Bit Allocation Laws for Multi-Antenna Channel Feedback Quantization: Multi-User Case

Behrouz Khoshnevis and Wei Yu

***Abstract*—This paper addresses the optimal design of limited-feedback downlink multi-user spatial multiplexing systems. A multiple-antenna base-station is assumed to serve multiple single-antenna users, who quantize and feed back their channel state information (CSI) through a shared rate-limited feedback channel. The optimization problem is cast in the form of minimizing the average transmission power at the base-station subject to users' target signal-to-interference-plus-noise ratios (SINR) and outage probability constraints. The goal is to derive the feedback bit allocations among the users and the corresponding channel magnitude and direction quantization codebooks in a high-resolution quantization regime. Toward this end, this paper develops an optimization framework using approximate analytical closed-form solutions, the accuracy of which is then verified by numerical results. The results show that, for channels in the real space, the number of channel direction quantization bits should be $(M-1)$ times the number of channel magnitude quantization bits, where M is the number of base-station antennas. Moreover, users with higher requested quality-of-service (QoS), i.e. lower target outage probabilities, and higher requested downlink rates, i.e. higher target SINR's, should use larger shares of the feedback rate. It is also shown that, for the target QoS parameters to be feasible, the total feedback bandwidth should scale logarithmically with the geometric mean of the target SINR values and the geometric mean of the inverse target outage probabilities. In particular, the minimum required feedback rate is shown to increase if the users' target parameters deviate from the corresponding geometric means. Finally, the paper shows that, as the total number of feedback bits B increases, the performance of the limited-feedback system approaches the perfect-CSI system as $2^{-B/M^2}$.**

***Index Terms*—Beamforming, bit allocation, channel quantization, limited feedback, multiple antennas, outage probability, power control, spatial multiplexing.**

I. Introduction

Multiple-input multiple-output (MIMO) technology can potentially provide significant performance improvements for wireless systems. More specifically, in the context of cellular communications, the availability of multiple antennas at the base-station allows it to simultaneously transmit to multiple users by multiplexing their data streams and hence improve the total downlink rate. Such systems are generally referred to as multi-user spatial multiplexing systems.

The performance of multi-user spatial multiplexing systems depends heavily on the amount of channel state information (CSI) at the base-station [1], [2]. Such information is necessary to form the downlink transmission beams at the base-station and to perform rate/power adaptation for each user. Acquiring channel information, however, is a challenging issue in practice especially for frequency-division duplex (FDD) systems, where the uplink and downlink channels use different frequency bands. In these systems, the users need to explicitly quantize their channels and send back the quantized information through a shared feedback channel. The feedback link in such a system is usually a rate-limited control channel, hence the term *limited-feedback systems*.

The scarcity of feedback bandwidth in limited-feedback systems necessitates efficient channel quantization codebook structures and an optimal allocation of feedback bits among those codebooks. This paper aims at deriving these structures and the corresponding bit allocation laws for limited-feedback multi-user spatial multiplexing systems.

A. Related Work

Multi-antenna communications with limited CSI is extensively studied in the literature for single-cell systems [1]–[10] and to some extent for cooperative multi-cell networks [11]–[13]. This paper focuses on single-cell multi-user systems.

The major advantage of multi-antenna multi-user systems with respect to single-user systems is the sum-rate *multiplexing gain*, which follows from the fact that multiple simultaneous transmissions can be established in a multi-antenna downlink. In order to preserve this gain in limited-feedback systems, the author of [2] shows that the total feedback rate should scale linearly with the signal-to-noise ratio (SNR) in dB scale. The work in [2] addresses a setup with small number of users. In a network with large number of users, there is another source of sum-rate improvement, referred to as *multi-user diversity gain*, which is realized by the base-station opportunistically scheduling users with favorable channel conditions. For scheduling, a well justified approach is to choose users with high channel gains and near-orthogonal channel directions [1], [3]–[7], [10]. The authors of [3] specifically show that one needs the channel gain information (CGI) in addition to the quantized channel direction information (CDI) in order to realize the multi-user diversity gain. The gain information however is assumed to be perfect in [3]. The split of feedback bits between CGI and CDI quantization introduces an interesting tradeoff between multiplexing gain and diversity gain. This tradeoff is studied by [7], where the authors numerically show that more bits should be used for CGI quantization in order to benefit from multi-user diversity gain as the number of users increases.

For the purpose of precoding the information intended to scheduled users, two distinct approaches are proposed

The material in this paper has been presented in part at the 44th Conference on Information Sciences and Systems, Princeton, NJ, Mar. 2010. This work was supported by the Natural Science and Engineering Research Council (NSERC) of Canada.

The authors are with the Edward S. Rogers Sr. Department of Electrical and Computer Engineering, University of Toronto, 10 King's College Road, Toronto, Ontario, Canada M5S 3G4 (email: bkhoshnevis@comm.utoronto.ca; weiyu@comm.utoronto.ca).

in the literature. The first group of work uses zero-forcing beamforming based on the quantized directions [3], [6]. The second group adopts a codebook (or multiple codebooks) of orthonormal beamforming vectors and selects beamforming directions based on the signal-to-interference-plus-noise ratio (SINR) feedbacks from the users [1], [4], [10]. The authors of [10] specifically claim that in the regime of large number of users, the orthogonal beamforming approach outperforms the zero-forcing method. Finally, the authors of [5] present a more thorough analysis of multi-user limited-feedback systems considering joint training, scheduling, and beamforming.

B. System Model

This paper considers a limited-feedback multi-user system with a M-antenna base-station and M single-antenna users. The users send their quantized CSI to the base-station through a feedback link with a capacity of B bits per each downlink transmission block. Based on this quantized information, the base-station then comes up with the downlink transmission powers and beamforming vectors.

Our goal is to optimize the system performance subject to the total feedback rate constraint. Although the proposed approach is rather generic, we impose several simplifying assumptions on the system model in order to achieve closed-form solutions. These assumptions are listed below and will be explicitly mentioned and justified whenever needed throughout the text:

Assumptions:

A1. Most of the analysis in this paper is based on the assumption of *high resolution quantization*, i.e. $B \to \infty$. This assumption is made mainly to achieve a tractable formulation for system optimization and has been used frequently in the literature of quantization theory [14], [15]. In addition, the paper investigates, both numerically and analytically, the system performance with moderate values of B, and examines the regime where the high resolution results are applicable.

A2. Users' channels are independent and identically distributed (i.i.d.).

A3. Users' channel directions are uniformly distributed over the M-dimensional unit hypersphere. This includes the Rayleigh i.i.d. channels as an special case [16].

A4. Users' channel magnitudes are independent from channel directions and can have an arbitrary distribution.

A5. We assume a product structure for the channel quantization codebook, i.e. channel magnitudes and channel directions are quantized independently. This product structure is justified in [8], [17]. According to the high resolution assumption, if we denote the magnitude and direction codebook sizes by $\dot{N}_k$ and $\ddot{N}_k$, we have $\dot{N}_k, \ddot{N}_k \to \infty$ for each user $1 \leq k \leq M$.

A6. The beamforming vectors are zero-forcing directions for the quantized directions. This assumption is to mimic the perfect-CSI case, where zero-forcing beamforming is shown to achieve asymptotically optimal sum-rate scaling with SNR [18], [19].

A7. The user channels are assumed to be real vectors. Although this assumption appears in the earlier literature, e.g. [20], [21], we use it mainly for the ease of geometric representation of the quantization regions and the corresponding calculations. The extension of the analysis to complex space is discussed in Section VIII.

C. Problem Formulation

We formulate the system design problem as the minimization of the average sum power subject to the users' outage probability constraints. In order to differentiate between the users' QoS requirements, we assume different target SINR's and outage probabilities across the users. Our goal is to derive the optimal split of feedback bits among the users and the corresponding magnitude and direction quantization codebooks.

Different variations of the power minimization formulation are used in the literature, e.g. in [22]–[26], as an appropriate formulation for fixed-rate delay-sensitive applications, e.g. voice over IP, video conferencing, and interactive gaming. The reliability of the fixed-rate link is usually achieved by applying power control at the base-station to compensate for the channel fading, as in Wideband Code Division Multiple Access (WCDMA) system standards [27].

An alternative problem formulation is to maximize the average sum rate subject to a power constraint [1]–[4], [6], [7], [10]. In most of these formulations, the transmission power is fixed and the quantized information is used only to adapt the transmission rates. This type of formulation is more appropriate for variable-rate communication systems, e.g. Worldwide Interoperability for Microwave Access (WiMAX) and 3GPP Long Term Evolution (LTE) system standards [28], [29].

We make the following comments on the existing rate-maximization formulations in the literature:

1. The sum-rate maximization problem assumes equal priority for all users in terms of their applications. Since the users are not differentiated based on their QoS measures, this formulation cannot answer the question of how to optimally split the feedback bits among users with different QoS requirements.
2. The channel gain information (CGI) is clearly an important factor in scheduling the users and also in setting the downlink rate for each user. However, most of the existing literature either assumes perfect channel gain information or completely ignores this information. It is therefore not clear how to optimally split the feedback bits between channel direction and channel gain quantizers in the context of sum-rate maximization problem. The works that look into this problem are mainly numerical and lack closed-form solutions [7], [10].

One solution to the first issue raised above is to prioritize users with certain weights and consider the weighted sum-rate maximization instead of the sum-rate itself. These weights can be set by the scheduler based on users' QoS requirements. The proportional fairness scheduler for example sets the weights based on users' backlogged traffic by assigning a higher weight to a user with larger backlog. This type of formulation appears for example in [30]–[33] for perfect-CSI systems. Generalizing this formulation to limited-feedback

systems however appears to be a difficult problem. We are not specifically aware of any work that addresses such a problem. For this reason, this paper resorts to a power minimization formulation, which can easily incorporate the QoS constraints by assuming different target SINRs and outage probabilities across the users. The power minimization formulation also simplifies the CGI/CDI bit allocation problem and provides insight to the system design by allowing for closed-form asymptotic bit allocation solutions.

D. Proposed Approach

Our approach for solving the power minimization problem is to fix the channel outage regions in advance and transform the problem to a robust optimization problem. Assuming zero-forcing beamforming vectors, we first formulate the robust power control problem in the form of a semi-definite programming (SDP) problem. Using an approximate upper bound solution to the SDP problem, we then derive the codebook structures and the corresponding bit allocations in the asymptotic regime where $B \to \infty$. Within the proposed approximate optimization framework, we show that for channels in real space:

1. The optimal number of channel direction quantization bits is $M-1$ times the number of channel magnitude quantization bits, where M is the number of base-station antennas.
2. The share of the kth user from the total feedback rate is controlled by $\log \gamma_k$ and $\log 1/q_k$, where γ_k and q_k are the user's target SINR and outage probability. As a general rule, a user with a lower target outage probability and higher target SINR needs a higher channel quantization resolution and therefore requires a larger share of the total feedback rate.
3. For the outage probability constraints to be feasible, the total feedback rate should scale logarithmically with $\bar{\gamma}$, the geometric mean of the target SINR values, and $1/\bar{q}$, the geometric mean of the inverse target outage probabilities. Moreover, the minimum required feedback rate increases if the users' target parameters deviate from the average parameters $\bar{\gamma}$ and $\bar{q}$, i.e. there is a feedback rate penalty for serving users with non-similar target parameters. The higher the deviation, the higher the penalty.
4. As the total feedback rate B increases, the performance of the limited CSI system approaches the performance of the perfect-CSI system as $2^{-\frac{B}{M2}}$.

These optimality results are based on minimizing an upper bound of the sum power. The closeness of the upper bound to the exact sum power is verified numerically in the paper.

E. Organization of the Paper

The remainder of this paper is organized as follows. Section II provides an overview of the perfect CSI system. Section III presents the system design problem in its general form and describes our approach in transforming it to a robust design problem. Section IV describes the product channel quantization codebook structure. In Section V, we study the power control optimization for fixed quantization codebooks and derive an upper bound for the average sum power. By using the sum power upper bound, we then optimize the product codebook structures in Section VI and derive the asymptotically optimal bit allocation laws. Finally, Section VII presents the numerical results and Section VIII concludes the paper.

Notations: Most of the computations in this paper are in real space. The logarithm functions are base 2. The angle between any two unit-norm vectors $\mathbf{u}$ and $\mathbf{v}$ is defined as $\angle(\mathbf{u},\mathbf{v}) = \arccos|\mathbf{u}^T\mathbf{v}|$ so that $0 \le \angle(\mathbf{u},\mathbf{v}) \le \pi/2$.

II. Multiuser Spatial Multiplexing System with Perfect CSI: Outage is Inevitable

We start by assuming perfect CSI at the base-station and show that, unlike a single-user system, outage is inevitable in the multi-user system even with perfect CSI. The difference with the single-user case is due to the fact that the base-station in a multi-user system needs to distinguish the users spatially and when the user channels are closely aligned it is not possible to satisfy the users' target SINR's with a bounded average transmission power.

Consider a multi-user downlink channel with M antennas at the base-station and M users each with a single antenna. Let $\mathbf{h}_k \in \mathbb{R}^M$, $\mathbf{v}_k \in \mathbb{R}^M$, P_k, and γ_k denote respectively, the user channel, the unit-norm beamforming vector, the allocated power, and the target SINR for the kth user, $1 \le k \le M$. The minimization of the transmission sum power subject the user SINR constraints is formulated as follows:

$$\min_{P_k,\mathbf{v}_k} \sum_{k=1}^{M} P_k \qquad (1)$$
$$\text{s.t. } \frac{P_k \left|\mathbf{h}_k^T\mathbf{v}_k\right|^2}{\sum_{l\neq k} P_l \left|\mathbf{h}_k^T\mathbf{v}_l\right|^2 + 1} \ge \gamma_k, \quad k = 1,2,\cdots,M$$

where the receiver noise power is assumed to be 1 for all users.

A suboptimal solution for problem (1) is to use zero-forcing (ZF) beamforming vectors $\mathbf{v}_k$ to eliminate the interference and find the power values P_k that satisfy the constraints with equality. This solution is asymptotically optimal in the high SNR regime [18], [19]. Clearly, for the zero-forcing solution to be applicable, the users' vector channels need to be linearly independent. Since the channels are assumed to be independent random vectors, this condition is satisfied almost surely, i.e., with probability one.

An important matter to consider with this solution is that the transmission powers would need to be extremely high when the users' channels are closely aligned, as the ZF beamforming vectors would be almost perpendicular to the corresponding channels in such cases. Therefore, it is not possible to always satisfy the SINR constraints with a bounded average power and as a result, a certain degree of outage must be tolerated by the users.

To see this rigorously, define $\theta_k = \angle(\mathbf{h}_k, \mathbf{H}_{-k})$, where $0 \le \theta_k \le \frac{\pi}{2}$, and $\mathbf{H}_{-k} = \text{span}(\{\mathbf{h}_l | l \neq k\})$. For zero-forcing beamforming vectors $\mathbf{v}_k$ we have $\angle(\mathbf{v}_k, \mathbf{h}_k) = \frac{\pi}{2} - \theta_k$. Assume

that the users' channels are i.i.d. with uniformly distributed directions and independent channel magnitudes (with arbitrary distributions). The average sum power of the zero-forcing method is given by

$$P_{\text{MU,CSI}}=\sum_{k=1}^{M}\mathbb{E}\left[\frac{\gamma_k}{\left|\mathbf{h}_k^T\mathbf{v}_k\right|^2}\right]=\sum_{k=1}^{M}\gamma_k\mathbb{E}\left[\frac{1}{\|\mathbf{h}_k\|^2}\right]\mathbb{E}\left[\frac{1}{\sin^2(\theta_k)}\right].$$

As θ_k is uniformly distributed in $[0,\frac{\pi}{2}]$, the expectation of $1/\sin^2(\theta_k)$ becomes unbounded.

To avoid unbounded transmit power, the users should tolerate certain degrees of outage. A reasonable approach is to declare outage for user k, i.e. set $P_k=0$, when

$$0\le\theta_k<\theta_k^\circ,$$

where $\theta_k^\circ\ll 1$ is the smallest acceptable angle between $\mathbf{h}_k$ and $\mathbf{H}_{-k}$. With this assumption the average sum power is given by

$$\begin{aligned}P_{\text{MU,CSI}}&=\sum_{k=1}^{M}\gamma_k\mathbb{E}\big[1/\|\mathbf{h}_k\|^2\big]\frac{1}{\pi/2}\int_{\theta_k^\circ}^{\pi/2}\frac{\mathrm{d}\theta_k}{\sin^2\theta_k}\\&=\frac{2}{\pi}\sum_{k=1}^{M}\gamma_k\mathbb{E}\big[1/\|\mathbf{h}_k\|^2\big]\cot\theta_k^\circ\approx\frac{2\rho_{\text{MU,CSI}}}{\pi}\sum_{k=1}^{M}\frac{\gamma_k}{\theta_k^\circ},\end{aligned}\quad(2)$$

where the approximation holds for $\theta_k^\circ\ll 1$ and

$$\rho_{\text{MU,CSI}}\stackrel{\text{def}}{=}\mathbb{E}\big[1/\|\mathbf{h}_k\|^2\big]\quad(3)$$

for i.i.d. users. The corresponding outage probabilities are $p_{\text{out},k}=2\theta_k^\circ/\pi$ for $1\le k\le M$.

Having studied the perfect CSI system, the next section describes a general framework for the limited-feedback system design. The insights achieved by studying this general form are used in the later sections for system design and optimization with product channel quantization codebooks.

III. System Design Problem and Vector Channel Quantization: General Form

To clarify the arguments, we start by some basic definitions. By a *vector channel quantization codebook* $\mathcal{C}$ of size N, we mean a partition of $\mathbb{R}^M$ into N disjoint *quantization regions* $S^{(n)}$, $1\le n\le N$:

$$\mathcal{C}=\{S^{(1)},S^{(2)},\cdots,S^{(N)}\}.$$

For every quantization codebook $\mathcal{C}$, we also define a *quantization function*

$$\mathcal{S}(\mathbf{h}):\mathbb{R}^M\to\mathcal{C},$$

which returns the quantization region that $\mathbf{h}\in\mathbb{R}^M$ belongs to.

Now, for each user $1\le k\le M$, associate a codebook $\mathcal{C}_k$ of size N_k and the corresponding quantization function $\mathcal{S}_k(\mathbf{h}_k)$, where $\mathbf{h}_k$ is the kth user's channel. Further, define the ordered M-tuples

$$\begin{aligned}&\mathbf{H}\stackrel{\text{def}}{=}[\mathbf{h}_1^T,\mathbf{h}_2^T,\cdots,\mathbf{h}_M^T]\in\mathbb{R}^{M^2},\\&\mathbf{S}(\mathbf{H})\stackrel{\text{def}}{=}[\mathcal{S}_1(\mathbf{h}_1),\mathcal{S}_2(\mathbf{h}_2),\cdots,\mathcal{S}_M(\mathbf{h}_M)]\in\prod_{k=1}^{M}\mathcal{C}_k.\end{aligned}$$

For a given total number of quantization (feedback) bits B, target SINR values γ_k, and target outage probabilities q_k, the system design problem is formulated as follows:

$$\min_{\substack{\mathcal{C}_k,N_k,\\P_k(\mathbf{S}(\mathbf{H})),\\\mathbf{v}_k(\mathbf{S}(\mathbf{H}))}}\ \mathbb{E}_{\mathbf{H}}\left[\sum_{k=1}^{M}P_k(\mathbf{S}(\mathbf{H}))\right]\quad(4)$$

$$\text{s.t.}\quad\prod_{k=1}^{M}N_k=2^B,$$

$$\text{prob}\left[\frac{P_k(\mathbf{S}(\mathbf{H}))\left|\mathbf{h}_k^T\mathbf{v}_k(\mathbf{S}(\mathbf{H}))\right|^2}{\sum\limits_{l\ne k}P_l(\mathbf{S}(\mathbf{H}))\left|\mathbf{h}_k^T\mathbf{v}_l(\mathbf{S}(\mathbf{H}))\right|^2+1}<\gamma_k\right]\le q_k,\quad k=1,2,\cdots,M$$

where the optimization is over the quantization codebooks $\mathcal{C}_k$, codebook sizes N_k, the power control functions $P_k(\mathbf{S}(\mathbf{H})):\prod_{k=1}^{M}\mathcal{C}_k\to\mathbb{R}_+$, and the beamforming functions $\mathbf{v}_k(\mathbf{S}(\mathbf{H})):\prod_{k=1}^{M}\mathcal{C}_k\to\mathfrak{U}_M$, where $\mathfrak{U}_M$ is the unit hypersphere in $\mathbb{R}^M$.

An exact solution to this problem is intractable. Our approach in simplifying the problem is to fix the outage scenarios in advance and transform the design problem to a robust design problem that guarantees the target SINR's for the no-outage scenarios.

Define the *outage region* $\Omega_k\subset\prod_k\mathcal{C}_k$ for user k such that $\text{prob}[\mathbf{S}(\mathbf{H})\in\Omega_k]=q_k$. Also define $I_k(\mathbf{S}(\mathbf{H}))$ as the *activity flag* for user k:

$$I_k(\mathbf{S}(\mathbf{H}))=\mathcal{I}(\mathbf{S}(\mathbf{H})\in\Omega_k^c),$$

where $\mathcal{I}(\cdot)$ is the logic true function. Whenever a user's channel resides outside the user's predefined outage region, the activity flag is on and the user must be served by the base-station, i.e. the user should not face an outage.

Let us fix the codebook sizes N_k for now. For a robust system design, we need to design the codebooks, the power control functions, and the beamforming functions such that the target SINR's are guaranteed whenever $I_k(\mathbf{S}(\mathbf{H}))=1$:

$$\min_{\substack{\mathcal{C}_k,\\P_k(\mathbf{S}(\mathbf{H})),\\\mathbf{v}_k(\mathbf{S}(\mathbf{H}))}}\ \mathbb{E}_{\mathbf{H}}\left[\sum_{k=1}^{M}P_k(\mathbf{S}(\mathbf{H}))\right]\quad(5)$$

$$\text{s.t.}\ \inf_{\mathbf{w}\in\mathcal{S}_k(\mathbf{h}_k)}\frac{P_k(\mathbf{S}(\mathbf{H}))\left|\mathbf{w}^T\mathbf{v}_k(\mathbf{S}(\mathbf{H}))\right|^2}{\sum\limits_{l\ne k}P_l(\mathbf{S}(\mathbf{H}))\left|\mathbf{w}^T\mathbf{v}_l(\mathbf{S}(\mathbf{H}))\right|^2+1}\ge\gamma_k I_k(\mathbf{S}(\mathbf{H})),\quad\forall\ \mathbf{H}\in\mathbb{R}^{M^2}\ \text{and}\ k=1,2,\cdots,M\quad(6)$$

Note that by including the activity flag in the constraint (6), this formulation guarantees the target SINR when $I_k=1$ and returns $P_k=0$ when $I_k=0$. Also note that the activity flags are fixed in advance such that $\text{prob}[I_k(\mathbf{S}(\mathbf{H}))=0]=q_k$.

The design problem in (5) is a complicated problem. In order to achieve a tractable reformulation, we accept two main simplifying assumptions (Assumptions A5 and A6 in Section I-B):

- We assume a product structure for the channel quantization codebook, where the channel magnitude and

the channel direction are quantized independently. Such a product structure, also known as shape-gain quantization in the literature [14], provides several practical advantages including faster quantization and lower storage requirement for the quantization codebooks. Product codebook structures are also shown to be sufficient structures for effective interference management in multi-user systems [8].

- We assume a fixed zero-forcing design for the following reason. Let us fix the quantization codebooks $\mathcal{C}_k$ and consider a time snapshot of the optimization problem in (5) with a fixed channel realization $\mathbf{H}$ and fixed quantized information $\mathbf{S}(\mathbf{H})$. For this specific point in time, we intend to optimize the beamforming vectors $\mathbf{v}_k$ and power levels P_k such that the worst-case SINR conditions in (6) are satisfied. It can be shown that this problem is non-convex with respect to P_k and $\mathbf{v}_k$ and therefore difficult to solve in general. However, if one fixes the vectors $\mathbf{v}_k$, the problem would be convex in terms of P_k, since the objective and constraints are linear in P_k. Similar observations are made in [34]–[36]. In order to preserve the convexity structure, we therefore assume that the beamforming vectors $\mathbf{v}_k(\mathbf{S}(\mathbf{H}))$ are fixed as zero-forcing vectors for the quantized channel directions. This is to mimic the zero-forcing beamforming with perfect CSI, where the base-station knows the exact channel directions. The exact definition of the quantized channel directions is provided later in Section IV-B.

With these simplifying assumptions, the robust design problem reduces to the following subproblems: 1) optimizing the power control function for fixed beamforming vectors and codebook structures; 2) optimizing the product codebook structure itself. The following sections address these subproblems.

IV. Product Quantization Codebook Structure

In this section, we describe the product quantization codebook structures and specify the corresponding outage regions. To be more exact, for a given target outage probability q_k, we specify the magnitude and direction outage regions such that

$$q_k = \dot{q}_k + \ddot{q}_k,$$

where the *magnitude outage probability* $\dot{q}_k$ is the probability that the channel magnitude resides in the specified *magnitude outage region* and the *direction outage probability* $\ddot{q}_k$ is the probability that the channel direction resides in the specified *direction outage region*.

A. Magnitude Quantization Codebook and Magnitude Outage Region

For each user $1 \leq k \leq M$, we use a magnitude quantization codebook

$$\mathbb{Y}_k = \left\{ y_k^{(1)}, y_k^{(2)}, \cdots, y_k^{(\dot{N}_k)} \right\}$$

for quantizing the channel magnitude squared $Y_k \stackrel{\text{def}}{=} \|\mathbf{h}_k\|^2$. Here $y_k^{(n)}$ are the quantization levels and $\dot{N}_k$ is the magnitude codebook size.

For a given *magnitude outage probability* $\dot{q}$, we define the *magnitude outage region* as the leftmost quantization interval $\left[0, y_k^{(1)}\right)$. The first quantization level is therefore fixed as

$$y_k^{(1)} = F^{-1}(\dot{q}_k), \tag{7}$$

where $F^{-1}(\cdot)$ is the inverse cumulative distribution function (cdf) of $Y_k \stackrel{\text{def}}{=} \|\mathbf{h}_k\|^2$.

We further define $\dot{\mathcal{C}}_k$ as the set of magnitude quantization regions, i.e. the set of quantization intervals for $\|\mathbf{h}_k\| = \sqrt{Y_k}$:

$$\dot{\mathcal{C}}_k = \left\{ J_k^{(1)}, J_k^{(2)}, \cdots, J_k^{(\dot{N}_k)} \right\}, \tag{8}$$

where $J_k^{(n)} = \left[\sqrt{y_k^{(n)}}, \sqrt{y_k^{(n+1)}} \right)$ and $y_k^{(\dot{N}_k+1)} \stackrel{\text{def}}{=} \infty$. Note that the definition uses the square root of the levels as the quantization levels $y_k^{(n)}$ are defined for quantizing $\|\mathbf{h}_k\|^2$.

Finally, for $Y_k \geq y_k^{(1)}$, we define the *quantized magnitude* $\tilde{Y}_k$ as the quantization level in $\mathbb{Y}_k$ that is in the immediate left of Y_k, i.e.,

$$\tilde{Y}_k = y_k^{(n)} \quad \text{if} \quad y_k^{(n)} \leq Y_k < y_k^{(n+1)}. \tag{9}$$

For reasons that are clarified later in Section VI, we are interested in a magnitude quantization codebook that minimizes $\mathbb{E}\left[1/\tilde{Y}_k\right]$. It is shown in [37] that the optimal codebook with such a criterion is uniform (in dB scale) in the asymptotic regime where $\dot{N}_k \to \infty$. We denote such an optimal codebook by $\mathbb{Y}_k^\star$ and refer to it as the *uniform magnitude quantization codebook* in the remainder of this paper.

The work in [37] further shows that the uniform codebook $\mathbb{Y}_k^\star$ satisfies the following upper bound:

$$\mathbb{E}\left[\frac{1}{\tilde{Y}_k^\star}\right] < \rho_{\text{MU,CSI}} \left(1 + \dot{N}_k^{-\zeta_k\left(\dot{N}_k\right)} + \omega \dot{N}_k^{-2\zeta_k\left(\dot{N}_k\right)}\right), \tag{10}$$

where

$$\rho_{\text{MU,CSI}} = \mathbb{E}[1/Y_k] = \mathbb{E}[1/\|\mathbf{h}_k\|^2]$$

as defined in (3). On the left hand side of (10), the variable $\tilde{Y}_k^\star$ is the quantized magnitude variable associated with the uniform codebook $\mathbb{Y}_k^\star$. On the right-hand side of (10),

$$\omega \stackrel{\text{def}}{=} \frac{\mathbb{E}[Y_k]}{\eta^2 \mathbb{E}[1/Y_k]},$$

where

$$\eta = \lim_{y \to \infty} -f(y)/f'(y)$$

and $f(\cdot)$ is the probability density function (pdf) of Y_k. The function $\zeta_k(n)$ depends on the magnitude outage probability $\dot{q}_k$ and is defined as the solution to the following equation:

$$n^{-\zeta_k(n)} \left(1 + n^{-\zeta_k(n)}\right)^{n-1} = \frac{\eta}{y_k^{(1)}},$$

where $y_k^{(1)} = F^{-1}(\dot{q}_k)$ as defined earlier. It can be shown that for any $\dot{q}_k > 0$ and hence $y_k^{(1)} > 0$, we have

$$\lim_{n \to \infty} \zeta_k(n) = 1. \tag{11}$$

The bound in (10) and the limit in (11) are used in Section VI for optimization of the product channel quantization codebook structure.

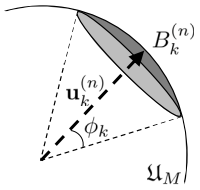


Fig. 1. Spherical cap $B_k^{(n)}$ around $\mathbf{u}_k^{(n)}$ on the unit hypersphere $\mathfrak{U}_M$.

B. Direction Quantization Codebook and Direction Outage Region

For each user $1\leq k\leq M$, we use a Grassmannian codebook $\mathbb{U}_k$ of size $\ddot{N}_k$ for direction quantization:

$$\mathbb{U}_k = \left\{ \mathbf{u}_k^{(1)}, \mathbf{u}_k^{(2)}, \cdots, \mathbf{u}_k^{(\ddot{N}_k)} \right\}, \tag{12}$$

where $\mathbf{u}_k^{(n)}$ vectors are M-dimensional unit-norm Grassmannian codewords.

Every channel realization $\mathbf{h}_k$ is mapped to a vector $\tilde{\mathbf{u}}_k(\mathbf{h}_k) \in \mathbb{U}_k$ that has the smallest angle with $\mathbf{h}_k$:

$$\tilde{\mathbf{u}}_k(\mathbf{h}_k) = \arg\min_{\mathbf{u}\in\mathbb{U}_k} \angle(\mathbf{h}_k, \mathbf{u}). \tag{13}$$

The vector $\tilde{\mathbf{u}}_k$ is referred to as the *quantized direction* for the channel realization $\mathbf{h}_k$. The corresponding quantization regions, according to the Gilbert-Varshamov argument [38], can be covered by the following spherical caps:

$$\ddot{\mathcal{C}}_k = \left\{ B_k^{(1)}, B_k^{(2)}, \cdots, B_k^{(\ddot{N}_k)} \right\}, \tag{14}$$

where

$$B_k^{(n)} = \left\{ \mathbf{w} \in \mathfrak{U}_M \,\middle|\, \angle(\mathbf{w}, \mathbf{u}_k^{(n)}) < \phi_k \right\}$$

is the spherical cap around $\mathbf{u}_k^{(n)}$ as shown in Fig. 1. In this definition,

$$\phi_k = \arcsin \delta_k$$

is the angular opening of the caps and δ_k is the minimum chordal distance of $\mathbb{U}_k$. It should be noted that covering the direction quantization regions with the spherical caps enlarges the quantization regions. By considering the constraint (6) in the robust design problem (5), such enlargement of the regions will lead to an upper bound for the average transmission power.

In order to describe the *direction outage regions*, define

$$\theta_k = \angle(\tilde{\mathbf{u}}_k, \tilde{\mathbf{U}}_{-k}),$$

where

$$\tilde{\mathbf{U}}_{-k} = \text{span}(\{\tilde{\mathbf{u}}_l | l \neq k\})$$

and $\tilde{\mathbf{u}}_k$ is the quantized direction for user k. This is a similar definition as in Section II, except that the exact channels $\mathbf{h}_k$ are replaced with the quantized directions $\tilde{\mathbf{u}}_k$. Similar to the discussion in Section II for the perfect CSI system, we say that user k is in *direction outage* if

$$0 \leq \theta_k < \theta_k^\circ,$$

where θ_k° is the minimum acceptable angle between $\tilde{\mathbf{u}}_k$ and $\mathbf{U}_{-k}$. This implicitly defines the direction outage regions of the users. Finally, assuming that θ_k is approximately uniform in $[0, \pi/2]$[1], the *direction outage probability* is given by

$$\ddot{q}_k \approx \frac{2}{\pi}\theta_k^\circ. \tag{15}$$

In Section VI, which addresses the product codebook optimization, we will need the following inequality, which describes the dependence between the angular opening ϕ_k and the direction codebook size $\ddot{N}_k$:

$$\phi_k \approx \sin\phi_k < 4\lambda_M \ddot{N}_k^{-\frac{1}{M-1}}, \tag{16}$$

where $\lambda_M = (\sqrt{\pi}\Gamma((M+1)/2)/\Gamma(M/2))^{\frac{1}{M-1}}$. This inequality holds for large enough values of $\ddot{N}_k$ and its proof is presented in [37]. The approximation on the left-hand side of (16) assumes $\phi \ll 1$, which is justified by the high resolution assumption $\ddot{N}_k \gg 1$ (Assumption A1 and A5 in Section I-B).

C. Product Codebook Structure

Using our definitions of the magnitude and direction quantization regions, we define product channel quantization codebook $\mathcal{C}_k$ for each user k as follows:

$$\mathcal{C}_k = (\dot{\mathcal{C}}_k \times \ddot{\mathcal{C}}_k) \cup \mathcal{O}_k, \tag{17}$$

where $\dot{\mathcal{C}}$ and $\ddot{\mathcal{C}}$ are the magnitude and direction quantization regions in (8) and (14), and

$$\mathcal{O}_k = \left\{ \mathbf{h} \,\middle|\, \|\mathbf{h}\| < \sqrt{y_k^{(1)}} \right\}$$

is a ball centered at origin corresponding to the magnitude outage region.

Based on the definitions of the magnitude and direction outage regions, the activity flag for user k is given by

$$I_k = \mathcal{I}\left(\theta_k \geq \theta_k^\circ \ \wedge \ \|\mathbf{h}\| \geq \sqrt{y_k^{(1)}} \right), \tag{18}$$

where $\mathcal{I}(\cdot)$ is the logic true function, and θ_k, θ_k°, and $y_k^{(1)}$ are defined in Sections IV-A and IV-B.

According to the activity flag expression in (18), the outage event for user k can be expressed as the union of the magnitude outage event, corresponding to the channel magnitude lying in the magnitude outage region, and the direction outage event, corresponding to the channel direction lying in the direction outage region. By using the union probability formula we have

$$\text{prob}[I_k = 0] \leq \dot{q}_k + \ddot{q}_k \approx F(y_k^{(1)}) + \frac{2}{\pi}\theta_k^\circ,$$

where $\dot{q}_k$ and $\ddot{q}_k$ are the magnitude and direction outage probabilities respectively and the approximation follows from (15). In order to satisfy the target outage probability q_k, we therefore impose the following constraint[2]:

$$F(y_k^{(1)}) + \frac{2}{\pi}\theta_k^\circ \leq q_k. \tag{19}$$

[1]This holds if the channel direction are uniformly distributed (Assumption A3 in Section I-B) and the codebooks $\mathbb{U}_k$ undergo sufficient random rotations.

[2]Note that this constraint is stronger than $\text{prob}[I_k = 0] \leq q_k$ and would therefore lead to an upper bound on the objective function (average sum power). This agrees with the direction of our analysis in achieving an upper bound for the sum power.

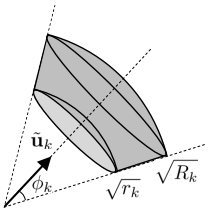


Fig. 2. Sector-type channel uncertainty region $S_k = S(R_k, r_k, \tilde{\mathbf{u}}_k, \phi_k)$ for user k.

Finally, the product channel quantization codebook size is given by

$$|\mathcal{C}_k| = N_k = \dot{N}_k \ddot{N}_k + 1, \tag{20}$$

where $\dot{N}_k$ and $\ddot{N}_k$ are the magnitude and direction codebook sizes respectively.

To summarize, for a given target outage probability q_k and a given codebook size N_k, we propose a product quantization structure and specify its outage regions and the corresponding activity flags such that prob$[I_k = 0] = q_k$. The proposed product codebook structure is parameterized by the magnitude and direction codebook sizes $\dot{N}_k$ and $\ddot{N}_k$, the minimum acceptable channel magnitude $y_k^{(1)}$, and the minimum acceptable directional separation θ_k°.

V. Optimization of the Power Control Function with Sector-Type Channel Uncertainty Regions

Assuming the product codebook structure in Section IV, this section addresses the optimization of the power control function. For this purpose, we fix the quantization codebooks $\mathcal{C}_k$ and the corresponding outage regions. Furthermore, as mentioned in Section III, we make the simplifying assumption that the beamforming vectors are the zero-forcing vectors for the quantized directions.

For the product quantization codebooks considered in this paper, the quantization (or channel uncertainty) regions are sector-type regions as shown in Fig. 2. A sector-type region is parameterized as

$$S(R, r, \tilde{\mathbf{u}}, \phi) = \left\{ \mathbf{h} \in \mathbb{R}^M \middle| \sqrt{r} \le \|\mathbf{h}\| < \sqrt{R},\ \angle(\mathbf{h}, \tilde{\mathbf{u}}) < \phi \right\},$$

where in the terminology of Section IV, $\tilde{\mathbf{u}}$ is the *quantized direction* and r is the *quantized magnitude*, which is denoted as $\tilde{Y}$ in (9).

For a specific point in time, consider the channel realizations $\mathbf{H} = [\mathbf{h}_1^T, \mathbf{h}_2^T, \cdots, \mathbf{h}_M^T]$ and the corresponding quantization (or channel uncertainty) regions

$$S_k = S(R_k, r_k, \tilde{\mathbf{u}}_k, \phi_k) \stackrel{\text{def}}{=} \mathcal{S}_k(\mathbf{h}_k),$$

where the *quantization functions* $\mathcal{S}_k(\mathbf{h}_k)$, $1 \le k \le M$, are defined in Section III. Also, let I_k denote the corresponding activity flags.

The goal is to optimize the power control function for the robust design problem (5). Therefore, for the current channel realizations $\mathbf{H}$, we have to find the transmission power levels P_k that minimize the instantaneous sum power subject to the worst-case SINR constraints:

$$\min_{P_k} \sum_{k=1}^{M} P_k \tag{21}$$

$$\text{s.t.} \inf_{\mathbf{w} \in S_k} \frac{P_k \left|\mathbf{w}^T \mathbf{v}_k\right|^2}{\sum_{l \ne k} P_l \left|\mathbf{w}^T \mathbf{v}_l\right|^2 + 1} \ge \gamma_k I_k, \quad k=1,2,\cdots,M \tag{22}$$

where the beamforming vectors $\mathbf{v}_k$ are fixed, since the quantized directions $\tilde{\mathbf{u}}_k$ are fixed. Let us refer to the users with $I_k = 0$ as the *silent* users and the users with $I_k = 1$ as the *active* users and let the set $\mathcal{K}$ denote the set of active users: $\mathcal{K} = \{1 \le k \le M | I_k = 1\}$. In general $\mathcal{K}$ is a random set depending on the channel realizations and the specified outage regions. Now, considering the power control problem in (21), we note that if a user k is silent, i.e. $I_k = 0$, the corresponding SINR constraint in (22) is redundant as the problem returns $P_k = 0$ for such a user. We therefore confine the SINR constraints in (22) to the set of active users, i.e. the indices $k \in \mathcal{K}$.

According to the robust design formulation in (5), for any channel realization $\mathbf{H} = [\mathbf{h}_1^T, \mathbf{h}_2^T, \cdots, \mathbf{h}_M^T]$, the base-station is required to serve (guarantee the target SINR's for) all active users $k \in \mathcal{K}$. The power control problem in (21) therefore must be feasible for the active users. The following theorem presents a sufficient condition that guarantees feasibility.

Theorem 1: To ensure the feasibility of the robust power control problem in (21), it is sufficient to have the following for all $1 \le k \le M$:

$$\ddot{N}_k \ge \left(4\lambda_M \Big/ \sin\left(\arctan\left(\frac{\sin\theta_k^\circ}{1 + \sqrt{(M-1)\gamma_k}} \right) \right) \right)^{M-1}. \tag{23}$$

Proof: See Appendix A. ■

This condition is referred to as the minimum quantization codebook size (MQCS) condition in the remainder of this paper.

In the high resolution regime, where the codebook sizes tend to infinity, the MQCS conditions are clearly satisfied; therefore, feasibility of the power control problem is not an issue as far as the high resolution analysis is considered. This condition however plays a key role in finding the minimum number of feedback bits B for which the asymptotic bit allocation laws are applicable. This issue is discussed in further detail in Theorem 5 of Section VI.

We are now ready to solve the power control problem in (21). An exact numerical solution to this problem can be obtained by transforming it into a semidefinite programming (SDP) problem as described in the following.

Theorem 2: The problem in (21) is equivalent to the following SDP problem for $M \ge 3$:

$$\min_{P_k, \lambda_k, \mu_k} \sum_{k \in \mathcal{K}} P_k \tag{24}$$

$$\text{s.t.} \quad \frac{1}{\gamma_k} P_k \mathbf{v}_k \mathbf{v}_k^T - \sum_{l \ne k} P_l \mathbf{v}_l \mathbf{v}_l^T \succeq (\lambda_k - \mu_k)\mathbf{I}_M + \frac{\mu_k}{\cos^2\phi_k} \tilde{\mathbf{u}}_k \tilde{\mathbf{u}}_k^T$$

$$\lambda_k \ge \frac{1}{r_k},\ \mu_k \ge 0,\ P_k \ge 0,\ k \in \mathcal{K},$$

where $\mathbf{I_M}$ is the $M\times M$ identity matrix.

Proof: The proof is based on the Polyak's theorem in [39]. See Appendix B for details. ■

Although the SDP reformulation provides a numerically efficient solution, it does not give the minimized sum power in a closed form. The availability of a closed-form expression for sum power is crucial in optimizing the quantization codebook structures. Moreover, for practical systems where the base-station needs to continuously compute and update the transmission powers for each fading block, the SDP reformulation would be of a limited application. We therefore resort to a suboptimal solution to the problem in (21) that provides a closed-form upper bound to the sum power. This upper bound solution is used later in Section VI as the objective function for optimization of the users' quantization codebooks.

First, we bound the SINR terms as follows. For the sector-type regions $S_k = S(R_k, r_k, \tilde{\mathbf{u}}_k, \phi_k)$, we have

$$\inf_{\mathbf{w}\in S_k} \frac{P_k \left|\mathbf{w}^T\mathbf{v}_k\right|^2}{\sum\limits_{l\neq k} P_l \left|\mathbf{w}^T\mathbf{v}_l\right|^2 + 1} \overset{\text{(a)}}{=} \inf_{\mathbf{w}\in S_k} \frac{P_k r_k |\hat{\mathbf{w}}^T\mathbf{v}_k|^2}{r_k\sum\limits_{l\neq k} P_l |\hat{\mathbf{w}}^T\mathbf{v}_l|^2+1}$$

$$\geq \frac{P_k r_k \inf_{\mathbf{w}\in S_k} |\hat{\mathbf{w}}^T\mathbf{v}_k|^2}{r_k\sum\limits_{l\neq k} P_l \sup_{\mathbf{w}\in S_k} |\hat{\mathbf{w}}^T\mathbf{v}_l|^2+1} \quad (25)$$

$$\overset{\text{(b)}}{=} \frac{P_k r_k \sin^2(\theta_k-\phi_k)}{(\sum_{l\neq k} P_l) r_k \sin^2\phi_k + 1}, \quad (26)$$

where $\hat{\mathbf{w}} = \mathbf{w}/\|\mathbf{w}\|$. The equality (a) holds since the SINR term is monotonic in $\|\mathbf{w}\|$, i.e. the minimum occurs on the spherical boundary region $\|\mathbf{w}\|=\sqrt{r_k}$ in Fig. 2. The equality (b) holds since $\mathbf{v}_k$'s are the zero-forcing directions for $\tilde{\mathbf{u}}_k$'s. To see this, consider $\theta_k = \angle(\tilde{\mathbf{u}}_k, \tilde{\mathbf{U}}_{-k})$ as defined earlier. By considering the zero-forcing principle, we have $\angle(\mathbf{v}_k, \tilde{\mathbf{u}}_k) = \frac{\pi}{2} - \theta_k$ and $\angle(\mathbf{v}_l, \tilde{\mathbf{u}}_k) = 0$ for $l \neq k$. Now, noting the definition of the sector-type region S_k, we have

$$\max_{\mathbf{w}\in\mathcal{S}_k} \angle(\mathbf{w}, \mathbf{v}_k) = \frac{\pi}{2} - \theta_k + \phi_k,$$
$$\min_{\mathbf{w}\in\mathcal{S}_k} \angle(\mathbf{w}, \mathbf{v}_l) = \frac{\pi}{2} - \phi_k.$$

By substituting the cosine of these angles in the numerator and denominator of (25), we achieve the final expression in (26).

In order to obtain an upper bound on the sum power, we set the last term in (26) to be equal to γ_k:

$$\frac{P_k r_k \sin^2(\theta_k-\phi_k)}{(\sum_{l\neq k} P_l) r_k \sin^2\phi_k + 1} = \gamma_k.$$

This is a set of linear equations in P_k, $k \in \mathcal{K}$, where $\mathcal{K}$ is the set of active users. By solving these equations and computing $\sum_{k\in\mathcal{K}} P_k$, we achieve the following upper bound for the sum power:

$$P_{\text{MU}} \overset{\text{def}}{=} \sum_{k\in\mathcal{K}} P_k = \frac{\sum_{k\in\mathcal{K}} \alpha_k/\beta_k}{1 - \sum_{k\in\mathcal{K}} \alpha_k}, \quad (27)$$

where $\alpha_k = \left(1 + \frac{\sin^2(\theta_k-\phi_k)}{\gamma_k \sin^2\phi_k}\right)^{-1}$ and $\beta_k = r_k \sin^2\phi_k$. The subscript MU in P_{MU} stands for *multi-user*. The closeness of the upper bound solution in (27) and the solution to the SDP problem in (24) is verified numerically in Section VII.

The upper bound in (27) is a bound on the instantaneous sum power for a single snapshot of channel realizations in time. As the users' channels change over time, the quantized magnitudes $r_k = \tilde{Y}_k$, the quantized directions $\tilde{\mathbf{u}}_k$ and the corresponding angles $\theta_k = \angle(\tilde{\mathbf{u}}_k, \tilde{\mathbf{U}}_{-k})$ all change with time. The variables α_k and β_k in (27) are therefore random variables. Since we are interested in the expected value of the sum power as the design objective in (5), we use the following sum-power upper bound approximation so that the expectation operation can be applied conveniently.

Theorem 3: In the asymptotic regime with large quantization codebook sizes and small values of ϕ_k, we have

$$P_{\text{MU}} = \sum_{k\in\mathcal{K}} e_k + \sum_{k\in\mathcal{K}} f_k\phi_k + \sum_{k\in\mathcal{K}} o(\phi_k), \quad (28)$$

where

$$e_k = \frac{\gamma_k}{r_k}(1+\zeta_k^2), \quad f_k = \frac{2\gamma_k}{r_k}(\zeta_k+\zeta_k^3), \quad \zeta_k = \cot\theta_k. \quad (29)$$

Here, the notation $g(\phi) = o(\phi)$, for an arbitrary function $g(\cdot)$, means that $\lim_{\phi\to 0} g(\phi)/\phi = 0$.

Proof: See Appendix C. ■

So far, we have only considered the active users $k \in \mathcal{K}$. In order to make the results applicable to the general case where some users might be in outage, we substitute γ_k with $\gamma_k I_k$ in definitions of e_k and f_k in (28) so that users with $I_k = 0$ contribute zero power to the sum-power upper bound. We therefore use the following expression for the average sum-power upper bound, where we have also replaced r_k by the quantized magnitude $\tilde{Y}_k$:

$$\mathbb{E}[P_{\text{MU}}] \approx \sum_k \mathbb{E}[e_k] + \sum_k \mathbb{E}[f_k]\phi_k, \quad (30)$$

where $\zeta_k = \cot\theta_k$ and e_k and f_k are redefined as follows:

$$e_k = \frac{\gamma_k I_k}{\tilde{Y}_k}(1+\zeta_k^2), \quad f_k = \frac{2\gamma_k I_k}{\tilde{Y}_k}(\zeta_k+\zeta_k^3).$$

This concludes the optimization of the power control function. In the next section, we use the average sum-power upper bound in (30) to optimize the product quantization codebook structures and to derive the asymptotic bit allocation laws.

VI. Product Codebook Optimization and Asymptotic Bit Allocation Laws

In this section, we study the quantization codebook optimization. For this purpose, we use the average transmission power bound in (30) in order to optimize the users' magnitude and direction codebook sizes for a given feedback link capacity constraint and to derive the optimal bit allocation across the users and their magnitude and direction quantization codebooks. The optimization process is asymptotic in the feedback rate B and assumes large quantization codebook sizes, $\dot{N}_k, \ddot{N}_k \gg 1$.

Consider the sum-power upper bound in (30). Assuming that $\theta_k = \angle(\tilde{\mathbf{u}}_k, \tilde{\mathbf{U}}_{-k})$ is approximately uniform in $[0, \pi/2]$

and using the definition of the activity flag in (18), we have

$$\mathbb{E}[e_k] = \gamma_k \mathbb{E}\left[1/\tilde{Y}_k\right] \frac{1}{\pi/2} \int_{\theta_k^\circ}^{\pi/2} 1+\zeta_k^2 \, \mathrm{d}\theta_k \approx \frac{2\gamma_k}{\pi\theta_k^\circ} \mathbb{E}\left[1/\tilde{Y}_k\right],$$

$$\mathbb{E}[f_k] = 2\gamma_k \mathbb{E}\left[1/\tilde{Y}_k\right] \frac{1}{\pi/2} \int_{\theta_k^\circ}^{\pi/2} \zeta_k+\zeta_k^3 \, \mathrm{d}\theta_k \approx \frac{2\gamma_k}{\pi{\theta_k^\circ}^2} \mathbb{E}\left[1/\tilde{Y}_k\right],$$

where the approximations hold for $\theta_k^\circ \ll 1$. By substituting these in (30), we achieve

$$\mathbb{E}[P_{\text{MU}}] \approx \frac{2}{\pi} \sum_{k=1}^{M} \frac{\gamma_k}{\theta_k^\circ} \, \mathbb{E}\left[\frac{1}{\tilde{Y}_k}\right] \left(1 + \frac{\phi_k}{\theta_k^\circ}\right). \tag{31}$$

The parameter ϕ_k in (31) is controlled by the direction codebook size $\ddot{N}_k$ as described by (16). The term $\mathbb{E}[1/\tilde{Y}_k]$ on the other hand is controlled by the magnitude codebook. The asymptotically optimal codebook that minimizes this term is the uniform (in dB) magnitude quantization codebook $\mathbb{Y}^\star$. By setting $\mathbb{Y} = \mathbb{Y}^\star$ and by using (10) and (16), we can bound the average sum power in (31) as follows:

$$\mathbb{E}\left[P_{\text{MU}}\right] < \frac{2\rho_{\text{MU,CSI}}}{\pi} \cdot \sum_{k=1}^{M} \frac{\gamma_k}{\theta_k^\circ} \left(1+\dot{N}_k^{-\zeta_k(\dot{N}_k)}+\omega\dot{N}_k^{-2\zeta_k(\dot{N}_k)}\right) \left(1+\frac{4\lambda_M}{\theta_k^\circ}\ddot{N}_k^{-\frac{1}{M-1}}\right) \tag{32}$$

where $\rho_{\text{MU,CSI}}$ is defined in (3) and the variable ω and the function $\zeta_k(\cdot)$ depend on the magnitude outage probability and the distribution of the channel magnitude as described in Section IV-A.

Our goal is to minimize the average sum-power upper bound in (32) in terms of the magnitude and direction quantization codebook parameters. The optimization constraints are as follows. Assuming a total number of feedback bits B, we have the following constraint on the codebook sizes:

$$\prod_{k=1}^{M} N_k = \prod_{k=1}^{M} \left(\dot{N}_k\ddot{N}_k + 1\right) = 2^B. \tag{33}$$

The two other constraints are the target outage probability constraints given by (19) and the MQCS conditions in (23). For the total feedback rate B, the target outage probabilities q_k, and the target SINR values γ_k, the product codebook optimization problem is therefore formulated as follows:

$$\min_{\substack{\dot{N}_k,\ddot{N}_k \\ y_k^{(1)},\theta_k^\circ}} \sum_{k=1}^{M} \frac{\gamma_k}{\theta_k^\circ} \left(1+\dot{N}_k^{-\zeta_k(\dot{N}_k)}+\omega\dot{N}_k^{-2\zeta_k(\dot{N}_k)}\right)\left(1+\frac{4\lambda_M}{\theta_k^\circ}\ddot{N}_k^{-\frac{1}{M-1}}\right) \tag{34}$$

$$\text{s.t.} \quad \prod_{k=1}^{M} \left(\dot{N}_k\ddot{N}_k + 1\right) = 2^B, \tag{35}$$

$$F(y_k^{(1)}) + \frac{2}{\pi}\theta_k^\circ \le q_k, \tag{36}$$

$$\ddot{N}_k \ge \left(4\lambda_M \Big/ \sin\left(\arctan\left(\frac{\sin\theta_k^\circ}{1+\sqrt{(M-1)\gamma_k}}\right)\right)\right)^{M-1}. \tag{37}$$

In order to obtain a closed-form solution for the optimal product structure, we simplify this problem as follows.

First, by assuming $\dot{N}_k, \ddot{N}_k \gg 1$ and using the fact that $\lim_{\dot{N}_k\to\infty} \zeta_k(\dot{N}_k) = 1$, we use the following approximation for the objective function:

$$\begin{aligned}\left(1 + \dot{N}_k^{-\zeta_k(\dot{N}_k)} + \omega\dot{N}_k^{-2\zeta_k(\dot{N}_k)}\right) \left(1 + \frac{4\lambda_M}{\theta_k^\circ}\ddot{N}_k^{-\frac{1}{M-1}}\right) &\approx \left(1 + \dot{N}_k^{-1}\right) \left(1 + \frac{4\lambda_M}{\theta_k^\circ}\ddot{N}_k^{-\frac{1}{M-1}}\right) \\ &\approx 1 + \dot{N}_k^{-1} + \frac{4\lambda_M}{\theta_k^\circ}\ddot{N}_k^{-\frac{1}{M-1}}. \end{aligned} \tag{38}$$

We therefore have the following approximate upper bound:

$$\mathbb{E}\left[P_{\text{MU}}\right] < \frac{2\rho_{\text{MU,CSI}}}{\pi} \sum_{k=1}^{M} \frac{\gamma_k}{\theta_k^\circ} \left(1+\dot{N}_k^{-1}+\frac{4\lambda_M}{\theta_k^\circ}\ddot{N}_k^{-\frac{1}{M-1}}\right). \tag{39}$$

Next, we simplify the optimization constraints as follows. We approximate the first constraint (35) as $\prod_{k=1}^{M} \dot{N}_k\ddot{N}_k = 2^B$. Regarding the outage constraint in (36), one can easily show that the objective function in (34) is a decreasing function of $y_k^{(1)}$ and θ_k°. The constraint in (36) should therefore be satisfied with equality at the optimum. In order to simplify this constraint, we make the assumption that the magnitude and direction outage probabilities are equally likely[3] :

$$\dot{q}_k = F(y_k^{(1)}) = \frac{q_k}{2} \tag{40}$$

$$\ddot{q}_k = \frac{2}{\pi}\theta_k^\circ = \frac{q_k}{2}. \tag{41}$$

According to this assumption, $y_k^{(1)} = F^{-1}(q_k/2)$ and

$$\theta_k^\circ = \frac{\pi}{4} q_k \tag{42}$$

are fixed and the codebook optimization is only over the codebook sizes. Finally, since the optimization is asymptotic in the codebook sizes $\dot{N}_k$ and $\ddot{N}_k$, the last constraint in (37) is redundant. This constraint however is used later to derive a lower bound on the total feedback rate B such that the target outage probabilities are feasible.

Now, by using the approximation in (38), the optimization problem in (34) simplifies to the following optimization problem:

$$\min_{\dot{N}_k,\ddot{N}_k} \sum_{k=1}^{M} \frac{\gamma_k}{\theta_k^\circ} \left(1 + \dot{N}_k^{-1} + \frac{4\lambda_M}{\theta_k^\circ}\ddot{N}_k^{-\frac{1}{M-1}}\right) \tag{43}$$

$$\text{s.t.} \quad \prod_{k=1}^{M} \dot{N}_k\ddot{N}_k = 2^B. \tag{44}$$

Define $\dot{B}_k \overset{\text{def}}{=} \log \dot{N}_k$ and $\ddot{B}_k \overset{\text{def}}{=} \log \ddot{N}_k$ as the number of magnitude and direction quantization bits respectively.

[3]It can be shown that any other division of the form $\dot{q}_k = \alpha q_k$ and $\ddot{q}_k = (1-\alpha)q_k$ with $0 < \alpha < 1$ only changes the bit allocations in Theorem 4 by a finite constant and therefore does not affect the asymptotic bit allocation results.

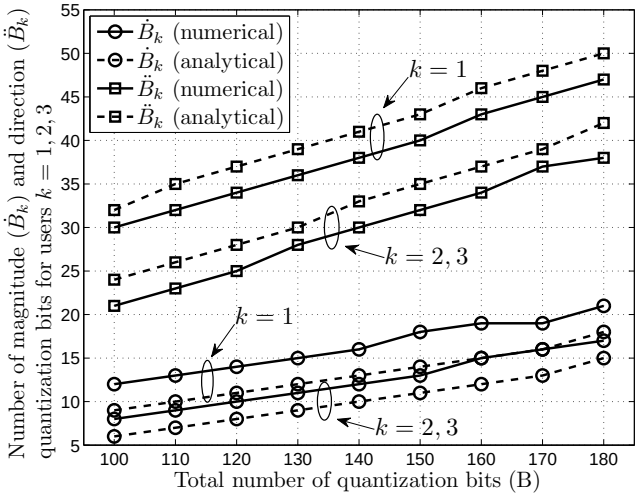


Fig. 3. Analytical magnitude and direction bit allocations in (47) and (48) vs. numerical bit allocations for three users with $\gamma_1 = 15$dB, $q_1 = 0.02$ and $\gamma_2 = \gamma_3 = 10$dB and $q_2 = q_3 = 0.05$.

Theorem 4: Define

$$\dot{B}_{\text{ave}} = \frac{1}{M^2}B - \frac{M-1}{M}\log\frac{1}{\bar{q}} - \kappa_{\text{MU}} \tag{45}$$

$$\ddot{B}_{\text{ave}} = \frac{M-1}{M^2}B + \frac{M-1}{M}\log\frac{1}{\bar{q}} + \kappa_{\text{MU}}, \tag{46}$$

where $\kappa_{\text{MU}} = \frac{M-1}{M}\log\frac{16\lambda_M}{\pi(M-1)}$ and $\bar{q} = \left(\prod_k q_k\right)^{1/M}$ is the geometric mean of the target outage probabilities. The optimal values of $\dot{B}_k$ and $\ddot{B}_k$ are given by

$$\dot{B}_k = \dot{B}_{\text{ave}} + \log\frac{\gamma_k}{\bar{\gamma}} + \log\frac{\bar{q}}{q_k} \tag{47}$$

$$\ddot{B}_k = \ddot{B}_{\text{ave}} + (M-1)\log\frac{\gamma_k}{\bar{\gamma}} + 2(M-1)\log\frac{\bar{q}}{q_k}, \tag{48}$$

where $\bar{\gamma} = \left(\prod_k \gamma_k\right)^{1/M}$ is the geometric mean of the target SINR values.

Proof: See Appendix D. ∎

Although the bit allocation laws in Theorem 4 are derived with the simplifying assumption of equal magnitude and direction outage probabilities in (40) and (41), the numerical results, as shown in Fig. 3, verify that the analytical results in (47) and (48) are close to the bit allocations derived by numerical minimization of (34). For the example in this figure, the base-station has $M = 3$ antennas and serves three users with the target parameters $\gamma_k = 15$dB and $q_k = 0.02$ for the first user ($k = 1$) and $\gamma_k = 10$dB and $q_k = 0.05$ for the two other users ($k = 2, 3$). The user channels are i.i.d. and $\mathbf{h}_k \sim \mathcal{N}(0, \mathbf{I}_M)$, where $\mathbf{I_M}$ is the $M \times M$ identity matrix. Also, the number of bits are rounded to the closest integer numbers[4].

Corollary 1: For each user k, the optimal number of magnitude and direction quantization bits are related as follows:

$$\ddot{B}_k = (M-1)\dot{B}_k + (M-1)\log\frac{1}{q_k} + M\kappa_{\text{MU}}, \tag{49}$$

[4]This is a popular approach in solving integer programming problems, where the integer conditions are relaxed and the optimum of the relaxed problem is rounded to the closed integer [40].

where κ_{MU} is defined in Theorem 4. Moreover the total number of quantization bits for user k is given by

$$B_k = \dot{B}_k + \ddot{B}_k = \frac{1}{M}B + M\log\frac{\gamma_k}{\bar{\gamma}} + (2M-1)\log\frac{\bar{q}}{q_k}. \tag{50}$$

As it is expected, if the users are homogenous in their requested target parameters, i.e. q_k and γ_k are the same for all users, each user takes an equal share of $\frac{1}{M}B$ of the total feedback rate. In the case of heterogenous users, on the other hand, a user with a higher QoS (lower target outage probability) and a higher target downlink rate (higher target SINR) uses a higher feedback rate B_k.

The bit allocation laws in Theorem 4 are asymptotic results in the feedback rate $B \to \infty$. In the following, in order to get a sense of how high the feedback rate should be, we determine a lower bound on B for which the target SINR values γ_k are in fact achievable with the target outage probabilities q_k.

Theorem 5: Assume $\gamma_k > 1$ and $q_k \ll 1$ and define

$$\mathcal{Q}_k \stackrel{\text{def}}{=} \frac{\sqrt{\gamma_k}}{q_k}.$$

For the target SINR's γ_k to be satisfied with outage probabilities q_k, the following total feedback rate B is sufficient:

$$B > \frac{1}{2}M^2\log\bar{\gamma} + (M^2-M)\log\frac{1}{\bar{q}} + M^2\log\Delta + b, \tag{51}$$

Here $\bar{\gamma}$ and $\bar{q}$ are the geometric means of γ_k's and q_k's respectively and

$$\Delta = \frac{\bar{\mathcal{Q}}}{\min\limits_{1\le k\le M}\mathcal{Q}_k},$$

where $\bar{\mathcal{Q}}$ is the geometric mean of $\mathcal{Q}_k$'s. The constant b in (51) is defined as $b = \frac{1}{2}M^2 + \frac{3}{2}M^2\log M + M^2\kappa_{\text{MU}}$, where κ_{MU} is defined in Theorem 4.

Proof: See Appendix E. ∎

Several interesting results can be extracted from Theorem 5. First, we observe that for the QoS constraints to remain feasible, the system feedback link capacity should scale logarithmically with the geometric mean of the target SINR values and the geometric mean of the inverse target outage probabilities. Second, if we compare the case of homogenous users with the case of heterogenous users, we see that heterogenous users impose an additional requirement, $M^2\log\Delta$, on the total feedback rate. If we think of $\mathcal{Q}_k$'s as users' QoS indicators, the variable Δ can be interpreted as a measure of discrepancy among users' QoS requirements. A higher QoS discrepancy requires a higher feedback bandwidth.

Finally, in order to study the performance of the limited-feedback system as the feedback rate increases, we substitute the optimal magnitude and direction codebook sizes given by Theorem 4 into the average sum-power upper bound in (39). The following theorem shows the scaling of the average sum power with the feedback rate B.

Theorem 6: For a limited-feedback system with a total number of B feedback bits, we have

$$\mathbb{E}\left[P_{\text{MU}}\right] < P_{\text{MU,CSI}}\left(1 + \frac{\sigma_{\text{MU}}}{\bar{q}}\cdot 2^{-\frac{B}{M^2}}\right), \tag{52}$$

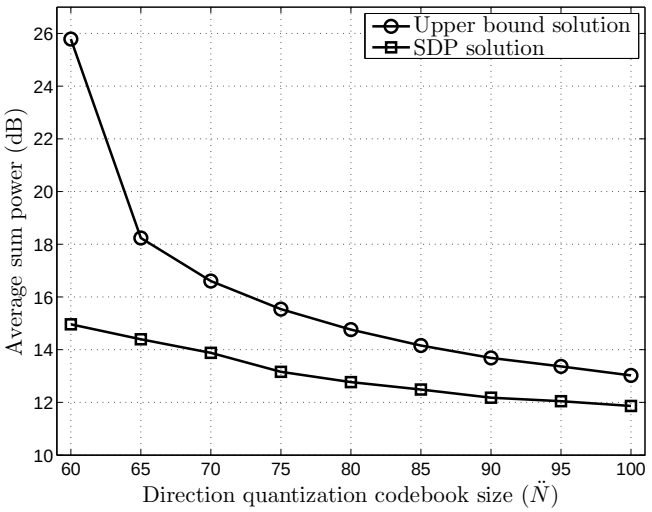


Fig. 4. SDP solution in (24) vs. the upper bound solution in (27) for three users with target SINR's of $\gamma_1 = 3$dB and $\gamma_2, \gamma_3 = 6$dB.

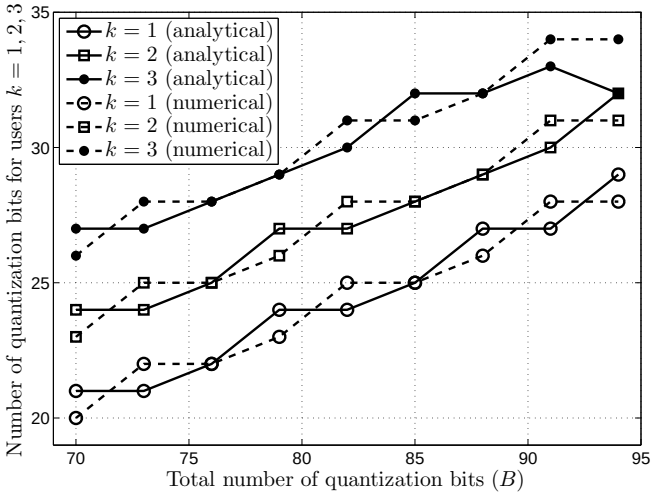


Fig. 5. Bit allocations B_k for three users with γ_1=2dB, γ_2=5dB, γ_3=8dB, and q_1=q_2=q_3=0.1.

where $P_{\text{MU,CSI}}$ is defined in (2), $\bar{q}$ is the geometric mean of the target outage probabilities, and

$$\sigma_{\text{MU}} = \frac{16M}{\pi(M-1)} \left(\frac{\pi^{3/2}(M-1)\Gamma((M+1)/2)}{16\Gamma(M/2)} \right)^{1/M}.$$

Proof: See Appendix F. ∎

If we define a quantization distortion measure as

$$D(B) = \frac{\mathbb{E}[P_{\text{MU}}] - P_{\text{MU,CSI}}}{P_{\text{MU,CSI}}}, \tag{53}$$

Theorem 6 implies that the distortion measure scales as $2^{-\frac{B}{M^2}}$ as $B \to \infty$.

VII. Numerical Validation

This section presents the numerical results that support and verify the analytical results in the earlier sections.

1) Upper Bound Approximation for The SDP Problem in (24)*:* The codebook optimizations in Section VI are based on the sum-power upper bound solution in (27) as an approximation of the solution to the SDP problem in (24). Here we investigate the accuracy of this approximation by comparing the two solutions for $M = 3$ users with target SINR's of $\gamma_1 = 3$dB and $\gamma_2, \gamma_3 = 6$dB.

To simplify the comparison, we assume perfect channel magnitude information, i.e. the quantized magnitude variables r_k in (24) and (27) are equal to the exact channel magnitudes. For direction quantization we use Grassmannian codebooks from [41]. The same codebook size is used for all users. The sum power values are averaged over 100 channel realizations for which the SDP problem is feasible. The user channels are i.i.d. and $\mathbf{h}_k \sim \mathcal{N}(0, \mathbf{I}_M)$.

Fig. 4 compares the two solutions as a function of the direction quantization codebook size $\ddot{N}$. As the figure shows, the two solutions converge as $\ddot{N}$ increases[5]. This justifies the use of the upper bound solution as the optimization objective in our high resolution analysis.

2) Bit Allocations and Distortion Scaling: Fig. 3 in Section VI compares the numerical and analytical bit allocations for $M = 3$ users with different target parameters. Here we repeat the process with a different set of parameters and record the share of each user B_k from the total number of feedback bits B. The target parameters are γ_1=2dB, γ_2=5dB, γ_3=8dB, and q_1=q_2=q_3=0.1. Channel models are similar to those used in Figs. 3 and 4 and the bit allocations are rounded to the closest integer numbers. As the figure verifies, users with higher target SINR's receive larger shares of the total feedback rate.

Finally, we investigate the system performance scaling with the number of feedback bits for the same set of parameters as in Fig. 5. For this purpose we use the average sum-power upper bound $\mathbb{E}[P_{\text{MU}}]$ in (32) and the definition of the distortion measure $D(B)$ in (53). Fig. 6 shows the distortion measure as a function of B when numerical and analytical bit allocations are utilized. As expected, the two bit allocations show close performances. The figure also shows the distortion upper bound in Theorem 6 for the purpose of comparison.

VIII. Concluding Remarks

We conclude this paper by comparing the asymptotic magnitude-direction bit allocation law for the multi-user system with that of the single-user system discussed in [37]. For the multi-user system and in the asymptotic regime where $B \to \infty$, the relation in (49) implies that the number of magnitude and direction quantization bits (for each user) are related as follows:

$$\ddot{B}_{\text{MU}} = (M-1)\dot{B}_{\text{MU}}. \tag{54}$$

The subscript MU in (54) stands for multi-user. For single-user systems, on the other hand, we have the following bit allocation law [37]:

$$\ddot{B}_{\text{SU}} = \frac{M-1}{2}\dot{B}_{\text{SU}}, \tag{55}$$

[5]This can made rigorous by showing that as $\ddot{N}_k \to \infty$ and $\phi_k \to 0$, the inequality in (25) is satisfied with an equality and therefore the upper bound solution is exact in the asymptotic high-resolution regime.

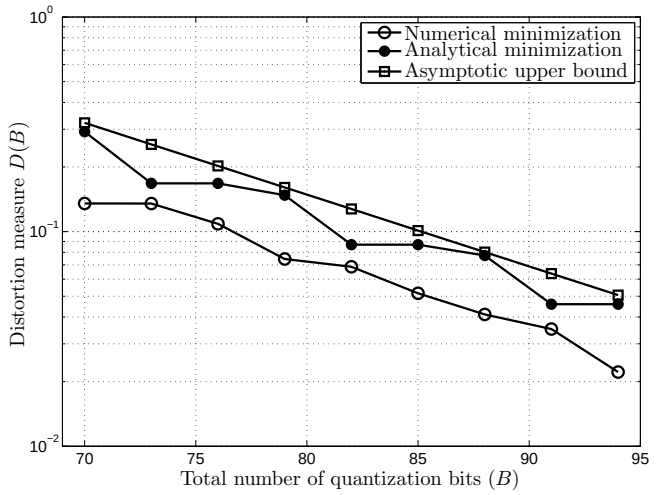


Fig. 6. Distortion measure with numerical and analytical bit allocations for three users with γ_1=2dB, γ_2=5dB, γ_3=8dB, and q_1=q_2=q_3=0.1.

with SU standing for single-user. If we define a relative quantization resolution as $\Upsilon = \ddot{B}/\dot{B}$, then we have

$$\Upsilon_{\text{MU}} = 2\ \Upsilon_{\text{SU}}, \tag{56}$$

which means that for the same number of magnitude quantization bits, the number of multi-user direction quantization bits is twice the number of single-user direction quantization bits. This is shown schematically in Fig. 7. As the figure implies, in order to make a single-user channel quantization codebook applicable to the multi-user system, each direction quantization region in the single-user codebook should be further quantized with the same resolution as the whole unit hypersphere.

As the final note we mention that the results in this paper are based on real channel space assumption mainly for the ease of geometric representation of the quantization regions and the corresponding sum-power calculations in Section V. However, the exact same approach introduced in this paper can be applied to complex space channels. The only main difference is in using the upper bound in (16) for real Grassmannian codebooks. For complex Grassmannian codebooks we need to use the following bound [16]:

$$\sin\phi_k < 2\ddot{N}_k^{-\frac{1}{2(M-1)}}, \tag{57}$$

By doing so, the rest of the analysis can be applied in a similar fashion to derive the bit allocation laws. In particular, the codebook optimization problem in (43) translates to the following problem for complex channels:

$$\min_{\dot{N}_k,\ddot{N}_k} \sum_{k=1}^{M} \frac{\gamma_k}{\theta_k^\circ}\left(1+\dot{N}_k^{-1}+\frac{2}{\theta_k^\circ}\ddot{N}_k^{-\frac{1}{2(M-1)}}\right)$$
$$\text{s.t.} \quad \prod_{k=1}^{M}\dot{N}_k\ddot{N}_k = 2^B.$$

By solving this problem, one can easily derive the bit allocation laws and the system performance scaling for the complex space similar to the ones in Theorems 4 and 6.

In particular, due to the difference between (16) and (57) in the exponents of the direction codebook sizes $\ddot{N}_k$, the asymptotic magnitude-direction bit allocation in the complex space turns out to be $\ddot{B} = 2(M-1)\dot{B}$ for the multi-user case and $\ddot{B} = (M-1)\dot{B}$ for the single-user case. Therefore, the complex-space quantization resolutions also satisfy the relation in (56).

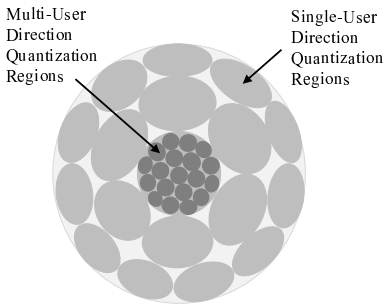


Fig. 7. Υ_{MU}=$2\Upsilon_{\text{SU}}$; single-user and multi-user direction quantization regions are shown as spherical caps on the unit hypersphere.

APPENDIX

A. Proof of Theorem 1

In order to prove the theorem, we use the following lemma:

Lemma 1: To ensure the feasibility of the power control problem in (21), it is sufficient to have

$$\frac{\tan\phi_k}{\sin\theta_k} < \frac{1}{1+\sqrt{(M-1)\gamma_k}}, \tag{58}$$

for all active users $k \in \mathcal{K}$.

Proof: The idea is to show that if the condition (58) holds, the upper bound solution in (27) is a valid solution to the power control problem in (21). For this purpose, it suffices to show that $\sum_{k\in\mathcal{K}}\alpha_k<1$, where α_k is defined in (27).

According to the condition (58), we have

$$\frac{\tan\phi_k}{\sin\theta_k} < \frac{1}{1+\sqrt{(M-1)\gamma_k}}.$$

Then

$$\frac{\sin(\theta_k-\phi_k)}{\sin\phi_k} = \frac{\sin\theta_k}{\tan\phi_k} - \cos\theta_k > \frac{\sin\theta_k}{\tan\phi_k} - 1 > \sqrt{(M-1)\gamma_k}. \tag{59}$$

Therefore

$$\alpha_k = \left(1+\frac{\sin^2(\theta_k-\phi_k)}{\gamma_k\sin^2\phi_k}\right)^{-1} < \frac{1}{M}, \tag{60}$$

and $\sum_{k\in\mathcal{K}}\alpha_k < 1$, since $|\mathcal{K}| \le M$. ■

The MQCS condition in (23) is equivalent to

$$4\lambda_M\ddot{N}_k^{-\frac{1}{M-1}} < \sin\left(\arctan\left(\frac{\sin\theta_k^\circ}{1+\sqrt{(M-1)\gamma_k}}\right)\right). \tag{61}$$

Combining this with the inequality in (16), we have

$$\sin\phi_k < \sin\left(\arctan\left(\frac{\sin\theta_k^\circ}{1+\sqrt{(M-1)\gamma_k}}\right)\right), \tag{62}$$

which leads to

$$\tan\phi_k < \frac{\sin\theta_k^\circ}{1+\sqrt{(M-1)\gamma_k}}. \tag{63}$$

Now for all active users $k \in \mathcal{K}$, we have $\theta_k \geq \theta_k^\circ$, since θ_k° is the smallest acceptable angle between user k's channel and the subspace spanned by all other users' channels. Combining $\theta_k \geq \theta_k^\circ$ with (63) leads to

$$\tan\phi_k < \frac{\sin\theta_k}{1+\sqrt{(M-1)\gamma_k}},$$

which according to Lemma 1 guarantees feasibility and therefore completes the proof.

B. Proof of Theorem 2

Consider the definition of the uncertainty region

$$S_k = \left\{ \mathbf{w} \in \mathbb{R}^M \middle| \sqrt{r} \leq \|\mathbf{w}\| < \sqrt{R},\ \angle(\mathbf{w}, \tilde{\mathbf{u}}) < \phi \right\}.$$

As far as the constraint in (22) in concerned, the constraint $\|\mathbf{w}\| < \sqrt{R}$ is redundant since the infimum occurs on the lower surface $\|\mathbf{w}\| = \sqrt{r}$ (see the process of deriving the upper bound solution in (25)).

Now define

$$T_0 = \left\{ \mathbf{w} \left| \frac{P_k \left|\mathbf{w}^T \mathbf{v}_k\right|^2}{\sum\limits_{l\neq k} P_l \left|\mathbf{w}^T \mathbf{v}_l\right|^2 + 1} > \gamma_k \right. \right\}, \tag{64}$$

$$T_1 = \left\{ \mathbf{w} \middle| \|\mathbf{w}\| \geq \sqrt{r} \right\}, \tag{65}$$

$$T_2 = \left\{ \mathbf{w} \middle| |\mathbf{w}^T \tilde{\mathbf{u}}_k| > \|\mathbf{w}\| \cos\phi \right\}, \tag{66}$$

where the condition in (66) is equivalent to $\angle(\mathbf{w}, \tilde{\mathbf{u}}) < \phi$. With these definitions, the constraint in (22) is equivalent to

$$T_1 \cap T_2 \ \subset \ T_0. \tag{67}$$

Now, define scalars τ_n and symmetric matrices A_n, for $n = 0, 1, 2$, as follows:

$$\begin{aligned} A_0 &= \sum_{l\neq k} P_l \mathbf{v}_l \mathbf{v}_l^T - \frac{1}{\gamma_k} P_k \mathbf{v}_k \mathbf{v}_k^T \\ A_1 &= -\mathbf{I} \\ A_2 &= -\frac{1}{\cos^2\phi} \tilde{\mathbf{u}}_k \tilde{\mathbf{u}}_k^T + \mathbf{I} \\ \tau_0 &= -1, \quad \tau_1 = -r, \quad \tau_2 = 0. \end{aligned}$$

With these definitions, the sets T_0, T_1, T_2 can be expressed as sublevels of quadratic functions:

$$T_n = \left\{ \mathbf{w} \middle| \mathbf{w}^T A_n \mathbf{w} \leq \tau_n \right\}, \quad n = 0, 1, 2.$$

We therefore can use the following theorem from [39] to replace the condition (67) with a SDP condition. This theorem is an important extension of what is known as S-procedure in the optimization literature [40].

Theorem 7: Let $M \geq 3$ and $A_n \in \mathbb{R}^{M\times M}$ be symmetric matrices for $n = 0, 1, 2$ and assume

$$\exists\ \nu_1, \nu_2 \in \mathbb{R} \ \text{ s.t. } \ \nu_1 A_1 + \nu_2 A_2 \succ 0.$$

Define the quadratic functions $f_n(\mathbf{w}) = \mathbf{w}^T A_n \mathbf{w}$. Then the following two statements are equivalent:

$$\text{I.} \quad f_1(\mathbf{w}) \leq \tau_1\ ,\ f_2(\mathbf{w}) \leq \tau_2 \quad \Rightarrow \quad f_0(\mathbf{w}) \leq \tau_0 \tag{68}$$

$$\text{II.} \quad \exists\ \lambda > 0,\ \mu > 0 \quad \text{s.t.} \quad \begin{cases} A_0 \preceq \lambda A_1 + \mu A_2 \\ \tau_0 \geq \lambda \tau_1 + \mu \tau_2 \end{cases} \tag{69}$$

For the problem in hand, one can easily find ν_1 and ν_2 such that the condition $\nu_1 A_1 + \nu_2 A_2 \succ 0$ is satisfied; therefore, the constraint in (67) translates to the SDP constraints in (24).

C. Proof of Theorem 3

From the definitions of α_k and β_k in (27), we have

$$\begin{aligned} \alpha_k &= \frac{\gamma_k \sin^2\phi_k}{\gamma_k \sin^2\phi_k + \sin^2(\theta_k - \phi_k)}, \\ \frac{\alpha_k}{\beta_k} &= \frac{\gamma_k / r_k}{\gamma_k \sin^2\phi_k + \sin^2(\theta_k - \phi_k)}. \end{aligned}$$

For small values of $\phi_k \ll 1$, it is easy to verify that

$$\begin{aligned} &\gamma_k \sin^2\phi_k = \gamma_k \phi_k^2 + o(\phi_k^2), \\ &\sin^2(\theta_k - \phi_k) = \sin^2\theta_k - (\sin 2\theta_k)\phi_k + o(\phi_k). \end{aligned}$$

After a few manipulations, one can show that

$$\begin{aligned} \alpha_k &= e_k r_k \phi_k^2 + o(\phi_k^2), \\ \frac{\alpha_k}{\beta_k} &= e_k + f_k \phi_k + o(\phi_k), \end{aligned}$$

and

$$P_{\text{MU}} = \frac{\sum_{k\in\mathcal{K}} \alpha_k/\beta_k}{1 - \sum_{k\in\mathcal{K}} \alpha_k} = \sum_{k\in\mathcal{K}} e_k + f_k\phi_k + o(\phi_k).$$

D. Proof of Theorem 4

By applying Lagrange multipliers method, we achieve

$$\dot{N}_k = \frac{1}{\Lambda} \cdot \frac{\gamma_k}{\theta_k^\circ} \tag{70}$$

$$\ddot{N}_k = \frac{1}{\Lambda^{M-1}} \cdot \left(\frac{4\lambda_M}{M-1} \cdot \frac{\gamma_k}{(\theta_k^\circ)^2} \right)^{M-1}, \tag{71}$$

where Λ is a Lagrange multiplier that should satisfy the constraint $\prod_{k=1}^M \dot{N}_k \ddot{N}_k = N$. By solving for Λ we get

$$\Lambda = \left(\frac{4\lambda_M}{M-1} \right)^{\frac{M-1}{M}} \left(\prod_{k=1}^M \gamma_k \right)^{\frac{1}{M}} \left(\prod_{k=1}^M \frac{1}{\theta_k^\circ} \right)^{\frac{2M-1}{M^2}} \cdot N^{-\frac{1}{M^2}}. \tag{72}$$

By substituting (72) in (70) and (71), and using $\theta_k^\circ = \frac{\pi}{4} q_k$ as in (42), and further manipulation, the optimal quantization resolutions $\dot{B}_k = \log \dot{N}_k$ and $\ddot{B}_k = \log \ddot{N}_k$ can be expressed as in (47) and (48).

E. Proof of Theorem 5

For the target parameters to be feasible, the optimal direction codebook sizes $\ddot{N}_k = 2^{\ddot{B}_k}$ are required to satisfy the MQCS conditions in (37). With the assumption of $q_k \ll 1$, we have $\sin\theta_k^\circ \approx \theta_k^\circ \ll 1$ and the MQCS conditions simplify to the following conditions:

$$\ddot{N}_k \geq \left(\frac{4\lambda_M}{\theta_k^\circ} \left(1 + \sqrt{(M-1)\gamma_k} \right) \right)^{M-1}. \tag{73}$$

In the following, we find a lower bound on B that guarantees the conditions in (73).

For $M \geq 2$ and $\gamma_k > 1$, we have the following inequality:

$$1 + \sqrt{(M-1)\gamma_k} < \sqrt{2M\gamma_k}. \tag{74}$$

To satisfy the conditions in (73), therefore, it is sufficient to satisfy

$$\ddot{N}_k \geq \left(\frac{4\lambda_M}{\theta_k^\circ}\sqrt{2M\gamma_k}\right)^{M-1}. \tag{75}$$

By substituting $\ddot{N}_k$ from (71) in (75), we achieve

$$\frac{1}{\Lambda} \geq (M-1)\frac{\theta_k^\circ}{\gamma_k}\sqrt{2M\gamma_k}. \tag{76}$$

Therefore it is sufficient to have

$$\frac{1}{\Lambda} \geq \sqrt{2}M^{3/2}\frac{\theta_k^\circ}{\sqrt{\gamma_k}}. \tag{77}$$

By substituting the expression for Λ in (72) into (77), we achieve the following constraints on $N = 2^B$ for $1 \leq k \leq M$:

$$N \geq C \cdot \left(\bar{\gamma} \cdot \bar{\theta^\circ}^{-\frac{2M-1}{M}} \cdot \frac{\theta_k^\circ}{\sqrt{\gamma_k}}\right)^{M^2} \tag{78}$$

where $\bar{\gamma} = \left(\prod_k \gamma_k\right)^{1/M}$ and $\bar{\theta^\circ} = \left(\prod_k \theta_k^\circ\right)^{1/M}$ and

$$C = \left(\sqrt{2}M^{3/2}\right)^{M^2}\left(\frac{4\lambda_M}{M-1}\right)^{M(M-1)}.$$

Since (78) is to be satisfied for all $1 \leq k \leq M$, we have the following sufficient bound on $N = 2^B$:

$$\begin{aligned} N &\geq C \cdot \left(\bar{\gamma} \cdot \bar{\theta^\circ}^{-\frac{2M-1}{M}} \cdot \max_{1\leq k\leq M}\frac{\theta_k^\circ}{\sqrt{\gamma_k}}\right)^{M^2} \\ &= C \cdot \left(\sqrt{\bar{\gamma}} \cdot \bar{\theta^\circ}^{-\frac{M-1}{M}} \cdot \max_{1\leq k\leq M}\frac{\theta_k^\circ/\sqrt{\gamma_k}}{\bar{\theta^\circ}/\sqrt{\bar{\gamma}}}\right)^{M^2}. \end{aligned} \tag{79}$$

By substituting $\theta_k^\circ = \frac{\pi}{4}q_k$ and $\bar{\theta^\circ} = \frac{\pi}{4}\bar{q}$ in (79) and taking the logarithm of the both sides we achieve the lower bound in (51), which completes the proof.

F. Proof of Theorem 6

By substituting the optimal values of $\dot{N}_k$ and $\ddot{N}_k$ given by (70) and (71) into the average sum-power upper bound in (32), we have

$$\begin{aligned} \mathbb{E}\left[P_{\text{MU}}\right] &< \frac{2\rho_{\text{MU,CSI}}}{\pi}\sum_{k=1}^{M}\frac{\gamma_k}{\theta_k^\circ}\left(1 + \dot{N}_k^{-1} + \frac{4\lambda_M}{\theta_k^\circ}\ddot{N}_k^{-\frac{1}{M-1}}\right) \\ &= \frac{2\rho_{\text{MU,CSI}}}{\pi}\sum_{k=1}^{M}\left(\frac{\gamma_k}{\theta_k^\circ} + \Lambda + (M-1)\Lambda\right) \\ &= \frac{2\rho_{\text{MU,CSI}}}{\pi}\left(\left[\sum_{k=1}^{M}\frac{\gamma_k}{\theta_k^\circ}\right] + M^2\Lambda\right) \\ &\overset{\text{(a)}}{=} \frac{2\rho_{\text{MU,CSI}}}{\pi}\left(\left[\sum_{k=1}^{M}\frac{\gamma_k}{\theta_k^\circ}\right] + M^2\chi\bar{\gamma}\left(\frac{1}{\bar{\theta^\circ}}\right)^{2-\frac{1}{M}}2^{-\frac{B}{M^2}}\right), \end{aligned} \tag{80}$$

where $\bar{\gamma} = \left(\prod_k \gamma_k\right)^{1/M}$ and $\bar{\theta^\circ} = \left(\prod_k \theta_k^\circ\right)^{1/M}$. For the equality (a) we have used the expression (72) for Λ with $N = 2^B$ and $\chi = \left(4\lambda_M/(M-1)\right)^{(M-1)/M}$.

By further manipulating (80), we have

$$\begin{aligned} &\mathbb{E}\left[P_{\text{MU}}\right] \\ &< \frac{2\rho_{\text{MU,CSI}}}{\pi}\left(\left[\sum_{k=1}^{M}\frac{\gamma_k}{\theta_k^\circ}\right] + M\chi\left[M\frac{\bar{\gamma}}{\bar{\theta^\circ}}\right]\left(\frac{1}{\bar{\theta^\circ}}\right)^{1-\frac{1}{M}}2^{-\frac{B}{M^2}}\right) \\ &\overset{\text{(b)}}{<} \frac{2\rho_{\text{MU,CSI}}}{\pi}\left(\left[\sum_{k=1}^{M}\frac{\gamma_k}{\theta_k^\circ}\right] + M\chi\left[\sum_{k=1}^{M}\frac{\gamma_k}{\theta_k^\circ}\right]\left(\frac{1}{\bar{\theta^\circ}}\right)^{1-\frac{1}{M}}2^{-\frac{B}{M^2}}\right) \\ &= \frac{2\rho_{\text{MU,CSI}}}{\pi}\left[\sum_{k=1}^{M}\frac{\gamma_k}{\theta_k^\circ}\right]\left(1 + \frac{M\chi}{\bar{\theta^\circ}}\bar{\theta^\circ}^{\frac{1}{M}}2^{-\frac{B}{M^2}}\right) \\ &\overset{\text{(c)}}{<} P_{\text{MU,CSI}}\left(1 + \frac{M\chi}{\bar{\theta^\circ}}2^{-\frac{B}{M^2}}\right), \end{aligned} \tag{81}$$

where in (b) we use the fact the geometric mean is smaller than arithmetic mean:

$$\frac{\bar{\gamma}}{\bar{\theta^\circ}} = \left(\prod_{k=1}^{M}\frac{\gamma_k}{\theta_k}\right)^{\frac{1}{M}} \leq \frac{1}{M}\sum_{k=1}^{M}\frac{\gamma_k}{\theta_k},$$

and in deriving (c), we use the definition of $P_{\text{MU,CSI}}$ in (2) and the fact that $\bar{\theta^\circ} = \frac{\pi}{4}\bar{q} < 1$, since $\bar{q}$ is a probability measure.

By substituting $\bar{\theta^\circ} = \frac{\pi}{4}\bar{q}$ in (81), we achieve the bound in (52) and the proof is complete.

REFERENCES

[1] M. Sharif and B. Hassibi, "On the capacity of MIMO broadcast channels with partial side information," *IEEE Trans. Inform. Theory*, vol. 51, no. 2, pp. 506-522, Feb. 2005.

[2] N. Jindal, "MIMO broadcast channels with finite rate feedback," *IEEE Trans. Inform. Theory*, vol. 52, no. 11, pp. 5045-5059, Nov. 2006.

[3] T. Yoo, N. Jindal, and A. Goldsmith, "Multi-antenna downlink channels with limited feedback and user selection," *IEEE J. Select. Areas Commun.*, vol. 25, no. 7, pp. 1478-1491, Sep. 2007.

[4] K. Huang, R. W. Heath, and J. G. Andrews, "Space division multiple access with a sum feedback rate constraint," *IEEE Trans. Sig. Proc.*, vol. 55, no. 7, pp. 3879-3891, Jul. 2007.

[5] G. Caire, N. Jindal, M. Kobayashi, and N. Ravindran, "Multiuser MIMO achievable rates with downlink training and channel state feedback," *IEEE Trans. Inform. Theory*, to appear.

[6] P. Ding, D. Love, and M. Zoltowski, "Multiple antenna broadcast channels with shape feedback and limited feedback," *IEEE Trans. Signal Process.*, vol. 55, no. 7, pp. 3417-3428, July 2007.

[7] M. Kountouris, R. de Francisco, D. Gesbert, D. Slock, and T. Salzer, "Multiuser diversity-multiplexing tradeoff in MIMO broadcast channels with limited feedback," in *Proc. IEEE Asilomar Conf. on Signals, Systems, and Comp. (ASILOMAR)*, Pacific Grove, CA, Oct.-Nov. 2006, pp. 364-368.

[8] B. Khoshnevis and W. Yu, "Structure of channel quantization codebook for multi-user spatial multiplexing systems," in *Proc. IEEE International Conference on Communications (ICC)*, Cape Town, South Africa, May 2010.

[9] B. Khoshnevis and W. Yu, "High resolution channel quantization rules for multi-user spatial multiplexing systems," in *Proc. Conference on Information Sciences and Systems (CISS)*, Princeton, NJ, Mar. 2010.

[10] K. Huang, R. Heath, and J. Andrews, "Performance of orthogonal beamforming for SDMA with limited feedback," *IEEE Trans. on Veh. Technol.*, vol. 58, no. 1, pp. 152-164, Jan. 2009.

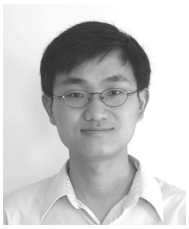
[11] R. Bhagavatula and R. Heath, "Adaptive limited feedback for sum-rate maximizing beamforming in cooperative multicell systems," *IEEE Trans. Signal Process.*, vol. 59, no. 2, pp. 800-811, Feb. 2011..

[12] R. Bhagavatula, R. Heath, and B. Rao, "Limited feedback with joint CSI quantization for multicell cooperative generalized eigenvector beamforming," in *Proc. IEEE Int. Conf. Acoust., Speech and Sig. Proc. (ICASSP)*, Dallas, TX, Mar. 2010, pp. 2838-2841.

[13] S. Han, C. Yang, M. Bengtsson, and A. Perez-Neira, "Channel norm based user scheduler in coordinated multi-point systems," in *Proc. IEEE Glob. Telecom. Conf. (GLOBECOM)*, Honolulu, HI, Nov.-Dec. 2009.

[14] A. Gersho and R. M. Gray, *Vector Quantization and Signal Compression*. Boston, MA: Kluwer, 1992.
[15] J. Zheng, E. Duni, and B. Rao, "Analysis of multiple-antenna systems with finite-rate feedback using high-resolution quantization theory," *IEEE Trans. Signal Process.*, vol. 55, no. 4, pp. 1461-1476, Apr. 2007.
[16] D. Love and R. Heath, "Grassmannian beamforming for multiple-input multiple-output wireless systems," *IEEE Trans. Inform. Theory*, vol. 49, no. 10, pp. 2735-2747, Oct. 2003.
[17] B. Khoshnevis and W. Yu, "Joint power control and beamforming codebook design for MISO channels with limited feedback," in *Proc. Global Telecommun. Conf. (GLOBECOM)*, Honolulu, HI, Nov.-Dec. 2009.
[18] N. Jindal, "A high SNR analysis of MIMO broadcast channels," in *Proc. IEEE Int. Symp. Information Theory (ISIT)*, Adelaide, Australia, Sep. 2005, pp. 2310-2314.
[19] T. Yoo and A. Goldsmith, "On the optimality of multiantenna broadcast scheduling using zero-forcing beamforming," *IEEE J. Select. Areas Commun.*, vol. 24, no. 3, pp. 528-541, Mar. 2006.
[20] W. Geyi, "Multi-antenna information theory," *Progress In Electromagnetics Research*, vol. 75, pp. 11-50, 2007.
[21] K. Pun, T. Nguyen, "Widely linear filter bank equalizer for real STBC", *IEEE Trans. on Signal Processing*, vol. 56, no. 9, pp. 4544-4548, Sep. 2008.
[22] M. Schubert and H. Boche, "Solution of the multi-user downlink beamforming problem with individual SINR constraints," *IEEE Trans. Veh. Technol.*, vol. 53, pp. 18-28, Jan. 2004.
[23] F. Rashid-Farrokhi, L. Tassiulas, and K. Liu, "Joint optimal power control and beamforming in wireless networks using antenna arrays," *IEEE Trans. Commun.*, vol. 46, no. 10, pp. 1313-1324, Oct. 1998.
[24] A. Wiesel, Y. Eldar, and S. Shamai, "Linear precoding via conic optimization for fixed MIMO receivers," *IEEE Trans. Signal Processing*, vol. 54, no. 1, pp. 161-176, Jan. 2006.
[25] V. Sharma and S. Lambotharan, "Multiuser downlink mimo beamforming using an iterative optimization approach," in *Proc. IEEE Vehicular Technology Conference (VTC)*, Montreal, Canada, Sep. 2006.
[26] A. Tarighat, M. Sadek, and A. Sayed, "A multi user beamforming scheme for downlink MIMO channels based on maximizing signal-to-leakage ratios," in *Proc. IEEE Int. Conf. Acoust., Speech and Sig. Proc. (ICASSP)*, Philadelphia, PA, Mar. 2005, pp. 1129-1132.
[27] H. Holma, and A. Toskala, *WCDMA for UMTS: HSPA Evolution and LTE*. John Wiley & Sons, 2007.
[28] *IEEE 802.16e Amendment: Physical and Medium Access Control Layers for Combined Fixed and Mobile Operation in Licensed Bands*, IEEE Standard 802.16, 2005.
[29] *3GPP, Technical specification group radio access network; (E-UTRA) and (E-UTRAN); overall description; stage 2*, Sep. 2008.
[30] M. Schubert and H. Boche, "QoS-based resource allocation and transceiver optimization," *Foundations and Trends in Communication and Information Theory*, vol. 2, no. 6, pp. 383-529, 2005.
[31] T. Girici, C. Zhu, J. Agre, and A. Ephremides, "Proportional fair scheduling algorithm in OFDMA-based wireless systems with QoS constraints," *Journal of Communications and Networks*, vol. 12, no. 1, pp. 30-42, Feb. 2010.
[32] E. Björnson, M. Bengtsson, and B. Ottersten, "Optimality properties and low-complexity solutions to coordinated multicell transmission," in *Proc. Global Telecommun. Conf. (GLOBECOM)*, Miami, FL, Dec. 2010.
[33] W. Yu, T. Kwon, and C. Shin, "Multicell coordination via joint scheduling, beamforming and power spectrum adaptation," in *Proc. IEEE Conference on Computer Communications (INFOCOM)*, Shanghai, China, Apr. 2011, pp. 2570-2578.
[34] M. Payaró, A. Pascual-Iserte, and M. A. Lagunas, "Robust power allocation designs for multi-user and multiantenna downlink communication systems through convex optimization," *IEEE J. Sel. Areas Commun.*, vol. 25, pp. 1390-1401, Sep. 2007.
[35] M. Botros Shenouda and T. N. Davidson, "Convex conic formulations of robust downlink precoder designs with quality of service constraints," *IEEE J. Select. Topics Signal Processing*, vol. 1, pp. 714-724, Dec. 2007.
[36] N. Vucic and H. Boche, "Downlink precoding for multi-user MISO systems with imperfect channel knowledge," in *Proc. IEEE Int. Conf. Acoust., Speech and Sig. Proc. (ICASSP)*, Las Vegas, NV, Mar.-Apr. 2008, pp. 3121-3124.
[37] B. Khoshnevis and Wei Yu, "Bit allocation law for multi-antenna channel feedback quantization: single-user case," *IEEE Trans. Signal Processing*, vol. 59, no. 5, pp. 2270-2283, May 2011.
[38] A. Barg and D. Y. Nogin, "Bounds on packings of spheres in the Grassmann manifold," *IEEE Trans. Inform. Theory*, vol. 48, no. 9, pp. 24502454, Sep. 2002.
[39] B. T. Polyak, "Convexity of quadratic transformations and its use in control and optimization," *J. Optim. Theory Appl.*, vol. 99, no. 3, pp. 553-583, Dec. 1998.
[40] S. Boyd and L. Vandenberghe, *Convex Optimization*. Cambridge University Press, 2004.
[41] N. Sloane. Packings in Grassmannian spaces. [Online] Available: http://www2.research.att.com/ njas/grass/grassTab.html

Behrouz Khoshnevis (S'07-M'10) received B.Sc. and M.Sc. degrees in Electrical Engineering from Sharif University of Technology, Tehran, Iran, in 2003 and 2005 respectively, and Ph.D. degree in Electrical and Computer Engineering from University of Toronto, Toronto, Ontario, Canada, in 2010. He is the recipient of Connaught Scholarship, Edward Rogers Sr. Graduate Scholarship, and Shahid U.H. Qureshi Memorial Prize in recognition of academic excellence from University of Toronto. He is currently a post-doctoral research fellow at University of Toronto. His research interests include MIMO communications, signal processing, information theory, and optimization theory. Dr. Behrouz Khoshnevis is a member of IEEE Communications Society.



Wei Yu (S'97-M'02-SM'08) received the B.A.Sc. degree in Computer Engineering and Mathematics from the University of Waterloo, Waterloo, Ontario, Canada in 1997 and M.S. and Ph.D. degrees in Electrical Engineering from Stanford University, Stanford, CA, in 1998 and 2002, respectively. Since 2002, he has been with the Electrical and Computer Engineering Department at the University of Toronto, Toronto, Ontario, Canada, where he is now an Associate Professor and holds a Canada Research Chair in Information Theory and Digital Communications. His main research interests include multiuser information theory, optimization, wireless communications and broadband access networks.

Prof. Wei Yu is currently an Editor for IEEE TRANSACTIONS ON INFORMATION THEORY and for IEEE TRANSACTIONS ON COMMUNICATIONS. He was an Editor for IEEE TRANSACTIONS ON WIRELESS COMMUNICATIONS from 2004 to 2007, and a Guest Editor for a number of special issues for the IEEE JOURNAL ON SELECTED AREAS IN COMMUNICATIONS and the EURASIP JOURNAL ON APPLIED SIGNAL PROCESSING. He is member of the Signal Processing for Communications and Networking Technical Committee of the IEEE Signal Processing Society. He received the IEEE Signal Processing Society Best Paper Award in 2008, the McCharles Prize for Early Career Research Distinction in 2008, the Early Career Teaching Award from the Faculty of Applied Science and Engineering, University of Toronto in 2007, and the Early Researcher Award from Ontario in 2006.