of ρ only effects the average number of users and the constant in front of $\log\log n$.

Note, from the above discussion and (9) it is natural to consider the effects of having users only feedback their channels if they meet both the norm constraint and the inner product constraint ϵ. Although this cannot be implemented in a decentralized manner, we may assume that a genie selects such a set or collection of sets. However, since this is no better in terms of the scaling law or in the number of users reporting their channel (outside of a constant scaling independent of n) we do not need such a genie.

A specific form of the proposed protocol for achieving the optimal scaling in rate while having the expected number of users that report back independent of n is as follows. The transmitter and receivers agree on ρ based on channel statistics and the total number of users n. We also assume the availability of a control channel over which the transmitter can send simple protocol directives to the receivers.

Protocol:

1. If n is small, set $\rho = 0$.
2. Otherwise, the transmitter picks ρ and sends it over the control channel to the receivers
3. if $\|\mathbf{h}_i\| > \rho$, receiver i sends its channel vector to the transmitter
4. the transmitter selects a set of users, $\mathcal{A}$, among those that reported, and transmits data to those users

Note that, with the above protocol, when $\rho > 0$, the number of users that report has a binomial distribution, which will exhibit a threshold in the number of users n for any given ρ. Also note that in step 3 of the protocol, the receivers can in practice quantize the vectors $\mathbf{h}_i$. Analyzing the tradeoff of sum rate with the number of quantization bits is the subject of our ongoing work.

5 Appendix

We will focus on linear multiplexing. That is where the instantaneous signal $\mathbf{x}$, can be represented as the linear combination

$$\mathbf{x} = \sum_{i \in \mathcal{A}} u_i \mathbf{w}_i = \mathbf{W}\mathbf{u}$$

where u_i is the message symbol for users i and $\mathbf{w}_i$ is the *beamforming* vector. This vector in general may be optimized for each transmission but may also come from some finite code book, say $\mathcal{C} = \{\mathbf{w}_1, \mathbf{w}_2, \ldots, \mathbf{w}_{2^k}\}$. If the receiver employs an MMSE receiver to maximize the receive SINR the resulting SINR for the ith user is [14]

$$\mathrm{SINR}_i(\mathcal{A}) = \frac{P_i |\mathbf{h}_i \mathbf{w}_i^\dagger|^2}{1 + \sum_{i \neq j} P_j |\mathbf{h}_i \mathbf{w}_j^\dagger|^2} \tag{10}$$

where P_i is the power allocated to user i.

It can be shown that in the case of maximizing $\mathbf{q} \cdot \mathbf{r}$, finding $\zeta_{\mathrm{sp}} = \zeta_{\mathrm{sp}}(\mathcal{A})$ that satisfies, for a fixed choice of $\mathcal{A}$,

$$\sum_{i \in \mathcal{A}} \left(q_i \zeta_{\mathrm{sp}} - \frac{1}{\mathrm{SINR}_i(\mathcal{A})} \right)_+ = P \tag{11}$$

where $(x)_+ = \max\{0, x\}$ yields the optimal power allocation. We will let $\mathcal{C}_{\mathrm{sp}}(\mathbf{H}, P)$ be the set of all rate vectors achievable by beamforming under a power constraint P and let $\mathcal{S}_{\mathrm{sp}}$ be the collection of user sets such that every user receives strictly positive power. That is,

$$\mathcal{S}_{\mathrm{sp}} \triangleq \left\{ \mathcal{A} \;\middle|\; \frac{1}{\mathrm{SINR}_{i_{\min}}(\mathcal{A})} \leq q_{i_{\min}} \zeta_{\mathrm{sp}}(\mathcal{A}) \right\} \tag{12}$$

The collection $\mathcal{S}_{\mathrm{sp}}$ is sufficiently large so that we can still find a set that achieves the maximal sum rate. Additionally, if $\mathcal{A} \in \mathcal{S}_{\mathrm{sp}}$, then the positivity constraint in the water filling equation (11) is always satisfied. This yields the following characterization of the weighted sum rate.

Proposition 5.1 ([9], Thrm. 2.2.1)**.** *Under a total power constraint P, the maximum weighted beamforming sum rate is*

$$\max_{\substack{\mathcal{A} \subset \mathcal{U} \\ \mathbf{r} \in \mathcal{C}_{\mathrm{sp}}(\mathbf{H}, \mathbf{q})}} \mathbf{q} \cdot \mathbf{r} = \max_{\mathcal{A} \in \mathcal{S}_{\mathrm{sp}}} \left(\sum_{i \in \mathcal{A}} q_i \right) f_{\mathrm{sp}}(\mathbf{H}_{\mathcal{A}}, \mathbf{q}(\mathcal{A}))$$

where

$$f_{\mathrm{sp}}(\mathbf{H}_{\mathcal{A}}, \mathbf{q}(\mathcal{A})) = \log \left(1 + \frac{P}{|\mathcal{A}|} \mathcal{H}(\{\mathrm{SINR}_i(\mathcal{A})\}) \right) + D(\, q(\mathcal{A}) \parallel \mathrm{SINR}(\mathcal{A})) \tag{13}$$

where $\mathcal{H}$ is the harmonic mean, $D(\cdot \| \cdot)$ is the Kullback Leibler distance and $\mathrm{SINR}(\mathcal{A})$ *and $q(\mathcal{A})$ are the empirical distributions of the* $\mathrm{SINR}_i(\mathcal{A})$ *and q_i restricted to the set $\mathcal{A}$.*

We now examine a special case of beamforming, zero-forcing multiplexing, in which we take the beamforming matrix to be the inverse of the channel matrix. That is, it simply inverts the channel at the transmitter by choosing a transmit vector $\mathbf{x} = \mathbf{H}_{\mathcal{A}}^{+}\mathbf{u}$, where $\mathbf{H}_{\mathcal{A}}^{+}$ is the pseudo-inverse of the channel matrix for the active user set $\mathcal{A}$ and $\mathbf{u}$ is the vector of message symbols. We assume throughout that $\mathbf{H}_{\mathcal{A}}$ is non-singular since this occurs with probability 1. In this special case we have $\text{SINR}_i(\mathcal{A}) = b_i(\mathcal{A})$ where

$$b_i(\mathcal{A}) = \frac{1}{\left(\mathbf{W}_{\mathcal{A}}^{-1}\right)_{i,i}} \quad \text{and} \quad \mathbf{W}_{\mathcal{A}} = \mathbf{H}_{\mathcal{A}}\mathbf{H}_{\mathcal{A}}^{\dagger} \tag{14}$$

Note that the sub-optimality of zero-forcing appears in the power price paid in inverting the channel. It can be shown that the power constraint becomes, $\sum_i P_i/b_i \leq P$. The b_i's have an important geometric interpretation as noted in [15] as the squared norm of user i's channel when projected away from every other users channel in the activation set $\mathcal{A}$. This suggests that we pay a large price in power if we have users who are nearly collinear. In this special case we have the following characterization of the sum rate.

Proposition 5.2 ([9], Thrm. 2.2.2)**.** *Under a total power constraint P, the maximum weighted zero-forcing sum rate is*

$$\max_{\substack{\mathcal{A}\subset\mathcal{U} \\ \mathbf{r}\in\mathcal{C}_{\text{zf}}(\mathbf{H},\mathbf{q})}} \mathbf{q}\cdot\mathbf{r} = \max_{\mathcal{A}\in\mathcal{S}_{\text{zf}}} \left(\sum_{i\in\mathcal{A}} q_i\right) f_{\text{zf}}(\mathbf{H}_{\mathcal{A}}, \mathbf{q}(\mathcal{A}))$$

where

$$f_{\text{zf}}(\mathbf{H}_{\mathcal{A}}, \mathbf{q}(\mathcal{A})) = \log\left(1 + \frac{P}{|\mathcal{A}|}\mathcal{H}\left(\{b_i(\mathcal{A})\}\right)\right) + D(q(\mathcal{A})\|b_{\mathcal{A}}) \tag{15}$$

where $\mathcal{H}$ is the harmonic mean, $D(\cdot\|\cdot)$ is the Kullback Leibler distance and $b_{\mathcal{A}}$ and $q(\mathcal{A})$ are the empirical distributions of the b_i and q_i restricted to the set $\mathcal{A}$.

Intuitively we know that difference between water filling and flat power allocation is small at high SNR. Since as we examine more and more users we expect the channel norms of the users to increase, and hence the SNR, it is reasonable to think that a constant power allocation will perform well in the multi-antenna downlink. Further, it is more analytically convenient if we lower bound $D(q(\mathcal{A})\|b_{\mathcal{A}})$ by zero. Close examination shows that this is equivalent to allocating each user of $\mathcal{A}$ a power of $P_i = P/\text{Tr}\left(\mathbf{W}_{\mathcal{A}}^{-1}\right)$. Since this allocation satisfies the power constraint, we instead consider scheduling users that maximize the zero forcing rate under this constant power allocation. That is, a scheduler that chooses the set $\mathcal{A}$ such that

$$\mathcal{A} = \arg\max_{\mathcal{A}\subset\mathcal{U}} |\mathcal{A}|\ \hat{f}_{\text{zf}}(\mathcal{A}) \quad \text{where} \quad \hat{f}_{\text{zf}}(\mathcal{A}) = \log\left(1 + \frac{P}{\text{Tr}(W_{\mathcal{A}}^{-1})}\right)$$

References

[1] H. Vishwanathan and K. Kumaran, "Rate scheduling in multiple antenna downlink wireless systems," in *Proc. 39th Annual Allerton Conf. on Communication, Control, and Computing*, Monticello, Illinois, 2001.

[2] M. Lopez, "Multiplexing, scheduling and multicasting strategies for antenna arrays in wireless networks," Ph.D. dissertation, MIT, Cambridge, Massachusetts, 2002.

[3] C. Swannack, E. Uysal-Biyikoglu, and G. W. Wornell, "Low complexity multiuser scheduling for maximizing throughput in the MIMO broadcast channel," in *Proc. 42nd Annual Allerton Conf. on Communication, Control, and Computing*, Monticello, Illinois, 2004.

[4] B. Hochwald, T. L. Marzetta, and V. Tarokh, "Multi-antenna channel-hardening and its implications for rate feedback and scheduling," *IEEE Transactions on Information Theory*, 2004, to appear.

[5] C. Swannack, E. Uysal-Biyikoglu, and G. W. Wornell, "Phase transition phenomena in MIMO broadcast scheduler-multiplexer design," *IEEE Transactions on Information Theory*, 2005, submitted (March 2005).

[6] T. Yoo and A. J. Goldsmith, "Optimality of zero-forcing beamforming with multiuser diversity," in *Proc. IEEE International Conf. on Communications(ICC)*, May 2005.

[7] ——, "Optimality of zero-forcing beamforming with multiuser diversity," *IEEE Journal on Selected Areas in Communications*, to appear.

[8] C. Swannack, E. Uysal-Biyikoglu, and G. W. Wornell, "Finding NEMO: Near mutually orthogonal sets and applications to MIMO broadcast scheduling," in *Proc. IEEE WIRELESSCOM 2005 : International Conference on Wireless Networks, Communications, and Mobile Computing*, Maui, Hawaii, USA, June 2005.

[9] C. Swannack, "Throughput and complexity tradeoffs in the multi-antenna downlink," Master's thesis, MIT, Cambridge, Massachusetts, June 2005.

[10] T. Yoo and A. J. Goldsmith, "Sum-rate optimal multi-antenna downlink beamforming strategy based on clique search," in *Proc. IEEE Globecom 2005*, November 2005, to appear.

[11] P. Koev and A. Edelman, "The efficient evaluation of the hypergeometric function of a matrix argument," *Mathematics of Computation*, to appear.

[12] A. Edelman, "Eigenvalues and condition numbers of random matrices," Ph.D. dissertation, MIT, Cambridge, Massachusetts, 1989.

[13] S. Janson, "Large deviations for sums of partly dependent random variables," *Random Structures Algorithms*, vol. 24, no. 3, pp. 234–248, 2004.

[14] E. Visotsky and U. Madhow, "Optimum beamforming using transmit antenna arrays," in *Proc. 1999 IEEE 49th Vehicular Technology Conference*, vol. 1, May 1999, pp. 851–856.

[15] G. Caire and S. Shamai (Shitz), "On the achievable throughput of a multiantenna Gaussian broadcast channel," *IEEE Transactions on Information Theory*, vol. 49, no. 7, pp. 1691–1706, July 2003.