Efficient Quantization for Feedback in MIMO Broadcasting Systems

Charles Swannack, Gregory W. Wornell
Dept. EECS, MIT
Cambridge MA 02139
Email: {swannack,gww}@mit.edu

Elif Uysal-Biyikoglu
Dept. ECE, Ohio State University
Columbus, OH 43210
Email: elif@ece.osu.edu

***Abstract*—**
We consider the problem of joint multiplexer-scheduler design for transmitting independent data streams over a Gaussian multiple-antenna broadcast channel in which feedback is used to convey channel state information from receivers to the transmitter. It is known that various low complexity strategies can achieve the optimal rate scaling as a function of receiver population size. In this work we develop a simple and efficient quantization strategy for use on the feedback link of such architectures.

I. Introduction and Background

There is growing interest in the development of efficient wireless broadcast systems for distributing independent data streams to different users over some geographical area. It is now widely appreciated that the use of a multiple-element antenna array at the transmitter can, in principle, greatly increase the capacity of such systems. When the number of users is no larger than the array size, the system design issues are rather well-understood. Moreover, when it is desirable for complexity or other reasons to restrict one's attention to case of linear multiplexing, the literature characterizing the associated performance tradeoffs is particularly extensive.

Recent approaches to this scheduling problem have examined the scaling behavior of the multiple-antenna broadcast channel in the large user limit with perfect channel channel state information [1]–[4] using various interference cancelling multiplexers and complexity constraints [2], [3], [5]. Here, we provide a simple architecture for scheduling over the Gaussian MIMO broadcast channel with quantized feedback. We have shown [6] the achieved rate of this architecture asymptotically equals that of the best multiplexer and scheduler. This was done by showing that there exists a group of users equal of size equal to the transmit dimension in which the mutual interference is negligible. In this paper we present a simple quantization scheme and show that good performance can be achieved when the number of users is only a small multiple of the user population. The single user version of the problem was discussed in [7]. We examine the effects of choosing groups of various sizes for a large user pool.

II. Channel and System Model

The system of interest consists of an m-element transmitter antenna array and a pool of n destinations (users). The transmitter has n collections of messages, each such collection destined for one of the n users. The collections are infinite in size, corresponding to an infinite backlog.

This work was supported in part by NSF under Grant No. CNS-0434974, Mitre Corporation, and by HP through the MIT/HP Alliance.

Our discrete-time channel model is a narrowband block fading one. Specifically, in any particular block, the signal $y_j(k)$ received by user j at time k in response to a signal $\mathbf{x}(k)$ transmitted from the array is of the form

$$y_j(k) = \mathbf{h}_j^\dagger \mathbf{x}(k) + z_j(k) \tag{1}$$

where $z_j(k)$ is independent identically distributed (i.i.d.) $\mathcal{CN}(0,1)$ noise, and where the (normalized) channel gain vectors $\mathbf{h}_j$ have i.i.d. $\mathcal{CN}(0,1/2m)$ elements. The noises and channel gains are independent from receiver to receiver, and from block to block.

Any message scheduled for delivery is transmitted within one block, and the blocks are long so that messages can be reliably received. Thus each block corresponds to a new signaling (and hence scheduling) interval. Within each signaling interval, the transmitter sends from its array a group of messages, one for each of a subset of the user pool. The transmitter is subject to an average total power constraint of P, i.e., $E\left[\|\mathbf{x}\|^2\right] \le P$ within each signaling interval. We will let $R(\mathbf{H}_\mathcal{A})$ be the achievable rate for the user set $\mathcal{A}$ under this power constraint.

In our model, channel gains in each signaling interval are known perfectly (i.e., measured to arbitrary accuracy) at the respective receivers at the beginning of each such interval. Moreover, a feedback link exists by which individual users can inform the transmitter of their channel gains (or more generally quantized versions thereof), also at the beginning of each associated signaling interval. The users do not know each other's channel gains, nor are they able to more generally share information between each other.

Finally, the performance criterion of interest in this work is average throughput (i.e., expected sum-rate), and our focus is on the small n regime (with m fixed).

III. System and Protocol Architecture

The architecture of interest is as illustrated in Fig. 1. The protocol is identical in each signaling interval, so we restrict our attention to a single arbitrary one. In such an interval, a subset $\mathcal{R}$ of users from the full population $\mathcal{U}$ sends a quantized representation of their respective channel gain vectors to the transmitter over the feedback link. The associated quantization codebook $\mathcal{C}$ is fixed and the same for all users. Its structure is such that the codewords $\mathbf{c} \in \mathcal{C}$ all lie on the unit sphere in m (complex) dimensions, and the quantization rule corresponds to

$$\hat{\mathbf{h}}_j = \arg\max_{\mathbf{c}\in\mathcal{C}} \left|\mathbf{c}^\dagger \mathbf{h}_j\right|, \tag{2}$$

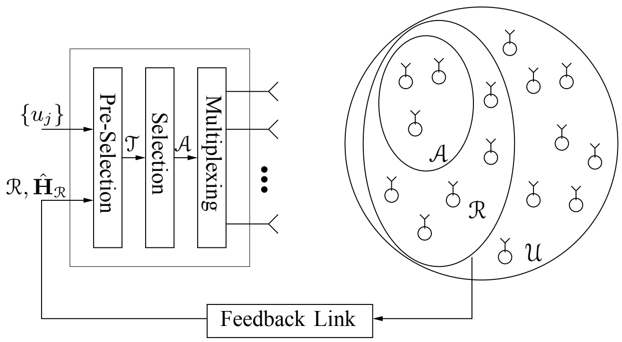

Fig. 1. MIMO system architecture. In each scheduling interval a subset $\mathcal{R}$ of the full user pool $\mathcal{U}$ of size n reports quantizations $\hat{\mathbf{H}}_{\mathcal{R}}$ of its channel gains to the transmitter via the feedback link using a decentralized (individual) criterion. From the set $\mathcal{R}$, the transmitter first forms a collection $\mathcal{T}$ of candidate user sets of size m using a pairwise criterion; this is the pre-selection phase. Next, a set $\mathcal{A} \in \mathcal{T}$ is chosen at random as the active set, whose messages $\{u_j, j \in \mathcal{A}\}$ are linearly multiplexed across the array for transmission.

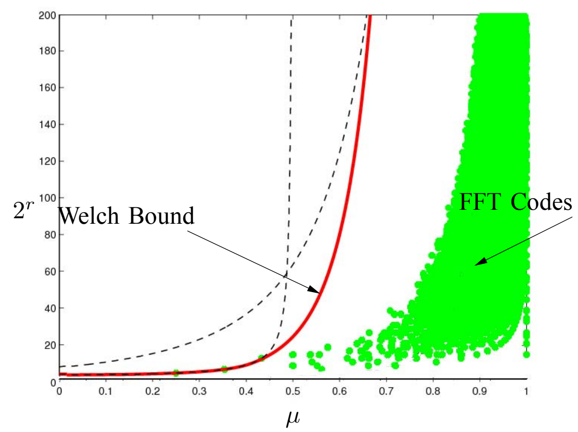

Fig. 2. A plot Welch bound and the coherence achieved by some FFT codes of size 4 to 200 in dimension 4. The Welch bound for $k = 1$ and large k are plot as dashed lines. The union over all k is plotted as the solid line.

where $\hat{\mathbf{h}}_j$ denotes the quantization of $\mathbf{h}_j$. We let r denote the number of bits to which a channel gain is quantized, so the the codebook is of size 2^r. We label the codewords in the codebook $\mathcal{C} = \mathcal{C}_r$ as $\mathbf{c}_1, \mathbf{c}_2, \ldots, \mathbf{c}_{2^r}$.

With this notation a key figure of merit for the codebook is its *coherence*

$$\mu(\mathcal{C}) = \max_{i \neq j} \left| \mathbf{c}_i^\dagger \mathbf{c}_j \right|. \quad (3)$$

Another key figure of merit for a codebook in a multi-user system is the maximum root-mean-square (RMS) inner product magnitude

$$\mu_{\text{rms}}(\mathcal{C}) = \max_j \sqrt{\sum_{i \neq j} \left| \mathbf{c}_i^\dagger \mathbf{c}_j \right|^2}. \quad (4)$$

In general, $0 \leq \mu \leq 1$, and, for a given r, smaller values of μ and μ_{rms} correspond to better codes. It is a natural to consider how large a code can be for a given coherence. In this direction we have the following bound from [8], [9].

Proposition 1: Let $\mathcal{C}_N \subset \mathbb{C}^m$ be an arbitrary code of size N. If $\mu = \mu(\mathcal{C}_N)$, then

$$N \leq \frac{1 - \mu^{2k}}{\binom{k+m-1}{k}^{-1} - \mu^{2k}}$$

for all k such that $\binom{k+m-1}{k}^{-1/(2k)} > \mu$.
This bound can be seen for the case of 4 transmit antennas in Figure 2.

At each scheduling interval the subset $\mathcal{R}$ is determined in a decentralized manner, i.e., based on an individual evaluation of each channel gain vector or a predetermined schedule. The particular criterion we consider corresponds to each user j computing the squared norm $\|\mathbf{h}_j\|^2$ of its channel gain vector, and the correlation $|\mathbf{h}_j^\dagger \hat{\mathbf{h}}_j|$ between the channel gain vector and its quantization $\hat{\mathbf{h}}_j$. If these factors fall within certain prescribed ranges, a user will convey its channel gain to the transmitter. More formally, the particular criterion we consider corresponds to

$$\mathcal{R}_{\rho,\sigma} \triangleq \{ j \in \mathcal{U} : \|\mathbf{h}_j^2\| \geq \rho \text{ and } |\tilde{\mathbf{h}}_j \hat{\mathbf{h}}_j^\dagger| \geq \sigma \}, \quad (5)$$

where $\tilde{\mathbf{h}}_j = \mathbf{h}_j / \|\mathbf{h}_j\|$, and where ρ and σ are prescribed parameters of the protocol. Furthermore, it suffices to restrict our attention to $\sigma \geq \mu(\mathcal{C})$.

At the transmitter, there are three relevant stages of processing. First, from the set $\mathcal{R}$ of reporting users, a collection $\mathcal{T}$ of candidate subsets of size l is formed; this is the pre-selection phase. Next, one of these subsets, denoted $\mathcal{A}$, is selected from $\mathcal{T}$ at random, and corresponds to the active user set for the signaling interval. Finally, one message for each of the active users is selected, and the resulting group of messages is multiplexed across the array for transmission.

The pre-selection phase is based on simple pairwise evaluation of the vectors in $\mathcal{R}$. The particular criterion we consider corresponds to

$$\mathcal{T}_{\epsilon,\rho,\sigma}^{(l)} \triangleq \{ \mathcal{A} \subset \mathcal{R}_{\rho,\sigma} : |\mathcal{A}| = l \text{ and } |\hat{\mathbf{h}}_i^\dagger \hat{\mathbf{h}}_j| \leq \epsilon, \ \forall\, i \neq j \in \mathcal{A} \},$$

where ϵ and l are pre-selection parameter chosen based on knowledge of $|\mathcal{R}_{\rho,\sigma}|$. In general, the parameters ρ, σ, and ϵ — and hence the sets $\mathcal{R}$, $\mathcal{T}$, and $\mathcal{A}$ — will all be functions of n. Our notation will only show this dependency explicitly when necessary.

IV. Quantization

We consider codebooks constructed from fast Fourier transform matrices[1]. This class of codes can be thought of as a subset of m rows of the $2^r \times 2^r$ FFT matrix [10]. More precisely, we let the FFT code with index set $\mathbf{g} = [g_1\, g_2\, \ldots\, g_m]$, denoted $\mathcal{C}_{\text{fft}}(\mathbf{g}, 2^r)$, be the codebook of size 2^r with codewords taken to be the columns of

$$\frac{1}{\sqrt{m}} \begin{bmatrix} 1 & 1 & \cdots & 1 \\ e^{j\frac{\pi}{2^{r-1}} g_1} & e^{j\frac{\pi}{2^{r-1}} g_2} & \cdots & e^{j\frac{\pi}{2^{r-1}} g_m} \\ e^{j\frac{\pi 2}{2^{r-1}} g_1} & e^{j\frac{\pi 2}{2^{r-1}} g_2} & \cdots & e^{j\frac{\pi 2}{2^{r-1}} g_m} \\ \vdots & \vdots & \ddots & \vdots \\ e^{j\frac{\pi(2^r-1)}{2^{r-1}} g_1} & e^{j\frac{\pi(2^r-1)}{2^{r-1}} g_2} & \cdots & e^{j\frac{\pi(2^r-1)}{2^{r-1}} g_m} \end{bmatrix}. \quad (6)$$

[1] We note that this theory is much deeper than considered here and that any FFT code can be considered as the image of a linear code. However, we will not mention this connection except when helpful in simplifying proofs.

r	$\mathbf{g}_4$
2	$\{1,2,3,4\}$
3	$\{1,2,3,5\}$
4	$\{1,3,4,8\}$
5	$\{1,2,6,13\}$
6	$\{1,2,50,54\}$

Fig. 3. The index sets for the FFT codes with minimal coherence in 4 dimensions for $2,3,4,5$ and 6 bit quantization.

These code books are known to achieve the smallest μ for a given r in very special cases. The index sets that minimize the coherence for $2,3,4,5$ and 6 bit quantization can be seen in Figure 3.

It is important to note that using this construction the RMS inner product magnitude is the same for all codewords. That is,

$$\mu_{\rm rms}(\mathcal{C}_{\rm fft}) = \max_j \sqrt{\sum_{i\neq j}\left|\mathbf{c}_i^\dagger \mathbf{c}_j\right|^2} = \sqrt{\sum_{i\neq k}\left|\mathbf{c}_i^\dagger \mathbf{c}_k\right|^2} \quad \forall \mathbf{c}_k \in \mathcal{C}_{\rm fft}$$

Moreover, the Voronoi regions of every FFT code are isomorphic so that under the assumption of uncorrelated Gaussian channel vectors each users channel vector is equally likely to be quantized to any code index. Lastly, it is clear from (6) that the correlation between every two users is only a function of the magnitude of the d difference of the indices. That is,

$$|\mathbf{c}_i^\dagger \mathbf{c}_j| = g(|i-j|)$$

for some function g. We will argue that it is these properties of FFT based codebooks that make this class of quantizers valuable when considering the scheduling of users with or without a complexity constraint. We address this after first discussing the multiplexing phase of our protocol.

V. Multiplexing

For the multiplexing phase of the protocol, we restrict our attention to linear multiplexers. Specifically, with $\mathbf{u}$ denoting the vector of m coded symbols $u_j, j \in \mathcal{A}$ for the m active users, the transmitted signal takes the form

$$\mathbf{x} = \sum_{j\in\mathcal{A}} u_j \mathbf{w}_j = \mathbf{W}_\mathcal{A}\mathbf{u} \tag{7}$$

where the $\mathbf{W}$ is a matrix whose columns are the unit-norm weight (i.e., beamforming) vectors $\mathbf{w}_j, j \in \mathcal{A}$. We further restrict our attention to uniform power allocation in the multiplexing, i.e., $E\left[|u_j|^2\right] = P/m$ for all $j \in \mathcal{A}$.

Among linear multiplexers, of primary interest for discussion will be interference-ignoring (II) multiplexers. The weight matrix in this case takes the form

$$\mathbf{W}_\mathcal{A}^{\rm II} = \hat{\mathbf{H}}_\mathcal{A}$$

where the columns of $\hat{\mathbf{H}}_\mathcal{A}$ are $\hat{\mathbf{h}}_j$, $j \in \mathcal{A}$. At the other end of the spectrum are interference-canceling (IC) multiplexers, for which

$$\mathbf{W}_\mathcal{A}^{\rm IC} = \hat{\mathbf{H}}_\mathcal{A}(\hat{\mathbf{H}}_\mathcal{A}^\dagger \hat{\mathbf{H}}_\mathcal{A})^{-1}. \tag{8}$$

We will only consider the case of interference-ignoring multiplexers. It is simple, using the relevant expressions in [6], to extend the following discussion to the class of interference-cancelling multiplexers. In the case of interference-ignoring multiplexers, for a given active set $\mathcal{A}$ and channel realization $\mathbf{H}_\mathcal{A}$, it is straightforward to verify that the achievable sum rate satisfies

$$R_{\mathbf{H}_\mathcal{A}}^{\rm II}(n) = \sum_{j\in\mathcal{A}} \log(1 + {\rm SINR}_j^{\rm II}) \tag{9}$$

where

$${\rm SINR}_j^{\rm II} = \frac{P\|\mathbf{h}_j\|^2|\sigma_j|^2}{m + P\|\mathbf{h}_j\|^2\|\boldsymbol{\sigma}_j^{\rm c}\|^2} \tag{10}$$

with $\sigma_j = \hat{\mathbf{h}}_j^\dagger \tilde{\mathbf{h}}_j$ and $\boldsymbol{\sigma}_j^{\rm c} = \hat{\mathbf{H}}_{\mathcal{A}\backslash j}^\dagger \tilde{\mathbf{h}}_j$.

It is clear that the SINR in (10) is dependent on the correlation between the channel vectors of the set $\mathcal{A}\backslash j$ and $\mathbf{h}_j$. Thus, the pre-selection parameters ϵ and l play an important role on the rate. We now turn to examining exactly how these parameters effect the expected sum rate.

VI. Pre-selection

Recall that using our architecture at each scheduling interval a random number of receivers report a quantized version of their channel vector. Since the transmitter is not informed of the magnitude of the receivers channel gain nor the correlation between the channel gain vector it is sufficient to only consider the quantization indices that are reported at each interval[2]. That is, for pre-selection we only need to consider the random subset of distinct quantization indices that are reported at each interval.

Pre-selection may be broken in to two steps. The first step groups the reporting users by their quantization index and the second does a pairwise evaluation for each group. We let $\mathcal{Q}$ be the random subset of quantization indices that are reported at each interval, i.e.

$$\mathcal{Q} = \left\{ i \,:\, \hat{\mathbf{h}}_j = \mathbf{c}_i \text{ for some } j \in \mathcal{R}_{\rho,\sigma} \right\}.$$

Remark: If the preselector searches for more than one user, say l many users, it is of interest to have $|\mathcal{Q}| \gg l$ so that a group of users may be found with low interference. Thus, one is interested in maximizing the expected number of distinct code indices at any scheduling interval. It is known that a uniform probability assignment to the code indices is optimal [11]. Thus, a code in which $\Pr(\hat{\mathbf{h}}_i = \mathbf{c}_j) = 1/2^r$ maximizes the expected number of distinct code indices.

Considering FFT based codes it should be clear that the probability that any user is quantized to any index, say i, is independent of i since the Voronoi regions are congruent for FFT based codes[3] and since Gaussian random vectors are isotropic. Thus, if one is interested in maximizing the expected value of $|\mathcal{Q}|$ at any interval and not the coherence, FFT based quantization code books are optimal. In this direction we have the following lemma.

Lemma 1: Let $\mathcal{C}_r$ be a codebook of size 2^r such that for any user, say user i, it is equally likely that the user's channel, $\mathbf{h}_i$, is quantized to any codeword. That is, $\Pr(\hat{\mathbf{h}}_i = \mathbf{c}_j) = 1/2^r$

[2]Note that had the users channel gains been quantized we can consider a user with the largest channel gain

[3]This is also true if we fix $\sigma \geq \mu(\mathcal{C})$.

for any $\mathbf{c}_j \in \mathcal{C}_r$. Then, the probability that at any scheduling interval there are k quantization indices given j users feed back is

$$\Pr\left(|\mathcal{Q}| = k \,|\, |\mathcal{R}_{\rho,\sigma}| = j\right) = \frac{(2^r)_k \left\{ {j \atop k} \right\}}{2^{rj}}$$

where $(x)_k = x(x-1)\cdots(x-k+1)$ is the falling factorial and where $\left\{ {j \atop k} \right\}$ is the Stirling number of the second kind[4].

Note, however, that even if $|\mathcal{Q}|$ is large at a particular scheduling interval improper choice of ϵ and l may cause $\mathcal{T}^{(l)}_{\epsilon,\rho,\sigma}$ to be empty. In particular, if the quantization codebook does not have a sub-code of size l with coherence less than ϵ then $\mathcal{T}^{(l)}_{\epsilon,\rho,\sigma}$ will always be empty. We let $k_\epsilon(\mathcal{C}_r, l)$ denote the number of codes of size l with coherence at most ϵ that can be constructed from expurgations of $\mathcal{C}_r$, i.e.,

$$k_\epsilon(\mathcal{C}_r, l) = \left|\left\{\mathcal{C}_{\log l} \in \mathcal{C}_r \,:\, \mu(\mathcal{C}_{\log l}) \le \epsilon\right\}\right|. \quad (11)$$

In general, the preselector should choose ϵ and l as to maximize the expected rate given $|\mathcal{Q}|$ and the structure of $\mathcal{C}$ (see [3] for a related discussion in the case of perfect feedback).

It is quite cumbersome in general to precisely characterize the trade off between the expected rate for a given choice of ϵ and l. We discuss how this may be done in the following but for now note that for any code we can provide simple bounds on the pre-selection success probability, $\Pr(|\mathcal{T}_{\epsilon,\rho,\sigma}| \neq 0)$. In this direction we note that for FFT codes it is clear that the probability that any subset of $\mathcal{Q}$ of size l is in $\mathcal{T}^{(l)}_{\epsilon,\rho,\sigma}$ is

$$p^{(l)}_{\epsilon|\rho,\sigma} = \frac{k_\epsilon(\mathcal{C}_r, l)}{\binom{2^r}{l}}.$$

In order to bound the probability that $|\mathcal{T}_{\epsilon,\rho,\sigma}| \neq 0$ we let

$$E(p, l) \triangleq \frac{1}{l}\max\left\{\frac{p}{2(1-p)}, \log\left(1 + \frac{p}{1-p}\right)\right\}. \quad (12)$$

Then, we have the following lemma from [6].

Lemma 2: Consider a randomly chosen finite set of quantization indices $\mathcal{Q}$. Then, the probability that $\mathcal{T}_{\epsilon,\rho,\sigma}$ is non-empty is lower bounded as:

$$\Pr(|\mathcal{T}_{\epsilon,\rho,\sigma}| \neq 0) \ge 1 - \exp\left(-l\left\lfloor\frac{|\mathcal{Q}|}{l}\right\rfloor E(p^{(l)}_{\epsilon|\rho,\sigma}, l)\right) \quad (13)$$

It is clear that the choice of l strongly effects the probability that $\mathcal{T}_{\epsilon,\rho,\sigma}$ is non-empty in Lemma 2. We now address what values we should choose for l in pre-selection in terms of the expected rate.

A. Single User Time Division and Opportunistic Scheduling

We begin by considering the most trivial version of our architecture where every user feeds back their channel vector and one user is selected randomly (or via a time division scheme), i.e. $l = 1, \rho = 0$ and $\sigma = 0$. This choice of parameters corresponds to single user beamforming with finite feed back rate. Such a scheme was presented and analyzed in [13]. In this case, by averaging over the correlation σ_1 the achieved rate is equal to

$$R_{\mathrm{mk}}(P) = E_{\sigma_1}\left[C(\sigma_1 P)\right]$$

where we let $C(x)$ be the ergodic capacity

$$C(x) = E_{\|\mathbf{h}\|^2}\left[\log\left(1 + x\|\mathbf{h}\|^2\right)\right].$$

However, it should be clear that for a sufficiently large pool of users one can do better than pure time division or random selection. Indeed, as we have mentioned it is useful to have the users self-select themselves based on their channel norms. Such a scheme incurs no additional complexity at the transmitter and, provided that a user meets the norm constraint $\|\mathbf{h}\|^2 > \rho$, the rate is strictly greater. We define the *bounded ergodic capacity* to be

$$C_\rho(P) = E\left[\log\left(1 + P\|\mathbf{h}\|^2\right) \,\middle|\, \|\mathbf{h}\|^2 > \rho\right]$$

It is clear that the expected single user rate with a norm constraint ρ and random selection then becomes

$$R_{\mathrm{su}}(\rho) = \Pr\left(\mathcal{R}_{\rho,\sigma} \neq 0\right) E_{\sigma_1}\left[C_\rho(\sigma_1 P)\right]$$

As before, when the user population becomes sufficiently large this scheme may be improved upon. Consider the case when the set $\mathcal{Q}$ has two quantization indices that correspond to code vectors with zero correlation. Then, for the interference ignoring multiplexer the expected rate is

$$2E_{\sigma_1}\left[C_\rho(\sigma_1 P/2)\right]$$

For large P this is clearly greater than $E_{\sigma_1}\left[C_\rho(\sigma_1 P)\right]$. However, when the quantization indices are not orthogonal the question is not as straight forward. We address this more general question in the following section.

B. Multiple User Scheduling

When there is sufficiently high SNR it is often desirable to transmit to more than one user. In this case the mutual interference between users becomes import. That is there is a trade off between the number of users that we select and the achieved sum rate. In order to combat mutual interference the transmitter may preselect groupings of users that have a low level of interference by choosing $\epsilon \ll 1$. However, if ϵ is chosen too small the expected sum rate will be approximately 0 due to the low probability that such a set exists.

In the case of the interference ignoring multiplexer we may lower bound (10) by

$$\mathrm{SINR}^{\mathrm{II}}_j \ge \frac{1 + \frac{P}{l}(|\sigma_j|^2 + \gamma^{\mathrm{c}}_{\max}(\epsilon, l))\|\mathbf{h}_j\|^2}{1 + \frac{P}{l}\gamma^{\mathrm{c}}_{\max}(\epsilon, l)\|\mathbf{h}_j\|^2} - 1 \quad (14)$$

$$= \frac{1 + \mathrm{SIR}^{\mathrm{N}}_j(\sigma_j, \epsilon, l)}{1 + \mathrm{SIR}^{\mathrm{D}}_j(\epsilon, l)} - 1 \quad (15)$$

where $\gamma^{\mathrm{c}}_{\max}(\epsilon, l)$ is the worst possible value of $\|\boldsymbol{\sigma}^{\mathrm{c}}_j\|^2$, i.e.

$$\gamma^{\mathrm{c}}_{\max}(\epsilon, l) \triangleq \min_{\mathcal{A}, j, \mathbf{H}\,:\,|\mathcal{T}|\neq 0,\, \mathcal{A}\in\mathcal{T},\, j\in\mathcal{A}} \|\boldsymbol{\sigma}^{\mathrm{c}}_j\|^2$$

[4]It is well known that the Stirling numbers of the second kind satisfy the recurrence relation

$$\left\{ {n \atop k} \right\} = \left\{ {n-1 \atop k-1} \right\} + k\left\{ {n-1 \atop k} \right\}$$

We refer the reader to [12] for an excellent introduction to these numbers. We note that the Stirling numbers of the second kind may similarly be defined through the relation $x^n = \sum_{k=0}^{\infty} \left\{ {n \atop k} \right\} (x)_k$.

and in turn where $\mathrm{SIR}_j^{\mathrm{N}}$ and $\mathrm{SIR}_j^{\mathrm{D}}$ are one less than the numerator and denominator of (14) respectively. Thus, any user in $\mathcal{T}_{\epsilon,\rho,\sigma}$ (if one exists) has an expected rate at least

$$R_{-}(\epsilon,l) \triangleq E_{\sigma_1}\left[C_\rho\left(\mathrm{SIR}_1^{\mathrm{N}}(\sigma_1,\epsilon,l)\right)\right] - C_\rho\left(\mathrm{SIR}_1^{\mathrm{D}}(\epsilon,l)\right).$$

Weighting $R_{-}(\epsilon,l)$ by the probability that $|\mathcal{T}_{\epsilon,\rho,\sigma}^{(l)}| \neq 0$ given $|\mathcal{Q}|$ yields the expected sum rate

$$R_{\mathrm{sum}}(\epsilon,l) \triangleq l \Pr\left(|\mathcal{T}_{\epsilon,\rho,\sigma}^{(l)}| \neq 0 \,|\, |\mathcal{Q}| = q\right) R_{-}(\epsilon,l).$$

Note that the expected rate $R_{-}(\epsilon,l)$ is a decreasing function of ϵ while the probability that $|\mathcal{T}_{\epsilon,\rho,\sigma}^{(l)}| \neq 0$ is increasing in ϵ. Asymptotically it is sufficient to fix $l = m$ and to take $\epsilon = 0$ to asymptotically achieve the optimal rate [6]. However, as previously noted, in practice for finite n it is better to choose ϵ and l based on the knowledge of $|\mathcal{Q}|$. In fact if $|\mathcal{Q}|$ is small then it is best that one take $\epsilon = 1$ and perform an optimal search over all sets. However, if $|\mathcal{Q}|$ is large this may be too complex. In this case, the preselector should choose ϵ and l as to maximize the expected rate. That is, the preselector chooses

$$(\epsilon^\star(q), l^\star(q)) = \arg\max_{\epsilon,l} R_{\mathrm{sum}}(\epsilon,l).$$

Then the resulting bound on the rate is then

$$R_{\mathrm{mu}}^{-}(\rho,\sigma,n) = \sum_{i<j} \Pr\left(|\mathcal{R}_{\rho,\sigma}| = j\right) \Pr\left(|\mathcal{Q}| = q \,|\, |\mathcal{R}_{\rho,\sigma}| = j\right) \times R_{\mathrm{sum}}(\epsilon^\star(q), l^\star(q)) \tag{16}$$

Note, however, that this bound uses the worst case value for the rate. It is natural to ask if this bound in practice is a fair characterization of the achievable rate and if not whether there is a simple way to improve on the achievable rate.

VII. Improving Performance with Greedy Search

It should be clear that the lower bound on the achievable rate (16) is close to the achieved rate if for a given ϵ a large fraction of the possible subsets of $|\mathcal{T}_{\epsilon,\rho,\sigma}^{(l)}|$ achieve approximately the same rate. That is, a large fraction of subsets achieve a rate approximately equal to $R_{\mathrm{sum}}(\epsilon^\star(q), l^\star(q))$. Note, that (16) assumes that $|\mathcal{Q}|$ is large so that pre-selection is needed. Otherwise, if $|\mathcal{Q}|$ is small one may select a set of users for transmission by computing the achieved rate for all subsets of $\mathcal{Q}$ and choosing the maximum. Thus, it is reasonable to assume that $|\mathcal{Q}|$ is large and hence $|\mathcal{C}|$ is large so that $|\sigma_j| \approx 1$ or alternatively that $\tilde{\mathbf{h}}_j \approx \hat{\mathbf{h}}_j$ for all j. Thus,

$$\sigma_j \approx \sum_{\substack{i\in\mathcal{A}\\ i\neq j}} \hat{\mathbf{h}}_i^\dagger \hat{\mathbf{h}}_j = \sum_{\substack{q_i\in Q(\mathcal{A})\\ q_i\neq Q(j)}} g(|q_i - q_j|) \triangleq s_j(\mathcal{A})$$

where we let $Q(\mathcal{A})$ be the set of quantization indices for the users in the set $\mathcal{A}$. Thus, the sum rate achieved by any set $\mathcal{A}$ is approximately equal to

$$R_{\mathrm{mu}}^a(\mathcal{A}) \triangleq \sum_{j\in\mathcal{A}} \log\left(1 + \frac{P\|\mathbf{h}_j\|^2|\sigma_j|^2}{|\mathcal{A}| + P\|\mathbf{h}_j\|^2 s_j(\mathcal{A})}\right)$$

Note, that this expression is only a function of the set of differences $|q_i - q_j|$ and it is clear that

$$R_{\mathrm{mu}}^a(\mathcal{A}+j) = R_{\mathrm{mu}}^a(\mathcal{A})$$

for all $j \in \mathbb{Z}$. Thus, when computing the rate achieved by a set $\mathcal{A}$ we may first translate $\mathcal{A}$ by its minimal element and then compute the rate. In this direction we let the set $\Delta\mathcal{A} = \mathcal{A} - \min_{j\in\mathcal{A}} j$. Then, we have that $R_{\mathrm{mu}}^a(\Delta\mathcal{A}) = R_{\mathrm{mu}}^a(\mathcal{A})$. It is important to note that the cardinality of the collection of translated user sets is a fraction of the collection of all user sets. Thus, if one uses a table look up to compute the rate achieve by a given set a constant speed up can be obtained. More importantly, however, using translated difference sets aids in a greedy search. Before proceeding to explain how this order is useful we note that the authors and others have reduced the search complexity by relating the scheduling problem to a search on a random graph [2], [3], [5]. In general finding maximal sub-structures in a random graph requires extensive search. However, finding a sub-graph of small size is much easier. The complexity of this problem is also much reduced if the underlying graph is strongly regular, which is the case for FFT based codes. In these cases, applications of backtracking algorithms work extremely well [14]. By greedily choosing sets based on the order of the translated sets one can improve on the worst case bound (16). Providing good bounds on this approach is the subject of ongoing work.

References

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

[2] 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, September 2004.

[3] ——, "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.

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

[5] ——, "Sum-rate optimal multi-antenna downlink beamforming strategy based on clique search," in *Proc. IEEE Globecom 2005*, November 2005, to appear.

[6] C. Swannack, G. W. Wornell, and E. Uysal-Biyikoglu, "MIMO broadcast scheduling with quantized channel state information," in *Proc. IEEE International Symposium on Information Theory*, Seattle, Washington, July 2006.

[7] P. Xia and G. Giannakis, "Design and analysis of transmit beamforming based on limited-rate feedback," *IEEE Transactions on Signal Processing*, vol. 54, no. 5, pp. 1853–1863, May 2006.

[8] L. R. Welch, "Lower bounds on the maximum cross correlation of signals," *IEEE Transactions on Information Theory*, vol. 20, no. 3, pp. 397 – 399, May 1974.

[9] G. A. Kabatyanskii and V. I. Lenenshtein, "Bounds for packings on a sphere and in space," *Problems Inform. Transm.*, vol. 14, pp. 1–17, 1978.

[10] P. Xia, S. Zhou, and G. Giannakis, "Achieving the welch bound with difference sets," *IEEE Transactions on Information Theory*, vol. 51, no. 5, pp. 1900– 1907, May 2005.

[11] N. L. Johnson and S. Kotz, *Urn models and their application: an approach to modern discrete probability theory*. New York: John Wiley & Sons, 1977.

[12] H. S. Wilf, *Generatingfunctionology*, 2nd ed. New York: Acemdemic Press, 1994.

[13] K. K. Mukkavilli, A. Sabharwal, E. Erkip, and B. Aazhang, "On beamforming with finite rate feedback in multiple-antenna systems," *IEEE Transactions on Information Theory*, vol. 49, no. 10, pp. 2562–2578, October 2003.

[14] M. Cimráková and V. Fack, "Clique algorithms for finding substructures in generalized quadrangles," *J. Combin. Math. Combin. Comput.*, 2006, (to appear).