

# Outage-Optimized Distributed Quantization of Multicast Beamforming Vectors

Erdem Koyuncu, *Member, IEEE*, Christian Remling,

Xiaoyi (Leo) Liu, *Student Member, IEEE*, and

Hamid Jafarkhani, *Fellow, IEEE*

## Abstract

We consider a slow-fading multicast channel with one  $T$ -antenna transmitter and  $K$  single-antenna receivers with the goal of minimizing channel outage probability using quantized beamforming. Our focus is on a distributed limited feedback scenario where each receiver can only quantize and send feedback information regarding its own receiving channels.

A classical result in point-to-point quantized beamforming is that a necessary and sufficient condition for full diversity is to have  $\lceil \log_2 T \rceil$  bits from the receiver. We first generalize this result to multicast beamforming systems and show that a necessary and sufficient condition to achieve full diversity for all receivers is to have  $\lceil \log_2 T \rceil$  bits from *each* receiver. Also, for a two-receiver system and with  $R$  feedback bits per receiver, we show that the outage performance with quantized beamforming is within  $O(2^{-\frac{R}{32T^2}})$  dBs to the performance with full channel state information at the transmitter (CSIT). This constitutes, in the context of multicast channels, the first example of a distributed limited feedback scheme whose performance can provably approach the performance with full CSIT. Numerical simulations confirm our analytical findings.

Erdem Koyuncu, Xiaoyi (Leo) Liu, and Hamid Jafarkhani are with the Center for Pervasive Communications and Computing, University of California, Irvine, Irvine CA 92697-2625 USA. E-mails: {ekoyuncu, xiaoyil3, hamidj}@uci.edu. Christian Remling is with the Department of Mathematics, University of Oklahoma, Norman OK 73019-3103 USA. E-mail: cremling@math.ou.edu.

Part of this work has been submitted for possible presentation at the IEEE International Symposium on Information Theory (ISIT) in 2016. This conference submission is available online at <https://webfiles.uci.edu/ekoyuncu/dbc.pdf>

This work was supported in part by the NSF Award CCF-1218771.

## I. INTRODUCTION

Multicasting refers to the transmission of common information to several physically-separated receivers. In the context of physical layer, a particularly well-investigated scenario is the multiple-input single-output (MISO) multicast channel, where a  $T$ -antenna transmitter wishes to communicate to  $K$  single-antenna receivers over fading channels [1]–[6]. In such a scenario, when channel state information (CSI) is available to the transmitter, one can maximize the overall performance (e.g., the ergodic capacity, or the outage probability) using beamforming or precoding. The capacity limits of MISO multicast channels with such CSI-adaptive transmission strategies have first been investigated in [1], where several scaling results have been derived with different assumptions on  $T$  and  $K$ . Other work on the capacity of multicast channels in the large system limit have studied the case of correlated channels [2], and the performance of antenna subset selection [3].

Unlike a point-to-point MISO system where the optimal transmitter covariance matrix is simply a beamformer along the channel direction, closed-form expressions for the optimal covariance matrices or beamforming vectors are not known for a general MISO multicast system. A significant amount of work thus also exists [4]–[6] on the numerical optimization of multicast covariance matrices and beamforming vectors, with closed-form optimal solutions being available for certain values of  $K$  and  $T$  [5]. In particular, it is known that beamforming is optimal for  $K \leq 3$ , or close to optimal when  $T$  is much larger than  $K$  [1], [5], [6].

Most of these previous studies assume that the transmitter has perfect knowledge of the CSI. In fact, CSI at the transmitter (CSIT) can be acquired through feedback from the receivers, each of which can acquire the knowledge of their own receiving channels through transmitter training sequences. On the other hand, since the CSI can assume any value in a multi-dimensional complex space, the assumption of perfect CSIT requires an “infinite number of feedback bits” from every receiver. In practice, each receiver can communicate only a finite number of bits per channel state as feedback information. A mathematical formulation of such a limited feedback scenario leads to a distributed quantization problem where each receiver quantizes only a part of the entire CSI.

A special case is a point-to-point MISO system with  $K = 1$ , where the distributed quantization problem boils down to a simple point-to-point quantization problem and several solutions are available [7]–[11]. However, very little work exists on the design of limited

feedback schemes when  $K > 1$ . In [12], the authors study a scenario where only the channel direction information is quantized with channel magnitude information still being perfectly available at the transmitter (this would again require infinitely many receiver feedback bits.). In [13], the performance of centralized (non-distributed) variable-length quantizers have been analyzed. To the best of our knowledge, the first “true” distributed quantizers for beamforming in multicast channels have been proposed in [14]. Although the quantizers in [14] can provide full diversity, they do not do so in a rate-optimal manner, and cannot provably approach the performance with full CSIT.

In this work, we consider outage-optimized distributed quantization of beamforming vectors for a MISO multicast channel. We show how to construct distributed quantizers that can achieve full diversity in a rate-optimal manner. For a two-receiver system, we also show how to design distributed quantizers that can approach the performance with full CSIT. Numerical simulations suggest that a similar result holds for more than two receivers, but a formal proof remains elusive.

The rest of this paper is organized as follows. In Section II, we introduce the system model and the distributed quantizers. In Section III, we construct rate-optimal distributed quantizers that can provide full diversity. In Section IV, we design quantizers that can approach the performance with full CSIT. Finally, in Section V, we draw our main conclusions.

*Notation:*  $\mathbb{C}^{m \times n}$  is the set of all  $m \times n$  complex matrices with  $\mathbb{C}^T \triangleq \mathbb{C}^{T \times 1}$ ,  $\mathbb{C} \triangleq \mathbb{C}^1$ .  $\Re(\cdot)$  and  $\Im(\cdot)$  are the real and imaginary parts of a complex number, respectively.  $\mathbf{0}_{m \times n}$  is the  $m \times n$  all zero matrix.  $\log_a$  is the base- $a$  logarithm;  $\log \triangleq \log_e$ .  $\|\cdot\|$  is the Euclidean norm,  $\|\cdot\|_1$  is the matrix 1-norm,  $\|\cdot\|_2$  is the matrix 2-norm, and  $|\langle \cdot, \cdot \rangle|$  is the inner product. Depending on the context,  $|\cdot|$  is the norm of a complex number or the cardinality of a set.  $\lceil \cdot \rceil$  is the ceiling function. For a set  $\mathcal{A}$ ,  $\mathcal{A}^k$  is its  $k$ th Cartesian power. For a logical statement  $S$ , we let  $\mathbf{1}(S) = 1$  if  $S$  is true, and otherwise, we let  $\mathbf{1}(S) = 0$ .

## II. PRELIMINARIES

### A. System Model

We consider a slow-fading MISO multicast channel with one transmitter with  $T$  antennas and  $K$  single-antenna receivers. Denote the channel from Transmitter Antenna  $t$  to Receiver  $k$  ( $t \in \{1, \dots, T\}$  and  $k \in \{1, \dots, K\}$ ) as  $h_{tk} \in \mathbb{C}$ . Also, let  $\mathbf{h}_k \triangleq [h_{1k} \cdots h_{Tk}]^T \in \mathbb{C}^{T \times 1}$  and

$\mathbf{h} \triangleq [\mathbf{h}_1 \cdots \mathbf{h}_K] \in \mathbb{C}^{T \times K}$  denote the channels from the transmitter to Receiver  $k$ , and the entire channel state, respectively. We assume that Receiver  $k$  knows the vector  $\mathbf{h}_k$  of its own receiving channels perfectly.

Let  $s \in \mathbb{C}$  denote the information-bearing symbol that we wish to multicast to the receivers. For a given channel state  $\mathbf{h}$ , the transmitter sends  $s\mathbf{x}^\dagger\sqrt{P}$  over its  $T$  antennas, where  $P$  is the transmitter short-term power constraint,  $\mathbf{x} \in \mathcal{X}$  is a beamforming vector, and  $\mathcal{X} \triangleq \{\mathbf{x} \in \mathbb{C}^{T \times 1} : \|\mathbf{x}\| = 1\}$  is the set of all beamforming vectors. The channel input-output relationships are  $y_k = s\langle \mathbf{h}_k, \mathbf{x} \rangle\sqrt{P} + \eta_k$ ,  $k = 1, \dots, K$ , where  $y_k \in \mathbb{C}$  and  $\eta_k \in \mathbb{C}$  are the received signal and the noise at Receiver  $k$ , respectively. We assume that  $\eta_1, \dots, \eta_K, h_{11}, \dots, h_{TK}$  are independent circularly-symmetric complex Gaussian random variables with variance 1.

The signal-to-noise ratio (SNR) at Receiver  $k$  can be calculated to be  $|\langle \mathbf{x}, \mathbf{h}_k \rangle|^2 P$ . We refer to the quantity  $\gamma(\mathbf{x}, \mathbf{h}) \triangleq \min_k |\langle \mathbf{x}, \mathbf{h}_k \rangle|^2 P$  as the “network SNR.” For a fixed  $\mathbf{h}$  and  $\mathbf{x}$ , the capacity of the multicast channel as defined above is then  $\log_2(1 + \gamma(\mathbf{x}, \mathbf{h}))$  bits/sec/Hz. Without loss of generality, we set the target data transmission rate to be 1 bit/sec/Hz, in which case an outage occurs if  $\gamma(\mathbf{x}, \mathbf{h}) < 1$ .

When  $\mathbf{h}$  is random, we consider a general scenario where the transmitter can utilize different beamforming vectors for different channel states. For this purpose, consider an arbitrary mapping  $\mathbf{M} : \mathbb{C}^{T \times K} \rightarrow \mathcal{X}$ , and suppose that the transmitter uses the beamforming vector  $\mathbf{M}(\mathbf{h})$  for a given  $\mathbf{h}$ . We define the outage probability with  $\mathbf{M}$  as  $\text{out}(\mathbf{M}) \triangleq \mathbb{P}(\gamma(\mathbf{M}(\mathbf{h}), \mathbf{h}) < 1)$ . We also let  $\mathbf{d}(\mathbf{M}) \triangleq \lim_{P \rightarrow \infty} -\frac{\log \text{out}(\mathbf{M})}{\log P}$  denote the diversity gain with  $\mathbf{M}$ , provided that the limit exists.

### B. The Full-CSIT System

If the transmitter somehow knows the entire channel state  $\mathbf{h}$  perfectly, we say that we have a “full-CSIT system.” In such a scenario, the transmitter can utilize an optimal beamforming vector, say  $\mathbf{F}(\mathbf{h})$ , for a given  $\mathbf{h}$ , so as to minimize the outage probability. It should be clear that the minimum-possible outage probability can be reached by maximizing the network SNR for every channel state. Hence, we define the corresponding full-CSIT mapping  $\mathbf{F}$  as

$$\mathbf{F}(\mathbf{h}) \triangleq \arg \max_{\mathbf{x} \in \mathcal{X}} \gamma(\mathbf{x}, \mathbf{h}) = \arg \max_{\mathbf{x} \in \mathcal{X}} \min_k |\langle \mathbf{x}, \mathbf{h}_k \rangle|^2, \quad (1)$$

with ties broken arbitrarily. For  $K = 1$ , we have a point-to-point MISO system where the optimal transmission strategy is simply beamforming along the direction of the sole channel state  $\mathbf{h}_1$ . We may thus set  $\mathbf{F}(\mathbf{h}) = \frac{\mathbf{h}_1}{\|\mathbf{h}_1\|} = \frac{\mathbf{h}}{\|\mathbf{h}\|}$  when  $K = 1$ . The first non-trivial case is when  $K = 2$ , for which a solution has been provided in [5]. No closed-form expression for  $\mathbf{F}(\mathbf{h})$  is known for  $K \geq 3$ , although numerical solution methods are available [4]. Moreover, no closed-form expression is known for the resulting minimum-possible outage probability  $\text{out}(\mathbf{F})$  unless we have the trivial case  $K = 1$ . However, it is straightforward to at least show that  $d(\mathbf{F}) = T$  for any  $K$ ; for completeness, a proof will be provided later on.

### C. Partial CSIT Systems via Distributed Limited Feedback

As evident from (1), the calculation of the optimal beamforming vector requires the knowledge of the entire channel state  $\mathbf{h}$ . On the other hand, none of the terminals in the network can acquire  $\mathbf{h}$  in its entirety. In fact, Receiver  $k$  can only acquire its own *local* channel states  $\mathbf{h}_k$  via transmitter training. To calculate  $\mathbf{F}(\mathbf{h})$ , the  $K$  parts  $\mathbf{h}_1, \dots, \mathbf{h}_K$  of the channel state  $\mathbf{h}$  should be available to the transmitter, which would require an “infinite rate of feedback” from all of the  $K$  receivers. Therefore, while a full-CSIT system provides the best possible performance, it is very difficult, if not impossible, to realize in a practical system.

We thus wish to design practical limited feedback schemes that can provably achieve, or at least approach, the performance with full CSIT. For this purpose, we consider a distributed limited feedback scenario that operates as follows: Given channel state  $\mathbf{h}$ , for every  $k \in \{1, \dots, K\}$ , Receiver  $k$  quantizes only its own local channel state  $\mathbf{h}_k$  and broadcasts the corresponding feedback bits. At this stage, one can imagine that there are  $K$  feedback messages “on the air,” with each message corresponding to each one of the receivers. We assume that the transmitter and the receivers acquire these  $K$  feedback messages without any errors or delays. The  $K$  feedback messages are jointly decoded at all the terminals to reproduce a common quantized beamforming vector, say,  $\mathbf{y}$  that depends on the feedback messages (Each terminal uses the same decoder, and thus each terminal obtains the same vector  $\mathbf{y}$ ). Finally, the transmitter begins data transmission via  $\mathbf{y}$ . The receivers can decode the transmitted symbols as they know  $\mathbf{y}$  and their own receiving channels.

Given  $n \in \mathbb{N}$ , let  $\mathcal{B}_n$  denote the set of all binary codewords of length  $n$  (e.g.,  $\mathcal{B}_2 =$

{00, 01, 10, 11}). Also, suppose Receiver  $k$  sends  $b_k$  bits of feedback for every channel state. We can then formalize the above distributed limited feedback scenario via a distributed channel quantizer  $\mathbf{Q} : \mathbb{C}^{T \times K} \rightarrow \mathcal{X}$ , which is uniquely specified by  $K$  encoders  $E_k : \mathbb{C}^T \rightarrow \mathcal{B}_{b_k}$ ,  $k = 1, \dots, K$ , with the  $k$ th encoder  $E_k$  available at the  $k$ th receiver, and a unique decoder  $\mathcal{D} : \prod_{k=1}^K \mathcal{B}_{b_k} \rightarrow \mathcal{X}$  that is available at all terminals. Given  $\mathbf{h}$ , the feedback message of Receiver  $k$  is  $E_k(\mathbf{h}_k)$ . The  $K$  messages  $E_1(\mathbf{h}_1), \dots, E_K(\mathbf{h}_K)$  are jointly decoded at all the receivers and the transmitter to reproduce the quantized beamforming vector  $\mathbf{Q}(\mathbf{h}) \triangleq \mathcal{D}(E_1(\mathbf{h}_1), \dots, E_K(\mathbf{h}_K))$ . We refer to the set  $\{\mathbf{Q}(\mathbf{h}) : \mathbf{h} \in \mathbb{C}^{T \times K}\}$  as the quantizer codebook.

The distributed quantizer  $\mathbf{Q}$  thus describes a mapping from the set of channel states to a set of beamforming vectors. In fact, it is a special case of the general mapping  $\mathbf{M}$  as described in Section II-A. The outage probability with  $\mathbf{Q}$  is thus  $\text{out}(\mathbf{Q}) = \mathbb{P}(\gamma(\mathbf{Q}(\mathbf{h}), \mathbf{h}) < 1)$ , which is achieved with a feedback rate of  $\mathbf{R}_k(\mathbf{Q}) \triangleq b_k$  bits per channel state at the  $k$ th receiver.

The adjective “distributed” signifies that there are potentially many ( $K > 1$ ) non-communicating quantizer encoders each of which quantizes only a part of the entire CSI. Note that in the special case of  $K = 1$ , we have a “non-distributed” quantizer with a single quantizer encoder at the sole receiver and a quantizer decoder at the transmitter. To gain initial insight on the problem of designing distributed quantizers, we discuss the existing non-distributed quantizer design methodology for  $K = 1$ , and show why the same design ideas cannot immediately be applied to the case of  $K > 1$ .

#### D. Non-Distributed vs. Distributed Quantization

For any given compact set (a codebook of beamforming vectors)  $\mathcal{C} \subset \mathcal{X}$ , let

$$\mathbf{M}_{\mathcal{C}}^*(\mathbf{h}) \triangleq \arg \max_{\mathbf{x} \in \mathcal{C}} \gamma(\mathbf{x}, \mathbf{h}). \quad (2)$$

It can be shown that  $\mathbf{M}_{\mathcal{C}}^*(\mathbf{h})$  is an optimal mapping for codebook  $\mathcal{C}$  in the sense that for any other mapping  $\mathbf{M} : \mathbb{C}^{T \times K} \rightarrow \mathcal{C}$ , we have  $\text{out}(\mathbf{M}_{\mathcal{C}}^*) \leq \text{out}(\mathbf{M})$ .

In a point-to-point system ( $K = 1$ ), the mapping (2) can easily be realized with limited feedback: The sole receiver can determine the SNR-maximizing beamforming vector  $\mathbf{M}_{\mathcal{C}}^*(\mathbf{h})$  in  $\mathcal{C}$ , and feed back  $\lceil \log_2 |\mathcal{C}| \rceil$  bits that can uniquely represent  $\mathbf{M}_{\mathcal{C}}^*(\mathbf{h})$ . Using these feedback bits, the transmitter can recover and transmit via the beamforming vector  $\mathbf{M}_{\mathcal{C}}^*(\mathbf{h})$ . Therefore, when  $K = 1$ , it is clear how to optimally design the quantizer encoding and decoding functions

for a given codebook. The problem of designing a good quantizer boils down to the design of good codebooks, and several constructions (e.g., Grassmannian codebooks) are available.

On the other hand, in a multicast network with more than one receiver ( $K > 1$ ), none of the receivers can, by itself, determine the beamforming vector  $\mathbf{M}_{\mathcal{C}}^*(\mathbf{h})$  that provides the highest network SNR. This is because the network SNR  $\gamma(\mathbf{x}, \mathbf{h}) = \min_k |\langle \mathbf{x}, \mathbf{h}_k \rangle|^2 P$  provided by a given beamforming vector  $\mathbf{x} \in \mathcal{C}$ , depends in general on all the  $KT$  channels from the transmitter to the  $K$  receivers. Therefore, when  $K > 1$ , for a general codebook  $\mathcal{C}$ , it is not immediately clear how to implement the optimal mapping in (2), or whether such an implementation is even possible.

The absence of a rate-limited distributed implementation of (2) is a fundamental difficulty in designing structured distributed quantizers. Our general quantizer design strategy is thus to forget about picking the best beamforming vector, and instead focus on not picking the worst beamforming vector(s) in a given codebook. For an optimal execution of this design strategy, we also take into account the specific performance measure at hand and the corresponding performance goal.

We first show how to achieve the full-CSIT diversity gain  $\mathbf{d}(\mathbf{F})$  using distributed limited feedback. Before proceeding, it is worth mentioning that one can achieve full diversity by transmitting independent complex Gaussian symbols over each antenna (without the need of any feedback) instead of the rank-1 beamforming strategy that we consider in this paper. The advantage of beamforming is that it provides the opportunity of using the already-available point-to-point codes for Gaussian channels for simpler encoding/decoding of data.

### III. DIVERSITY GAINS OF DISTRIBUTED QUANTIZERS

In this section, we design distributed quantizers that can achieve the full-CSIT diversity gain  $T$ . For a point-to-point MISO system with beamforming, it is a well-known fact that a necessary and sufficient condition to achieve full diversity is to have  $\lceil \log_2 T \rceil$  feedback bits from the receiver [8]. Here, we generalize this result to the multicast setting by showing that a necessary and sufficient condition to achieve the full-diversity gain  $T$  is to have  $\lceil \log_2 T \rceil$  feedback bits from *every* receiver.

Let us first verify that indeed we have  $\mathbf{d}(\mathbf{F}) = T$  for any  $K$ . Let  $\mathbf{e}_t \triangleq [\mathbf{0}_{1 \times (t-1)} \ 1 \ \mathbf{0}_{1 \times T-t}]$ ,  $t = 1, \dots, T$  denote the antenna selection vectors, and  $\mathcal{E} \triangleq \{\mathbf{e}_1, \dots, \mathbf{e}_T\}$  represent their codebook.

**Proposition 1.** *For any  $K$ , we have  $\mathbf{d}(\mathbf{F}) = T$ .*

*Proof.* For any  $\mathbf{x}$  and  $\mathbf{h}$ , we have  $\gamma(\mathbf{x}, \mathbf{h}) = \min_k |\langle \mathbf{x}, \mathbf{h}_k \rangle|^2 P \leq \min_k \|\mathbf{h}_k\|^2 P \leq \|\mathbf{h}_1\|^2 P$ . Hence,  $\gamma(\mathbf{F}(\mathbf{h}), \mathbf{h}) \leq \|\mathbf{h}_1\|^2 P$ . This final upper bound provides a diversity gain of  $T$ , which implies  $\mathbf{d}(\mathbf{F}) \leq T$ . On the other hand, we have  $\gamma(\mathbf{F}(\mathbf{h}), \mathbf{h}) = \max_{\mathbf{x} \in \mathcal{X}} \min_k |\langle \mathbf{x}, \mathbf{h}_k \rangle|^2 P \geq \max_{\mathbf{x} \in \mathcal{E}} \min_k |\langle \mathbf{x}, \mathbf{h}_k \rangle|^2 P = \max_t \min_k |\langle \mathbf{e}_t, \mathbf{h}_k \rangle|^2 P = \max_t \min_k |h_{kt}|^2 P$ . Since the random variables  $\min_k |h_{kt}|^2 P$ ,  $t = 1, \dots, T$  are independent with each providing a diversity gain of 1, their maximum provides a diversity gain of  $T$ . Hence,  $\mathbf{d}(\mathbf{F}) \geq T$ . Combining this with the inequality  $\mathbf{d}(\mathbf{F}) \leq T$  we have already proved, we obtain  $\mathbf{d}(\mathbf{F}) = T$ , as desired.  $\square$

The proof of Proposition 1 above also suggests the following possible strategy for the construction of a quantizer that can achieve full diversity: If one can select the best antenna with the highest network SNR for every channel state, then one can achieve full diversity. This is, however, equivalent to using the mapping  $\mathbf{M}_{\mathcal{E}}^*$ , which does not admit a distributed implementation as discussed in Section II-D. In order to design a distributed quantizer that achieves full diversity, we recall our general design strategy (see Section II-D): Instead of trying to pick the best beamforming vector in a given codebook (such as  $\mathcal{E}$ ), we shall instead focus on not picking the worst beamforming vector(s) in a given codebook.

For this purpose, for any given  $n \geq T$ , let  $\mathcal{C}_n \triangleq \{\mathbf{x}_1, \dots, \mathbf{x}_n\} \subset \mathcal{X}$  be an arbitrary set of beamforming vectors such that for any  $n \geq T$ , any  $T$  of the vectors in  $\mathcal{C}_n$  (chosen without repetition) are linearly independent.<sup>1</sup> For example, for  $T = 2$ ,  $n = 3$ , and  $\mathcal{C}_3 = \{\mathbf{x}_1, \mathbf{x}_2, \mathbf{x}_3\}$ , (i) the vectors  $\mathbf{x}_1$  and  $\mathbf{x}_2$  should be linearly independent, (ii) the vectors  $\mathbf{x}_1$  and  $\mathbf{x}_3$  should be linearly independent, and (iii) the vectors  $\mathbf{x}_2$  and  $\mathbf{x}_3$  should be linearly independent.

Consider now a MISO system where  $K = 1$ . For any given  $\mathbf{h} \in \mathbb{C}^T$  and  $n \geq T$ , the vectors in  $\mathcal{C}_n$  can be ordered from the worst to the best in terms of the SNR provided by each. In other words, we have  $|\langle \mathbf{x}_{i_1}, \mathbf{h}_1 \rangle|^2 P \leq \dots \leq |\langle \mathbf{x}_{i_n}, \mathbf{h}_1 \rangle|^2 P$  for some  $i_1, \dots, i_n$  with  $\{i_1, \dots, i_n\} = \{1, \dots, n\}$ . Note that the ordering indices  $i_1, \dots, i_n$  depend on the channel state  $\mathbf{h} = \mathbf{h}_1$ . We consider the mapping  $\mathbf{w}_n(\mathbf{h}_1) \triangleq \mathbf{x}_{i_T}$  that chooses the “ $T$ th worst” beamforming vector

<sup>1</sup>The following argument guarantees the existence of such codebooks for any  $n \geq T$ : Let the random vectors  $\mathbf{u}_1, \dots, \mathbf{u}_n$  be independent and uniformly distributed on  $\mathcal{X}$ . Let  $p$  be the probability that there are  $T$  vectors in  $\{\mathbf{u}_1, \dots, \mathbf{u}_n\}$  which are not linearly independent. It is sufficient to show  $p = 0$ . For this purpose, let  $p'$  be the probability that the  $T$  vectors  $\mathbf{u}_1, \dots, \mathbf{u}_T$  are not linearly independent. We have  $p' = 0$ . By a union bound, we obtain  $p \leq \binom{n}{T} p' = 0$ , which means  $p = 0$ .



in  $\mathcal{C}_n$ . For example, for  $T = 2$ ,  $n = 4$ , and  $\mathcal{C}_4 = \{\mathbf{x}_1, \mathbf{x}_2, \mathbf{x}_3, \mathbf{x}_4\}$ , suppose that  $|\langle \mathbf{x}_3, \mathbf{h}' \rangle|^2 \leq |\langle \mathbf{x}_1, \mathbf{h}' \rangle|^2 \leq |\langle \mathbf{x}_2, \mathbf{h}' \rangle|^2 \leq |\langle \mathbf{x}_4, \mathbf{h}' \rangle|^2$  for some  $\mathbf{h}' \in \mathbb{C}^2$ . Then, we have  $\mathbf{w}_4(\mathbf{h}') = \mathbf{x}_1$ .

The mapping  $\mathbf{w}_n$  just avoids the  $T - 1$  worst beamforming vectors in  $\mathcal{C}_n$ . We now show that it provides full diversity.

**Proposition 2.** *Let  $K = 1$ . Then,  $\mathbf{d}(\mathbf{w}_n) = T, \forall n \geq T$ .*

*Proof.* Suppose  $n = T$ . Then, by definition,  $\mathbf{w}_T(\mathbf{h}_1) = \arg \max_{\mathbf{x} \in \mathcal{C}_T} |\langle \mathbf{x}, \mathbf{h}_1 \rangle|^2$ . In other words, one chooses the best beamforming vector out of  $T$  linearly independent beamforming vectors. It is known (see [8]) that this mapping provides full diversity. For  $n > T$ , let  $\mathfrak{J}$  represent the collection of all subsets of  $\{1, \dots, n\}$  with cardinality  $T$ . We have  $|\mathfrak{J}| = \binom{n}{T}$ . For any  $\mathcal{J} \in \mathfrak{J}$ , let  $\mathcal{A}(\mathcal{J}) = \{\mathbf{h}_1 : \mathcal{J} = \{i_1, \dots, i_T\}\} \subset \mathbb{C}^T$ . In other words,  $\mathcal{A}(\mathcal{J})$  represents the set of channel states for which the  $T$  worst beamforming vectors have the indices in  $\mathcal{J}$ . Note that if  $\mathbf{h}_1 \in \mathcal{A}(\mathcal{J})$ , then  $\mathbf{w}_n$  chooses the best beamforming vector in codebook  $\{\mathbf{x}_j : j \in \mathcal{J}\}$ . In other words, if  $\mathbf{h}_1 \in \mathcal{A}(\mathcal{J})$ , then  $\mathbf{w}_n(\mathbf{h}_1) = \mathbf{M}_{\{\mathbf{x}_j : j \in \mathcal{J}\}}^*(\mathbf{h}_1)$ . Therefore,

$$\text{out}(\mathbf{w}_n) = \sum_{\mathcal{J} \in \mathfrak{J}} \mathbb{P} \left( |\langle \mathbf{M}_{\{\mathbf{x}_j : j \in \mathcal{J}\}}^*(\mathbf{h}_1), \mathbf{h}_1 \rangle|^2 < \frac{1}{P}, \mathbf{h}_1 \in \mathcal{A}(\mathcal{J}) \right) \quad (3)$$

$$\leq \sum_{\mathcal{J} \in \mathfrak{J}} \mathbb{P} \left( |\langle \mathbf{M}_{\{\mathbf{x}_j : j \in \mathcal{J}\}}^*(\mathbf{h}_1), \mathbf{h}_1 \rangle|^2 < \frac{1}{P} \right) \quad (4)$$

$$= \sum_{\mathcal{J} \in \mathfrak{J}} \text{out} \left( \mathbf{M}_{\{\mathbf{x}_j : j \in \mathcal{J}\}}^* \right). \quad (5)$$

This implies  $\mathbf{d}(\mathbf{w}_n) \geq \min_{\mathcal{J} \in \mathfrak{J}} \mathbf{d}(\mathbf{M}_{\{\mathbf{x}_j : j \in \mathcal{J}\}}^*)$ . For any  $\mathcal{J} \in \mathfrak{J}$ , the  $T$  vectors  $\mathbf{x}_j, j \in \mathcal{J}$  are linearly independent by the construction of the codebook  $\mathcal{C}_n$ . By invoking the already-established special case of the proposition for  $n = T$ , we have  $\mathbf{d}(\mathbf{M}_{\{\mathbf{x}_j : j \in \mathcal{J}\}}^*) = T$  for any  $\mathcal{J} \in \mathfrak{J}$ , and thus  $\mathbf{d}(\mathbf{w}_n) \geq T$ . Since (obviously)  $\mathbf{d}(\mathbf{w}_n) \leq T$  as well, we obtain  $\mathbf{d}(\mathbf{w}_n) = T$ .  $\square$

The proposition shows that in the MISO setting, it is not necessary to choose the best beamforming vector in a codebook to achieve full diversity. One just has to avoid the  $T - 1$  worst beamforming vectors and pick at least the  $T$ th worst vector in the given codebook. We now show how this observation can be applied to the multicast setting for designing a distributed quantizer that can achieve full diversity. We first provide an example for  $T = 4$ ,  $K = 2$ , and then state and prove the general case.

**Example 1.** *Let  $T = 4$  and  $K = 2$ . We design a distributed quantizer that achieves full diversity by using 2 feedback bits per receiver per channel state. Consider the codebook  $\mathcal{C}_{16} =$*

$\{\mathbf{x}_1, \dots, \mathbf{x}_{16}\}$ . Note that any 4 of the 16 vectors in  $\mathcal{C}_{16}$  are linearly independent. We imagine the vectors in  $\mathcal{C}_{16}$  as cells of a  $4 \times 4$  grid as shown in Fig. 1. In a sense that is to be made precise in the following, Receiver 1 will be “working on” the columns of the grid, while Receiver 2 will work on the rows of the grid. Each row/column is uniquely represented by one of the 2-bit binary codewords. All the data in Fig. 1 will be available at both receivers and the transmitter.

		Receiver 1			
		00	01	10	11
Receiver 2	00	$\mathbf{x}_1$	$\mathbf{x}_2$	$\mathbf{x}_3$	$\mathbf{x}_4$
	01	$\mathbf{x}_5$	$\mathbf{x}_6$	$\mathbf{x}_7$	$\mathbf{x}_8$
	10	$\mathbf{x}_9$	$\mathbf{x}_{10}$	$\mathbf{x}_{11}$	$\mathbf{x}_{12}$
	11	$\mathbf{x}_{13}$	$\mathbf{x}_{14}$	$\mathbf{x}_{15}$	$\mathbf{x}_{16}$

Fig. 1: An example quantizer for  $T = 4$ ,  $K = 2$ .

Consider now a distributed quantizer, namely  $\tilde{\mathcal{Q}}$ , that operates as follows: Given channel state  $\mathbf{h} = [\mathbf{h}_1 \ \mathbf{h}_2]$ , Receiver 1 calculates and sorts its SNR values as  $|\langle \mathbf{x}_{i_1}, \mathbf{h}_1 \rangle|^2 P \leq \dots \leq |\langle \mathbf{x}_{i_{16}}, \mathbf{h}_1 \rangle|^2 P$  for some  $\{i_1, \dots, i_{16}\} = \{1, \dots, 16\}$ . Then, for Receiver 1, as far as its received SNR is concerned, the 3 worst beamforming vectors are  $\mathbf{x}_{i_1}, \mathbf{x}_{i_2}$  and  $\mathbf{x}_{i_3}$ . In the grid representation of the 16 beamforming vectors in Fig. 1, there exists a column index, say  $I_c \in \{1, \dots, 4\}$ , that does not contain any one of the 3 worst beamforming vectors  $\mathbf{x}_{i_1}, \mathbf{x}_{i_2}$  and  $\mathbf{x}_{i_3}$ . Receiver 1 feeds back the 2-bit binary codeword that represents  $I_c$  (For example, if  $i_1 = 9, i_2 = 7$  and  $i_3 = 16$ , we have  $I_c = 2$ , and Receiver 1 feeds back 01.). Similarly, Receiver 2 calculates and sorts its SNR values as  $|\langle \mathbf{x}_{j_1}, \mathbf{h}_2 \rangle|^2 P \leq \dots \leq |\langle \mathbf{x}_{j_{16}}, \mathbf{h}_2 \rangle|^2 P$  for some  $\{j_1, \dots, j_{16}\} = \{1, \dots, 16\}$ . There exists, this time, a row index  $I_r \in \{1, \dots, 4\}$  that does not contain any one of the 3 worst beamforming vectors  $\mathbf{x}_{j_1}, \mathbf{x}_{j_2}, \mathbf{x}_{j_3}$  for Receiver 2. Receiver 2 feeds back 2 bits that represents  $I_r$ . The transmitter recovers the indices  $I_r$  and  $I_c$ , and transmits over the beamforming vector in the  $I_r^{\text{th}}$  row,  $I_c^{\text{th}}$  column of the grid in Fig. 1.

We now analyze the diversity gain with  $\tilde{\mathcal{Q}}$ . Using a union bound over all receivers, we have

$$\text{out}(\tilde{\mathcal{Q}}) = \text{P} \left( \min_{k \in \{1, 2\}} |\langle \tilde{\mathcal{Q}}(\mathbf{h}), \mathbf{h}_k \rangle|^2 P < 1 \right) \quad (6)$$

$$\leq \sum_{k=1}^2 \mathbb{P}(|\langle \tilde{\mathbf{Q}}(\mathbf{h}), \mathbf{h}_k \rangle|^2 P < 1). \quad (7)$$

On the other hand, by construction, the quantizer  $\tilde{\mathbf{Q}}$  avoids any of the 3 worst beamforming vectors for any of the receivers. Hence, by Proposition 2, for any  $k \in \{1, 2\}$ , we have  $\mathbb{P}(|\langle \tilde{\mathbf{Q}}(\mathbf{h}), \mathbf{h}_k \rangle|^2 P < 1) \in O(P^{-4})$ . This implies  $d(\tilde{\mathbf{Q}}) = 4$ .  $\square$

The construction in Example 1 extends to the case of an arbitrary number of receivers and transmitter antennas in a straightforward manner. In fact, let  $\mathcal{I}$  be the collection of all  $K$ -dimensional vectors whose components are elements of the set  $\{1, \dots, T\}$ . We use the notation  $\mathbf{i} = [i_1 \cdots i_K] \in \mathcal{I}$  for members of  $\mathcal{I}$ . Consider now the reindexing of the codebook  $\mathcal{C}_{TK}$  as  $\{\mathbf{x}_{\mathbf{i}} : \mathbf{i} \in \mathcal{I}\} = \mathcal{C}_{TK}$ . For any  $\mathbf{h}$ , and any  $k \in \{1, \dots, K\}$ , there is an index  $I_k \in \{1, \dots, T\}$  such that the set  $\{\mathbf{x}_{\mathbf{i}} : i_k = I_k, \mathbf{i} \in \mathcal{I}\}$  does not contain any of the  $T - 1$  worst beamforming vectors for Receiver  $k$ . The index  $I_k$  can be calculated at Receiver  $k$  and fed back using  $\lceil \log_2 T \rceil$  bits. The transmitter uses the beamforming vector  $\mathbf{x}_{[I_1 \dots I_K]}$ . Let  $\tilde{\mathbf{Q}}_{\mathcal{C}_{TK}}$  denote the corresponding distributed quantizer. Proposition 2 applied to a union bound over all receivers reveals that  $\tilde{\mathbf{Q}}_{\mathcal{C}_{TK}}$  provides full diversity. We summarize this result by the following proposition.

**Proposition 3.** *For any  $T$  and  $K$ , there is a quantizer  $\mathbf{Q}$  with  $d(\mathbf{Q}) = T$  and  $R_k(\mathbf{Q}) = \lceil \log_2 T \rceil, \forall k$ .*

Let us now state the converse result. The proof is provided in Appendix A.

**Proposition 4.** *For any quantizer  $\mathbf{Q}$  with  $R_k(\mathbf{Q}) < \lceil \log_2 T \rceil$  for some  $k \in \{1, \dots, K\}$ , we have  $d(\mathbf{Q}) < T$ .*

The main result of this section is then the following combined restatement of Propositions 3 and 4.

**Theorem 1.** *A necessary and sufficient condition to achieve the full diversity gain in a quantized multicast beamforming system is to have  $\lceil \log_2 T \rceil$  feedback bits from each receiver.*

This generalizes the classical result for point-to-point MISO systems to multicast systems.

#### IV. APPROACHING THE FULL-CSIT PERFORMANCE WITH DISTRIBUTED QUANTIZATION

We now consider the design of distributed quantizers whose outage probabilities can be made arbitrarily close to that of a full-CSIT system. This will be accomplished via the following two steps. As the first step, in Section IV-A, we will show that for any arbitrary codebook  $\mathcal{C}$ , one can synthesize a distributed quantizer that can achieve the same performance as the centralized quantizer  $\mathbf{M}_{\mathcal{C}}^*$ . Then, as the second step, in Section IV-B, we will show that the outage probabilities of centralized quantizers with well-designed codebooks can approach the full-CSIT outage probability. A combined restatement of our results in these two steps will show the existence of distributed quantizers whose performances can approach that of a full-CSIT system.

##### *A. The Synthesis of a Distributed Quantizer out of a Centralized Quantizer*

As we have mentioned in Section II-D, the fundamental difficulty in designing distributed quantizers is the absence of a rate-limited distributed implementation of the optimal centralized quantizer  $\mathbf{M}_{\mathcal{C}}^*(\mathbf{h}) = \arg \max_{\mathbf{x} \in \mathcal{C}} \min_k |\langle \mathbf{x}, \mathbf{h}_k \rangle|$  for a given codebook  $\mathcal{C}$ . If this difficulty could be overcome, the problem of designing a good distributed quantizer would boil down to the much easier problem of designing a good quantizer codebook.

Fortunately, for the outage probability performance measure, we do not need to implement  $\mathbf{M}_{\mathcal{C}}^*$  as it is. In fact, an outage event with  $\mathbf{M}_{\mathcal{C}}^*$ , i.e. the event  $\max_{\mathbf{x} \in \mathcal{C}} \min_k |\langle \mathbf{x}, \mathbf{h}_k \rangle|^2 P < 1$ , occurs if and only if  $\max_{\mathbf{x} \in \mathcal{C}} \min_k \mathbf{1}(|\langle \mathbf{x}, \mathbf{h}_k \rangle|^2 P < 1) = 1$ . Hence, for

$$\bar{\mathbf{Q}}_{\mathcal{C}}(\mathbf{h}) \triangleq \arg \max_{\mathbf{x} \in \mathcal{C}} \min_k \mathbf{1}(|\langle \mathbf{x}, \mathbf{h}_k \rangle|^2 P < 1) \quad (8)$$

(with ties broken arbitrarily), we have  $\text{out}(\bar{\mathbf{Q}}_{\mathcal{C}}) = \text{out}(\mathbf{M}_{\mathcal{C}}^*)$ . But now, in contrast to the mapping  $\mathbf{M}_{\mathcal{C}}^*$ , the new mapping  $\bar{\mathbf{Q}}_{\mathcal{C}}$  can be realized as a distributed quantizer. In fact, Receiver  $k$  can calculate the  $|\mathcal{C}|$  binary values  $\mathbf{1}(|\langle \mathbf{x}, \mathbf{h}_k \rangle|^2 P < 1)$ ,  $\mathbf{x} \in \mathcal{C}$  and feed them back using  $|\mathcal{C}|$  feedback bits. The transmitter can then determine  $\bar{\mathbf{Q}}_{\mathcal{C}}(\mathbf{h})$  for every given  $\mathbf{h}$  via  $|\mathcal{C}|$  feedback bits from each receiver. We summarize these results by the following proposition.

**Proposition 5.** *For any codebook  $\mathcal{C}$ , we have  $\text{out}(\bar{\mathbf{Q}}_{\mathcal{C}}) = \text{out}(\mathbf{M}_{\mathcal{C}}^*)$  with  $\mathbf{R}_k(\bar{\mathbf{Q}}_{\mathcal{C}}) = |\mathcal{C}|$ ,  $\forall k$ .*

There is one particular disadvantage of our construction so far: In general, the synthesis of the distributed version of an  $R$ -bit centralized quantizer (with a codebook of cardinality

$2^R$ ) requires  $2^R$  bits per receiver. In other words, there is an exponential rate amplification while transitioning from the centralized to the distributed. For example, an 8-bit centralized quantizer requires 256 bits per receiver to be realized in a distributed manner.

In order to resolve the exponential rate amplification problem, we revisit the operation of the encoders of  $\bar{\mathbf{Q}}_{\mathcal{C}}$  at the receivers. Consider, for example, the quantizer encoding operation at Receiver 1. For a given codebook  $\mathcal{C} = \{\mathbf{x}_i : i = 1, \dots, |\mathcal{C}|\}$ , what Receiver 1 feeds back can be thought as a *configuration*  $\{i : |\langle \mathbf{x}_i, \mathbf{h}_1 \rangle|^2 P \leq 1\}$ , i.e. a set of beamforming vectors in  $\mathcal{C}$  that result in outage at Receiver 1 given that the channel state from the transmitter to Receiver 1 is  $\mathbf{h}_1$ . Now, let

$$\chi(\mathcal{C}) \triangleq \left| \left\{ \{i : |\langle \mathbf{x}_i, \mathbf{h}_1 \rangle|^2 P \leq 1\} : \mathbf{h}_1 \in \mathbb{C}^{T \times 1} \right\} \right| \quad (9)$$

denote the cardinality of the collection of all configurations given  $\mathcal{C}$ . In order to convey the binary values  $\mathbf{1}(|\langle \mathbf{x}, \mathbf{h}_k \rangle|^2 P < 1)$ ,  $\mathbf{x} \in \mathcal{C}$  to the transmitter, it is then sufficient for each receiver to send  $\lceil \log_2 \chi(\mathcal{C}) \rceil$  feed back bits for every channel state. The reason  $|\mathcal{C}|$  feed back bits is sufficient for this purpose is a result of the trivial estimate  $\chi(\mathcal{C}) \leq 2^{|\mathcal{C}|}$ .

As it turns out, for codebooks with large enough cardinalities, the quantity  $\chi(\mathcal{C})$  is in fact much smaller than  $2^{|\mathcal{C}|}$ , i.e. most of the configurations are, in fact, not feasible. To show this, we will utilize an existing result on hyperplane arrangements on real Euclidean spaces. Let  $\mathcal{F} = \{\mathbf{f}_i : i = 1, \dots, |\mathcal{F}|\} \in \mathbb{R}^d$  be an arbitrary codebook of  $d$ -dimensional real vectors. A hyperplane  $\{\mathbf{f} \in \mathbb{R}^d : \langle \mathbf{f}, \mathbf{d} \rangle = b\}$  where  $\mathbf{d} \in \mathbb{R}^d - \{\mathbf{0}\}$  and  $b \in \mathbb{R}$  then induces the configuration  $\{i : \langle \mathbf{f}_i, \mathbf{d} \rangle \leq b\}$  on codebook  $\mathcal{F}$ . We let

$$\bar{\chi}(\mathcal{F}) \triangleq \left| \left\{ \{i : \langle \mathbf{f}_i, \mathbf{d} \rangle \leq b\} : b \in \mathbb{R}, \mathbf{d} \in \mathbb{R}^d \right\} \right|. \quad (10)$$

One can readily observe that when  $d = 1$ , we have  $\bar{\chi}(\mathcal{F}) \leq 2|\mathcal{F}|$  with equality if and only if the elements of  $\mathcal{F}$  are all distinct. For a general  $d$ , Harding [16] has proved the tight bound

$$\bar{\chi}(\mathcal{F}) \leq 2 \sum_{i=0}^d \binom{|\mathcal{F}| - 1}{i}. \quad (11)$$

whenever  $|\mathcal{F}| \geq d + 1$ . This implies that for any non-empty  $\mathcal{F}$ , we have  $\bar{\chi}(\mathcal{F}) \in O(|\mathcal{F}|^d)$ . This greatly improves upon the trivial estimate  $\bar{\chi}(\mathcal{F}) \leq 2^{|\mathcal{F}|}$ , especially when  $|\mathcal{F}|$  is large.

Coming back to our problem, the quantity  $\chi(\mathcal{C})$  in (9) can be thought as a “complex version” of  $\bar{\chi}(\mathcal{F})$ . Intuition suggests that a similar bound on  $\chi(\mathcal{C})$  should hold. We verify

this intuition via the following proposition, which essentially follows from Harding’s result. The proof of the proposition can be found in Appendix B.

**Proposition 6.** *For any codebook  $\mathcal{C}$ , we have*

$$\chi(\mathcal{C}) \leq \min \left\{ 2^{|\mathcal{C}|}, 16 \left( \sum_{i=0}^{2T} \binom{|\mathcal{C}| - 1}{i} \right)^4 \right\}. \quad (12)$$

Note that the upper bound in (12) is  $O(|\mathcal{C}|^{8T})$  as  $|\mathcal{C}| \rightarrow \infty$ . Therefore, for any fixed codebook  $\mathcal{C}$  and a given channel state  $\mathbf{h}$ , the binary values  $\mathbf{1}(|\langle \mathbf{x}, \mathbf{h}_k \rangle|^2 P < 1)$ ,  $\mathbf{x} \in \mathcal{C}$  can be losslessly conveyed from Receiver  $k$  to the transmitter using  $8T \log_2 |\mathcal{C}| + O(1)$  bits. Using these feedback bits, the transmitter can determine  $\bar{\mathbf{Q}}_{\mathcal{C}}(\mathbf{h})$  in the same manner as discussed in the beginning of this section. The resulting quantizer, which we shall refer to as  $\mathbf{Q}_{\mathcal{C}}$  from now on, achieves  $\text{out}(\mathbf{Q}_{\mathcal{C}}) = \text{out}(\bar{\mathbf{Q}}_{\mathcal{C}}) = \text{out}(\mathbf{M}_{\mathcal{C}}^*)$ . This establishes the following main result of this section.

**Theorem 2.** *For any codebook  $\mathcal{C}$ , we have  $\text{out}(\mathbf{Q}_{\mathcal{C}}) = \text{out}(\mathbf{M}_{\mathcal{C}}^*)$  with  $\mathbf{R}_k(\mathbf{Q}_{\mathcal{C}}) \in 8T \log_2 |\mathcal{C}| + O(1)$ ,  $\forall k$ .*

Hence, for any rate- $R$  centralized quantizer, we can synthesize a distributed quantizer that achieves the same performance as the centralized quantizer and can operate with roughly  $8TR$  bits per receiver.

### B. The Existence of Good Centralized Quantizer Codebooks

In Section IV-A, we have shown how to synthesize a distributed quantizer out of a centralized quantizer in a rate-efficient manner. Therefore, to design a good distributed quantizer whose performance can approach the full-CSIT performance, it is sufficient to design a good centralized quantizer. Also, since the performance of an optimal centralized quantizer depends only on its codebook, we just need to design a good quantizer codebook.

Since the full-CSIT beamforming vector can take any value in  $\mathcal{X}$ , we wish to design our codebook, say  $\mathcal{C}$ , in such a way that for any beamforming vector  $\mathbf{x} \in \mathcal{X}$ , there should be a codebook element, say  $\mathbf{y} \in \mathcal{C}$ , that is “close enough” to  $\mathbf{x}$ . This way, we hope to minimize the losses due to quantization of an optimal beamforming vector. Also, a desirable property is to have some control on the precision of quantization, i.e., how close we want  $\mathbf{y}$  to be to  $\mathbf{x}$ . In this context, high precision translates to low quantization losses but high feedback

rates, while low precision means high losses but low rates. Such a control over precision thus allows us to determine the dependence of the achievable performance on the feedback rate. These properties that we wish to have in our codebook design lead to the following notion of a “ $\delta$ -covering codebook.”

**Definition 1.** Let  $\delta \in (0, 1)$ . We call  $\mathcal{D}_\delta$  a  $\delta$ -covering codebook if  $\forall \mathbf{x} \in \mathcal{X}$ ,  $\exists \mathbf{y} \in \mathcal{D}$  such that  $|\langle \mathbf{y}, \mathbf{x} \rangle|^2 \geq 1 - \delta$ .

Though not vital for our discussions in this paper, explicit constructions of  $\delta$ -covering codebooks for any  $\delta \in (0, 1)$  is available [11]. We summarize the construction in [11] below.

**Example 2.** Given  $\delta \in (0, 1)$ , let  $s_\delta \triangleq 2^{\lceil \log_2(2T/\delta) \rceil + 1}$ ,  $\mathcal{S}_\delta \triangleq \{-1 + ks_\delta, k = 0, \dots, 2s_\delta^{-1}\}$ , and  $\mathcal{Y}_\delta \triangleq \{\mathbf{y}/\|\mathbf{y}\| : \Re y_1, \Im y_1, \dots, \Re y_T, \Im y_T \in \mathcal{S}_\delta, \text{ and } 0 < \|\mathbf{y}\| \leq 1\}$ . According to [11, Proposition 3], for any  $\delta \in (0, 1)$ , the codebook  $\mathcal{Y}_\delta$  is a  $\delta$ -covering codebook with  $|\mathcal{Y}_\delta| \in O(\delta^{-2t})$ .

Let us now discuss how to utilize the  $\delta$ -covering codebooks in a point-to-point MISO system with  $K = 1$ . Consider the optimal quantizer  $\mathbf{M}_{\mathcal{D}_\delta}^*(\mathbf{h}) = \arg \max_{\mathbf{x} \in \mathcal{D}_\delta} |\langle \mathbf{x}, \mathbf{h} \rangle|^2$  for codebook  $\mathcal{D}_\delta$ . Note that  $\mathbf{M}_{\mathcal{D}_\delta}^*$  can be implemented using  $\lceil \log_2 |\mathcal{D}_\delta| \rceil$  feedback bits. By the definition of  $\mathcal{D}_\delta$ , for any given channel state  $\mathbf{h}_1$ , there exists  $\mathbf{y} \in \mathcal{D}_\delta$  such that  $|\langle \frac{\mathbf{h}_1}{\|\mathbf{h}_1\|}, \mathbf{y} \rangle|^2 \geq 1 - \delta$ . This leads to the lower bound

$$|\langle \mathbf{M}_{\mathcal{D}_\delta}^*(\mathbf{h}_1), \mathbf{h}_1 \rangle|^2 P \geq |\langle \mathbf{F}(\mathbf{h}_1), \mathbf{h}_1 \rangle|^2 P(1 - \delta) \quad (13)$$

$$= \|\mathbf{h}_1\|^2 P(1 - \delta). \quad (14)$$

Therefore, in the worst case scenario, a well-designed quantizer results in a uniformly bounded multiplicative SNR loss. This is a very useful property as most performance measures (such as outage probability or ergodic capacity) are monotonic functions of the SNR. We can thus conclude that the performance of a rate- $\lceil \log_2 |\mathcal{D}_\delta| \rceil$  quantized beamforming system at power  $P$  is at least that of a full-CSIT system at power  $P(1 - \delta)$ . In particular, for the codebook  $\mathcal{Y}_\delta$ , the performance with  $R$  bits of feedback at power  $P$  is no worse than the full-CSIT performance at power  $P(1 - O(2^{-\frac{R}{2t}}))$ .

Then, a fundamental question is to determine whether or not the SNR loss due to quantization can similarly be uniformly bounded for a general multicast system with  $K > 1$  receivers. The positive answer is provided by the following theorem for the special case of  $K = 2$  receivers. The proof of the theorem can be found in Appendix C.

**Theorem 3.** *Let  $K = 2$ . We have  $\forall \mathbf{h} \in \mathbb{C}^{T \times K}$ ,  $\exists \mathbf{y} \in \mathcal{Y}_\delta$ ,  $\gamma(\mathbf{y}, \mathbf{h}) \geq \gamma(\mathbf{F}(\mathbf{h}), \mathbf{h})(1 - O(\sqrt{\delta}))$ .*

Hence, for a two-receiver system, for any channel state, one can quantize the full-CSIT beamforming vector  $\mathbf{F}(\mathbf{h})$  to a quantized vector  $\mathbf{y} \in \mathcal{D}_\delta$  such that the SNR loss is always within  $(1 - O(\sqrt{\delta}))$  of the SNR with full CSIT. A major open problem is to study whether a similar result holds for more than two receivers. We shall note the following in this context.

**Remark 1.** Let  $\vec{\mathbf{h}}_k = \frac{\mathbf{h}_k}{\|\mathbf{h}_k\|}$ ,  $k = 1, \dots, K$ . In [1, Section III.B], it is claimed that (with our notation)

$$|\langle \mathbf{F}(\mathbf{h}), \vec{\mathbf{h}}_k \rangle|^2 \geq \frac{1}{K^2}, \forall k, \forall \mathbf{h}. \quad (15)$$

If (15) were true, then Theorem 3 could easily be shown to hold for any  $K$ . In fact, suppose (15) holds. Due to the  $\delta$ -covering property of  $\mathcal{Y}_\delta$ , there is a beamforming vector  $\mathbf{y} \in \mathcal{Y}_\delta$  such that  $|\langle \mathbf{F}(\mathbf{h}), \mathbf{y} \rangle|^2 \geq 1 - \delta$ . Using Lemma 1 in Appendix C, we could then obtain  $|\langle \mathbf{y}, \vec{\mathbf{h}}_k \rangle|^2 \geq |\langle \mathbf{F}(\mathbf{h}), \vec{\mathbf{h}}_k \rangle|^2 (1 - K^2 \sqrt{\delta})$ ,  $\forall k, \forall \mathbf{h}$ . Multiplying each side of this inequality by  $\|\mathbf{h}_k\|^2$  and then taking the minimum over all  $k$ , we would obtain  $\gamma(\mathbf{y}, \mathbf{h}) \geq \gamma(\mathbf{F}(\mathbf{h}), \mathbf{h})(1 - K^2 \sqrt{\delta})$ , which generalizes Theorem 3 to any number of receivers  $K$ .

Unfortunately, the claim in (15) does not hold (despite the fact that  $\min_k |\langle \mathbf{F}(\mathbf{h}), \mathbf{h}_k \rangle|^2 \geq \frac{1}{K^2} \min_k \|\mathbf{h}_k\|^2$  holds for every  $\mathbf{h}$  as shown in [15, Claim 2.4.2(i)]). As a counterexample, let  $T = K = 2$  with  $\mathbf{h}_1 = [2 \ 0]^T$ ,  $\mathbf{h}_2 = [0 \ 1]^T$ . For  $\mathbf{z} = [z_1 \ z_2]^T = \mathbf{F}([\mathbf{h}_1 \ \mathbf{h}_2])$ , suppose  $|z_1|^2 = |\langle \mathbf{z}, \vec{\mathbf{h}}_1 \rangle|^2 \geq \frac{1}{K^2} = 0.25$ . Then, since  $|z_1|^2 + |z_2|^2 \leq 1$ , we have  $|z_2|^2 \leq 0.75$ . This implies  $\gamma(\mathbf{z}, \mathbf{h}) = \min_{k \in \{1,2\}} |\langle \mathbf{z}, \mathbf{h}_k \rangle|^2 \leq |\langle \mathbf{z}, \mathbf{h}_2 \rangle|^2 = |z_2|^2 \leq 0.75$ . On the other hand, for the beamforming vector  $\mathbf{z}' = [\sqrt{0.2} \ \sqrt{0.8}]^T$ , we have  $\gamma(\mathbf{z}', \mathbf{h}) = 0.8 > 0.75$ , which contradicts the optimality of  $\mathbf{z}$ . In fact, it can be shown (we omit the proof here as it is not relevant to our major focus) that for any  $K$  and  $T$  there exists a set  $\mathcal{H} \subset \mathbb{C}^{T \times K}$  of channel states such that  $P(\mathcal{H}) > 0$  and  $|\langle \mathbf{F}(\mathbf{h}), \vec{\mathbf{h}}_k \rangle|^2 \geq \frac{1}{K^2}$ ,  $\forall k$  fails to hold for any  $\mathbf{h} \in \mathcal{H}$ .  $\square$

Now, let  $\text{out}(\mathbf{M}; P)$  denote the outage probability of mapping  $\mathbf{M}$  for a given transmitter power constraint  $P$ . Combining Theorems 2 and 3, we obtain the following main result for a two-user multicast system with distributed quantized beamforming.

**Theorem 4.** *Let  $K = 2$ . With  $R$  feedback bits per receiver per channel state, an outage performance of  $\text{out}(\mathbf{F}; P(1 - O(2^{-\frac{R}{32T^2}})))$  is achievable at any  $P$ .*



*Proof.* Theorem 3 implies that for the optimal centralized quantizer  $\mathbf{M}_{\mathcal{Y}_\delta}^*$  for  $\mathcal{Y}_\delta$ , we have  $\text{out}(\mathbf{M}_{\mathcal{Y}_\delta}^*; P) \leq \text{out}(\mathbf{F}; P(1 - O(\sqrt{\delta})))$ . On the other hand, by Theorem 2,  $\mathbf{M}_{\mathcal{Y}_\delta}^*$  can be realized as the distributed quantizer  $\mathbf{Q}_{\mathcal{Y}_\delta}$  using  $8T \log_2 |\mathcal{D}_\delta| + O(1) = 16T^2 \log_2 \frac{1}{\delta} + O(1)$  bits per each receiver. The equality follows since  $|\mathcal{Y}_\delta| \in O(\delta^{-2T})$  as shown in Example 2. Setting  $R = 16T^2 \log_2 \frac{1}{\delta} + O(1)$  and solving for  $\delta$ , we obtain the statement of the theorem.  $\square$

One way to visualize the outage probability loss due to quantization is that in the usual graph of  $P$  in the horizontal axis versus the outage probability in the vertical axis (where both axes are in the logarithmic scale), the outage probability with  $R$  bits of feedback per receiver is at most the full-CSIT curve shifted  $-10 \log_{10}(1 - O(2^{-\frac{R}{32T^2}}))$  dBs to the right. Equivalently, since as  $x \rightarrow 0$ ,  $-\log(1 - O(x)) = O(x)$ , the outage probability with  $R$  bits of feedback is within  $O(2^{-\frac{R}{32T^2}})$  dBs to the outage probability with full CSIT.

It is also instructive to compare our result for  $K = 2$  to that of a point-to-point MISO system where  $K = 1$ . For the case  $K = 1$ , it is known (see e.g. [9]) that the performance with quantized beamforming is at most within  $O(2^{-\frac{R}{T-1}})$  dBs to the outage probability with full CSIT. Hence, despite the complicated distributed nature of the quantizer design problem for  $K = 2$ , the performance loss due to quantization can still be made to decay exponentially with the per-receiver feedback rate  $R$  as  $O(2^{-\frac{R}{32T^2}})$ . Here, we also note that the factor  $32T^2$  is likely not the best possible, and can perhaps be improved (made smaller) with more work.

Let us now discuss the case of more than two receivers  $K \geq 3$ . In this case, our results are not strong enough to prove that one can uniformly approach the full-CSIT performance using distributed feedback. The difficulty is to show the existence of good centralized quantizers whose performances can uniformly approach the full-CSIT performance. The existence of such quantizers can be proved e.g. by an extension of Theorem 3 to more than two receivers. Nevertheless, if such an extension had been available, it would have been straightforward to synthesize good distributed quantizers out of the good centralized quantizers via the same arguments as in Section IV-A (Note that the results of Section IV-A hold for any number of receivers.). On the other hand, intuition suggests, without much room for doubt, that a sequence of optimal centralized quantizers for a sequence of  $\delta$ -covering codebooks should uniformly approach the full-CSIT performance as  $\delta \rightarrow 0$ . Part of the next section verifies this intuition with numerical simulations for the special case of three receivers.

## V. NUMERICAL RESULTS

In this section, we provide numerical simulations that verify our analytical results. We first show examples of quantizers that can achieve full diversity in a rate-optimal manner for a three-receiver system with two or three transmitter antennas.

We recall from Section III that the necessary and sufficient condition to achieve full diversity is to have  $\lceil \log_2 2 \rceil = 1$  bit of feedback per receiver when  $T = 2$ , and  $\lceil \log_2 3 \rceil = 2$  bits of feedback per receiver when  $T = 3$ . Such a performance can be achieved with the quantizer  $\tilde{\mathbf{Q}}_{\mathcal{C}_{TK}}$ , where  $\mathcal{C}_{TK}$  is an arbitrary codebook of cardinality  $T^K$  with the property that any of its  $T - 1$  elements are linearly independent. Hence, for the case  $T = 2$  and  $K = 3$ , we have constructed one codebook  $\mathcal{C}_{8,1}$  of cardinality  $2^3 = 8$  by drawing 8 samples independently and uniformly at random on  $\mathcal{X}$ . We have constructed a second codebook  $\mathcal{C}_{8,2}$  of cardinality 8 via the same procedure. For the case  $T = 3$  and  $K = 3$ , we have constructed the codebooks  $\mathcal{C}_{27,1}$  and  $\mathcal{C}_{27,2}$  of cardinality  $3^3 = 27$  in the same manner. The performance of the corresponding quantizers  $\tilde{\mathbf{Q}}_{\mathcal{C}_{8,1}}$ ,  $\tilde{\mathbf{Q}}_{\mathcal{C}_{8,2}}$ ,  $\tilde{\mathbf{Q}}_{\mathcal{C}_{27,1}}$  and  $\tilde{\mathbf{Q}}_{\mathcal{C}_{27,2}}$  are then as shown in Fig. 2 together with the performance of the open-loop system ( $R = 0$ ) with no feedback<sup>2</sup> and the full-CSIT systems ( $R = \infty$ ). In the figure, the horizontal axis represents  $P$  in decibels, and the vertical axis represents the outage probability.

We can observe that the open-loop system can only achieve a diversity gain of 1, the quantizers  $\tilde{\mathbf{Q}}_{\mathcal{C}_{8,1}}$  and  $\tilde{\mathbf{Q}}_{\mathcal{C}_{8,2}}$  can achieve a diversity gain of 2, and the quantizers  $\tilde{\mathbf{Q}}_{\mathcal{C}_{27,1}}$  and  $\tilde{\mathbf{Q}}_{\mathcal{C}_{27,2}}$  achieve a diversity gain of 3. In other words, the quantizers  $\tilde{\mathbf{Q}}_{\mathcal{C}_{8,1}}$ ,  $\tilde{\mathbf{Q}}_{\mathcal{C}_{8,2}}$ ,  $\tilde{\mathbf{Q}}_{\mathcal{C}_{27,1}}$  and  $\tilde{\mathbf{Q}}_{\mathcal{C}_{27,2}}$  achieve the maximal diversity gains of their respective systems. Obviously, and as we can also observe from Fig. 2, the array gains of quantizers of the form  $\tilde{\mathbf{Q}}_{\mathcal{C}}$  will depend on the exact values of the elements of  $\mathcal{C}$ . Further optimizations of the codebook  $\mathcal{C}$  in this context is an interesting direction for future work.

We now verify that we can approach the full-CSIT performance by increasing the per-receiver feedback rates. Also, in order to demonstrate that our constructions can be applied to different codebook designs, we consider here Grassmannian codebooks. Let  $\mathcal{G}_{T,N} \triangleq \arg \min_{\mathcal{C} \in \mathcal{X}^N} \max\{|\langle \mathbf{x}, \mathbf{y} \rangle| : \mathbf{x}, \mathbf{y} \in \mathcal{C}, \mathbf{x} \neq \mathbf{y}\}$  denote a cardinality- $N$  Grassmannian codebook

<sup>2</sup>The open-loop system refers to the scenario where the transmitter uses a unique beamforming vector, say  $\mathbf{x}_o \in \mathcal{X}$ , for every channel state. The resulting outage probability can be shown to be independent of the number of transmitter antennas  $T$  and the choice of the beamforming vector  $\mathbf{x}_o$ .

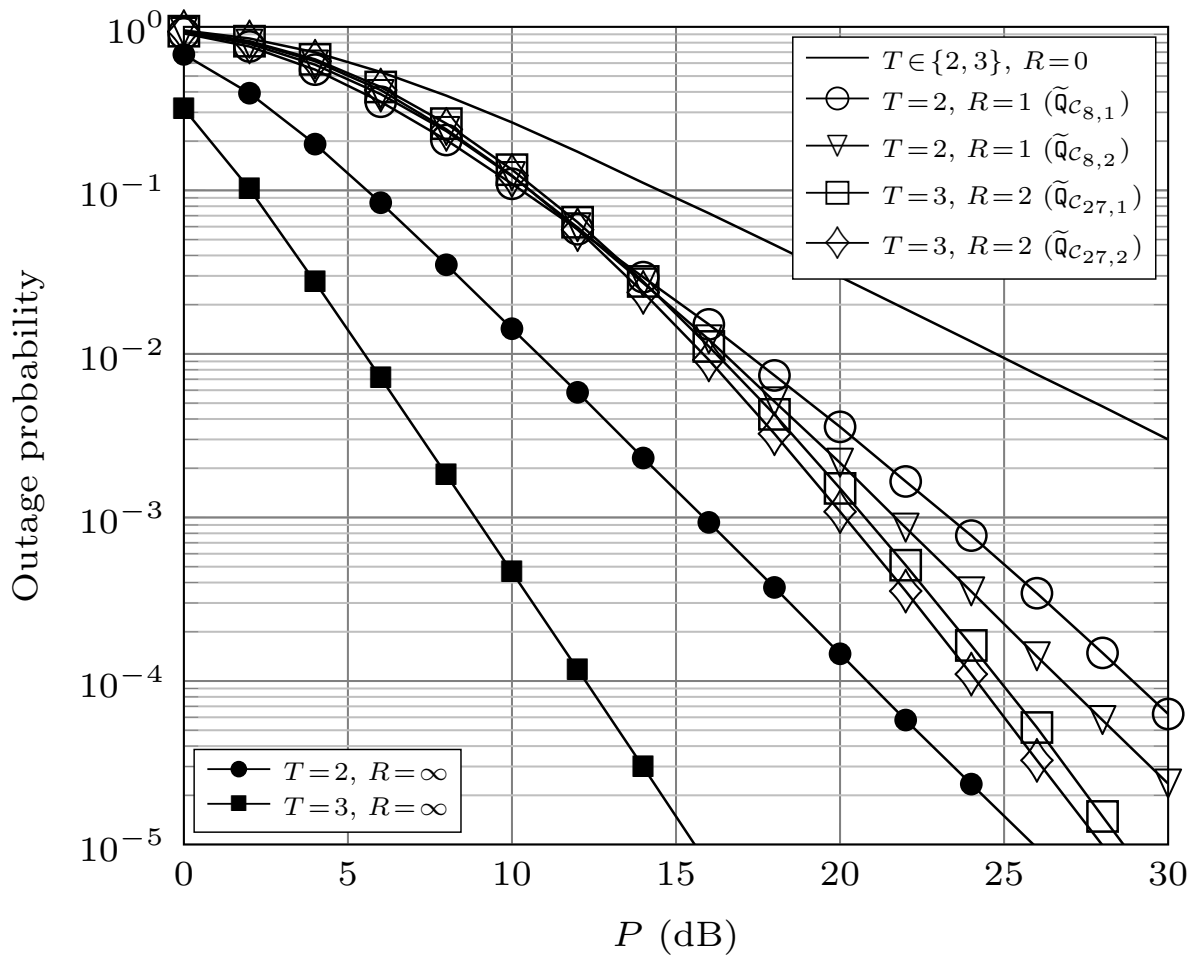


Fig. 2: Achieving full diversity in a rate-optimal manner using distributed limited feedback.

for a system with  $T$  transmitter antennas. Note that by definition, the codebook  $\mathcal{G}_{T,N}$  provides the maximum minimum distance between the distinct codebook elements among all codebooks with cardinality  $N$ .

We have constructed Grassmannian codebooks  $\mathcal{G}_{T,N}$  for  $(T, N) \in \{(2, 2), (2, 4), (2, 16), (3, 4), (3, 8), (3, 16), (3, 256)\}$  via numerical methods [8]. The performance of the corresponding distributed quantizers  $\bar{\mathcal{Q}}_{\mathcal{G}_{T,N}}$  (which can be implemented using  $R = N$  bits per receiver)

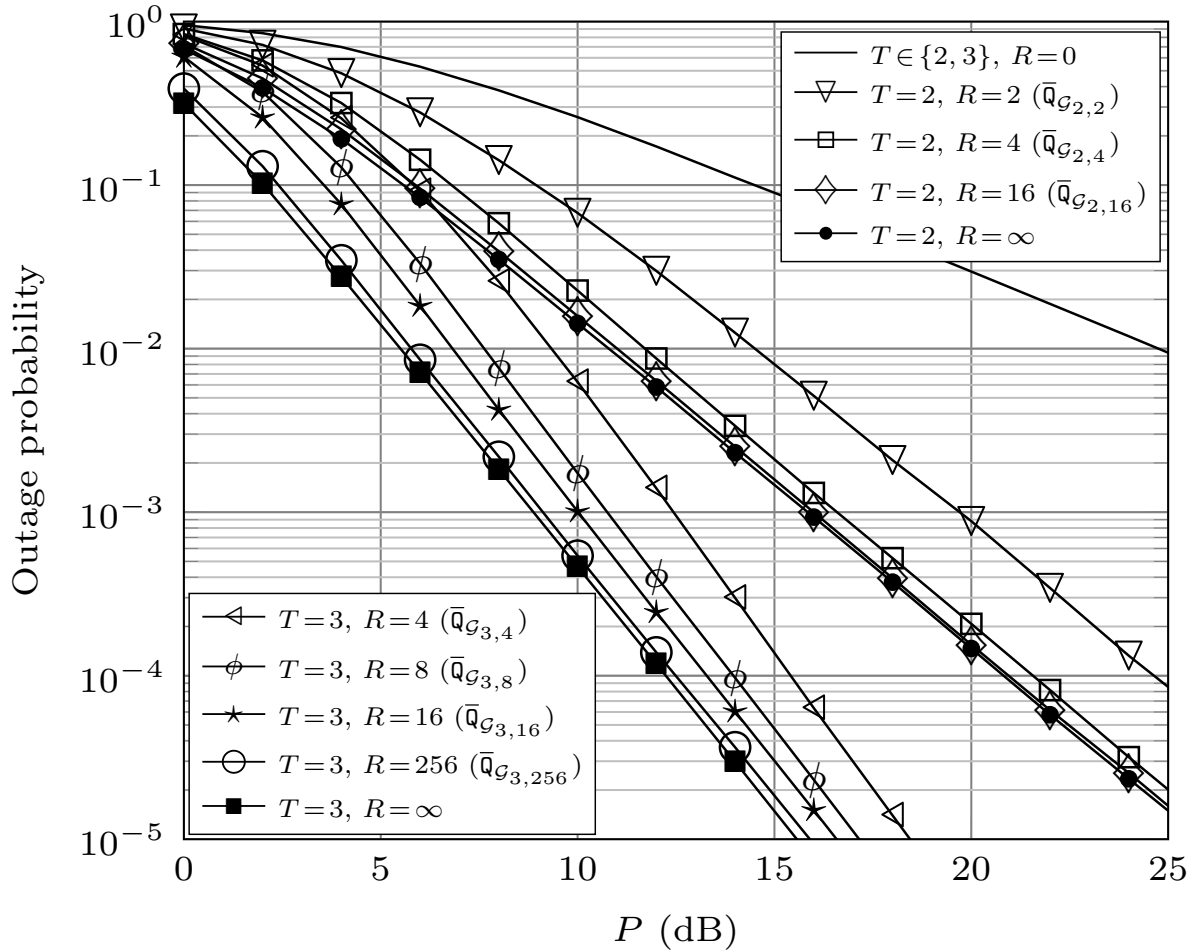


Fig. 3: Approaching the full-CSIT performance using distributed limited feedback.

are shown in Fig. 3 for a three-receiver system with either  $T = 2$  and  $T = 3$ . We can observe that for both cases  $T = 2$  and  $T = 3$ , as the per-receiver feedback rate  $R = N$  increases, the performance of the distributed quantizers approaches uniformly to the full-CSIT performance. This suggests that Theorem 4 will also hold for  $K = 3$ . We have obtained similar results for a two-receiver system  $K = 2$ , and have thus verified Theorem 4.

We now consider the performance of the quantizers  $\mathcal{Q}_{\mathcal{G}_{T,N}}$  for different  $T$  and  $N$ ; we refer to

Section IV-A for the definition of  $\mathbf{Q}_{\mathcal{C}}$  for a given codebook  $\mathcal{C}$ . Note that for any  $T$  and  $N$ , the quantizer  $\mathbf{Q}_{\mathcal{G}_{T,N}}$  can achieve the exact same performance as the quantizer  $\bar{\mathbf{Q}}_{\mathcal{G}_{T,N}}$ , and it uses only  $\lceil \log_2 \chi(\mathcal{G}_{T,N}) \rceil \in O(\log^4 N)$  feedback bits instead of the  $N$  bits as required by  $\bar{\mathbf{Q}}_{\mathcal{G}_{T,N}}$ . In general, implementing the quantizer  $\mathbf{Q}_{\mathcal{C}}$  requires one to determine the collection of all possible configurations  $\mathfrak{C} \triangleq \{i : |\langle \mathbf{x}_i, \mathbf{h}_1 \rangle|^2 P \leq 1\} : \mathbf{h}_1 \in \mathbb{C}^{T \times 1}\}$  given  $\mathcal{C}$ . Therefore, as a first step, we have estimated  $\mathfrak{C}$  via  $\mathfrak{C}' \triangleq \{i : |\langle \mathbf{x}_i, \mathbf{h}_{1j} \rangle|^2 \leq 1\} : j = 1, \dots, J\}$ , where  $\mathbf{h}_{1j}$ ,  $j = 1, \dots, J$  is a sequence of circularly-symmetric complex Gaussian random vectors with unit variance for each component, and  $J$  is chosen to be a sufficiently large number so as to (hopefully) observe all configurations. Using this method, we have identified 4 configurations for  $\mathcal{G}_{2,2}$ , 16 for  $\mathcal{G}_{2,4}$ , 1090 for  $\mathcal{G}_{2,16}$ , 16 for  $\mathcal{G}_{3,4}$ , 256 for  $\mathcal{G}_{3,8}$ , and finally, 14496 for  $\mathcal{G}_{3,16}$ . We have then simulated the actual communication system as follows: Given codebook  $\mathcal{C}$ , each receiver calculates the configuration  $c$  corresponding to the given (generated) channel state. This is followed by each receiver feeding back the index of the configuration in  $\mathfrak{C}'$  with the smallest Hamming distance to  $c$ . The quantizer decoder uses the beamforming vector that avoids outage at all the receivers according to the received configurations. If no such beamforming vector exists, the quantizer decoder uses an arbitrary beamforming vector. Note that the performance of the resulting quantizer will be the same as  $\mathbf{Q}_{\mathcal{C}}$  (and thus  $\bar{\mathbf{Q}}_{\mathcal{C}}$ ) provided that  $\mathfrak{C}' = \mathfrak{C}$ . In fact, for the Grassmannian codebooks, the resulting simulated outage probabilities was an exact match with the outage probabilities in Fig. 3. This shows, for example, that the cost of implementing  $\mathcal{G}_{2,16}$  can be lowered down to  $\lceil \log_2 1090 \rceil = 11$  bits per receiver instead of 16 bits per receiver.

Unfortunately, determining the set of all configurations of a codebook with our “exhaustive Monte Carlo” method is not a feasible task when the codebook cardinality is large. For example, interpolation suggests that the codebook  $\mathcal{G}_{3,256}$  will have around  $2^{40}$  configurations (so that it can be implemented with 40 bits per receiver instead of 256 bits per receiver.). One solution may be to find good structured codebook designs that will induce a structured set of configurations. Finding such codebook designs, or, in general, designing efficient algorithms for finding good distributed quantizers will remain as challenging open problems.

## VI. CONCLUSIONS

We have studied the design of outage-optimal distributed quantizers for beamforming in MISO multicast channels. We have constructed rate-optimal quantizers that can achieve

full diversity. For the special case of a two-receiver system, we have also designed quantizers that can provably approach the outage probability with full CSIT. Determining whether a similar result holds for more than two receivers remains as an open problem.

Beamforming is not optimal in general, e.g. for more than 3 receivers. Therefore, the extensions of our results for general covariance transmission is another important future research direction. Also, the problem of how to design (analytically or numerically) distributed quantizers for different performance measures should be studied in detail. In this context, as far as formal analytical results are concerned, the main obstacle is the construction of a distributed quantizer that provides a uniformly bounded multiplicative SNR loss. Numerical design of distributed quantizers also prove to be very challenging. For example, Lloyd-like algorithms incur a computational complexity that grows exponentially with the number of receivers, and are thus simply not feasible as numerical design methods.

Finally, we would also like to mention that some of our results can be extended to “outage-like” performance measures with little or no additional work. One example is the scenario where one wishes to minimize the maximum error probability among all receivers for a given specific channel code. Our diversity results can easily be extended to this scenario with no additional work. Similarly, with  $T$  transmitter antennas,  $\lceil \log_2 T \rceil$  feedback bits from each receiver can be shown to be necessary and sufficient.

## APPENDIX A

### PROOF OF PROPOSITION 4

Suppose (without loss of generality) that  $R_1(\mathbf{Q}) < \lceil \log_2 T \rceil$ . Let  $\mathcal{B}_k \triangleq \{E_k(\mathbf{h}) : \mathbf{h} \in \mathbb{C}^T\}$  denote the image of the encoding function  $E_k$  at the  $k$ th receiver. We have  $|\mathcal{B}_1| < T$ . Also, let  $\mathcal{B}_{k,\mathbf{b}} \triangleq \{\mathbf{h} : E_k(\mathbf{h}) = \mathbf{b}\}$ . We can then find a lower bound on the outage probability with  $\mathbf{Q}$  as

$$\text{out}(\mathbf{Q}) = \mathbb{P}(\min_k |\langle \mathbf{Q}(\mathbf{h}), \mathbf{h}_k \rangle|^2 P < 1) \quad (16)$$

$$\geq \mathbb{P}(|\langle \mathbf{Q}(\mathbf{h}), \mathbf{h}_1 \rangle|^2 P < 1) \quad (17)$$

$$= \sum_{\mathbf{b}_K \in \mathcal{B}_K} \int_{\mathcal{B}_{K,\mathbf{b}_K}} \cdots \sum_{\mathbf{b}_1 \in \mathcal{B}_1} \int_{\mathcal{B}_{1,\mathbf{b}_1}} \mathbf{1}(|\langle D(\mathbf{b}_1, \dots, \mathbf{b}_K), \mathbf{h}_1 \rangle|^2 P < 1) d\mathbf{h}_1 \cdots d\mathbf{h}_K \quad (18)$$

$$\geq \sum_{\mathbf{b}_K \in \mathcal{B}_K} \int_{\mathcal{B}_{K,\mathbf{b}_K}} \cdots \sum_{\mathbf{b}_1 \in \mathcal{B}_1} \int_{\mathcal{B}_{1,\mathbf{b}_1}} \mathbf{1}(|\langle \mathbf{y}_{\mathbf{b}_1}, \mathbf{h}_1 \rangle|^2 P < 1) d\mathbf{h}_1 \cdots d\mathbf{h}_K \quad (19)$$

$$= \sum_{\mathbf{b}_1 \in \mathcal{B}_1} \int_{\mathcal{B}_{1,\mathbf{b}_1}} \mathbf{1} \left( |\langle \mathbf{y}_{\mathbf{b}_1}, \mathbf{h}_1 \rangle|^2 P < 1 \right) d\mathbf{h}_1 \quad (20)$$

$$\geq \sum_{\mathbf{b}_1 \in \mathcal{B}_1} \int_{\mathcal{B}_{1,\mathbf{b}_1}} \mathbf{1} \left( \max_{\mathbf{b} \in \mathcal{B}_1} |\langle \mathbf{y}_{\mathbf{b}}, \mathbf{h}_1 \rangle|^2 P < 1 \right) d\mathbf{h}_1 \quad (21)$$

$$= \mathbb{P} \left( \max_{\mathbf{b} \in \mathcal{B}_1} |\langle \mathbf{y}_{\mathbf{b}}, \mathbf{h}_1 \rangle|^2 P < 1 \right), \quad (22)$$

where  $\mathbf{y}_{\mathbf{b}} \triangleq \arg \min_{\mathbf{x} \in \mathcal{X}} \mathbb{P}(|\langle \mathbf{x}, \mathbf{h} \rangle|^2 < 1 | \mathbf{h} \in \mathcal{B}_{1,\mathbf{b}})$ ,  $\mathbf{b} \in \mathcal{B}_1$ . Let  $\mathbf{Y}$  denote the  $T \times |\mathcal{B}_1|$  matrix whose columns are the vectors  $\mathbf{y}_{\mathbf{b}}$ ,  $\mathbf{b} \in \mathcal{B}_1$ . Also, let  $\mathbf{U}_1 \mathbf{D} \mathbf{U}_2 = \mathbf{Y}$  be the singular value decomposition of  $\mathbf{Y}$ , where  $\mathbf{U}_1 \in \mathbb{C}^{T \times T}$  and  $\mathbf{U}_2 \in \mathbb{C}^{|\mathcal{B}_1| \times |\mathcal{B}_1|}$  are unitary matrices, and  $\mathbf{D} \in \mathbb{R}^{T \times |\mathcal{B}_1|}$  is a diagonal matrix with non-negative entries. Let  $d_i$ ,  $i = 1, \dots, |\mathcal{B}_1|$  denote the entry in the  $i$ th row,  $i$ th column of  $\mathbf{D}$ . We have

$$\max_{\mathbf{b} \in \mathcal{B}_1} |\langle \mathbf{y}_{\mathbf{b}}, \mathbf{h}_1 \rangle|^2 = \|\mathbf{h}_1^\dagger \mathbf{Y}\|_1^2 \quad (23)$$

$$= \|\mathbf{h}_1^\dagger \mathbf{D} \mathbf{U}_2 \mathbf{U}_2^\dagger \mathbf{U}_2\|_1^2 \quad (24)$$

$$\leq \|\mathbf{h}_1^\dagger \mathbf{D} \mathbf{U}_2 \mathbf{U}_2^\dagger\|_1^2 \|\mathbf{U}_2\|_1^2 \quad (25)$$

$$= \|\mathbf{h}_1^\dagger \mathbf{D}\|_1^2 \|\mathbf{U}_2\|_1^2 \quad (26)$$

$$\leq |\mathcal{B}_1| \|\mathbf{h}_1^\dagger \mathbf{D}\|_1^2 \|\mathbf{U}_2\|_2^2 \quad (27)$$

$$= |\mathcal{B}_1| \|\mathbf{h}_1^\dagger \mathbf{D}\|_1^2 \quad (28)$$

$$= |\mathcal{B}_1| \max_{i \in \{1, \dots, |\mathcal{B}_1|\}} d_i^2 |h_{1i}|^2 \quad (29)$$

where the first inequality follows from the submultiplicity of the 1-norm, and the second inequality follows from the inequality  $\|\mathbf{A}\|_1 \leq \sqrt{n} \|\mathbf{A}\|_2$ ,  $\mathbf{A} \in \mathbb{C}^{n \times n}$ . This implies

$$\text{out}(\mathbf{Q}) \geq \mathbb{P} \left( |\mathcal{B}_1| \max_{i \in \{1, \dots, |\mathcal{B}_1|\}} d_i^2 |h_{1i}|^2 P < 1 \right). \quad (30)$$

Since this final lower bound involves the maximum of only  $|\mathcal{B}_1|$  exponential random variables, we have  $\text{d}(\mathbf{Q}) \leq |\mathcal{B}_1| < T$ .

## APPENDIX B

### PROOF OF PROPOSITION 6

We begin with the following definitions. For any given  $\mathbf{y} \in \mathbb{C}^{T \times 1}$ , let

$$\mathbf{y}^{\text{R}} \triangleq [\Re y_1 \ \Im y_1 \ \Re y_2 \ \Im y_2 \ \cdots \ \Re y_T \ \Im y_T] \in \mathbb{R}^{2T \times 1}, \quad (31)$$

$$\mathbf{y}^{\text{I}} \triangleq [-\Im y_1 \ \Re y_1 \ -\Im y_2 \ \Re y_2 \ \cdots \ -\Im y_T \ \Re y_T] \in \mathbb{R}^{2T \times 1}. \quad (32)$$

Note that for any  $\mathbf{y}_1, \mathbf{y}_2 \in \mathbb{C}^{T \times 1}$ , we have  $\Re\langle \mathbf{y}_1, \mathbf{y}_2 \rangle = \langle \mathbf{y}_1^R, \mathbf{y}_2^R \rangle$  and  $\Im\langle \mathbf{y}_1, \mathbf{y}_2 \rangle = \langle \mathbf{y}_1^R, \mathbf{y}_2^I \rangle$ . We also let  $\mathcal{C}^R \triangleq \{\mathbf{x}^R : \mathbf{x} \in \mathcal{C}\}$ .

Now, without loss of generality, suppose that  $P = 1$  (The collection of all possible configurations remain the same unless  $P = 0$ ). We have

$$\chi(\mathcal{C}) = \left| \left\{ \{i : |\langle \mathbf{x}_i, \mathbf{y} \rangle| \leq 1\} : \mathbf{y} \in \mathbb{C}^{T \times 1} \right\} \right| \quad (33)$$

$$= \left| \left\{ \{i : -u \leq \langle \mathbf{x}_i^R, \mathbf{y}^R \rangle \leq u, -v \leq \langle \mathbf{x}_i^R, \mathbf{y}^I \rangle \leq v\} : \mathbf{y} \in \mathbb{C}^{T \times 1}, u, v \geq 0, u^2 + v^2 \leq 1 \right\} \right| \quad (34)$$

$$\leq \left| \left\{ \{i : -u \leq \langle \mathbf{x}_i^R, \mathbf{y}^R \rangle \leq u, -v \leq \langle \mathbf{x}_i^R, \mathbf{y}^I \rangle \leq v\} : \mathbf{y} \in \mathbb{C}^{T \times 1}, u, v \geq 0 \right\} \right|, \quad (35)$$

where the inequality follows since omitting a condition (which, in the above derivation, is the condition  $u^2 + v^2 \leq 1$ ) on the configurations cannot decrease the total number of configurations. Consider now the general problem of estimating the cardinality of the collection

$$\mathfrak{A} \triangleq \{ \{i : \mathbf{S}_1(i, z) \text{ and } \dots \text{ and } \mathbf{S}_L(i, z)\} : z \in \mathcal{Z} \}, \quad (36)$$

where  $i$  takes values on a finite set,  $\mathbf{S}_1(i, z), \dots, \mathbf{S}_L(i, z)$  are  $L$  arbitrary logical statements whose truth (or falsity) depend on  $i$  and  $z$ , and  $\mathcal{Z}$  is some arbitrary space where  $z$  takes its values. Let

$$\mathfrak{B}_\ell \triangleq \{ \{i : \mathbf{S}_\ell(i, z)\} : z \in \mathcal{Z} \}, \quad (37)$$

and

$$\mathfrak{B} \triangleq \prod_{\ell=1}^L \mathfrak{B}_\ell \quad (38)$$

with the understanding that the product is Cartesian. We claim that the map  $[\mathcal{B}_1 \cdots \mathcal{B}_L] \mapsto \bigcap_{\ell=1}^L \mathcal{B}_\ell$  is a surjection from  $\mathfrak{B}$  to  $\mathfrak{A}$ . In fact, if  $\mathcal{A} \in \mathfrak{A}$ , then  $\mathcal{A} = \bigcap_{\ell=1}^L \{i : \mathbf{S}_\ell(i, z_0)\}$  for some  $z_0 \in \mathcal{Z}$ , and for  $\mathcal{B}'_\ell = \{i : \mathbf{S}_\ell(i, z_0)\}$ ,  $\ell = 1, \dots, L$ , we have  $[\mathcal{B}'_1 \cdots \mathcal{B}'_L] \mapsto \mathcal{A}$ . The surjectivity implies  $|\mathfrak{A}| \leq |\mathfrak{B}|$ . Moreover, since  $\mathfrak{B}_\ell, \ell = 1, \dots, L$  are finite collections of sets, we have  $|\mathfrak{B}| = \prod_{\ell=1}^L |\mathfrak{B}_\ell|$ , and therefore

$$|\mathfrak{A}| \leq \prod_{\ell=1}^L |\mathfrak{B}_\ell| = \prod_{\ell=1}^L |\{ \{i : \mathbf{S}_\ell(i, z)\} : z \in \mathcal{Z} \}|. \quad (39)$$

In particular, for the expression in (35), we can identify 4 different conditions ( $L = 4$ ) corresponding to the 4 inequalities. We can then obtain

$$\chi(\mathcal{C}) \leq \left| \left\{ \{i : -u \leq \langle \mathbf{x}_i^R, \mathbf{y}^R \rangle\} : \mathbf{y} \in \mathbb{C}^{T \times 1}, u, v \geq 0 \right\} \right| \times$$



$$\left| \left\{ \{i : \langle \mathbf{x}_i^R, \mathbf{y}^R \rangle \leq u\} : \mathbf{y} \in \mathbb{C}^{T \times 1}, u, v \geq 0 \right\} \right| \times \left| \left\{ \{i : -v \leq \langle \mathbf{x}_i^R, \mathbf{y}^I \rangle\} : \mathbf{y} \in \mathbb{C}^{T \times 1}, u, v \geq 0 \right\} \right| \times \left| \left\{ \{i : \langle \mathbf{x}_i^R, \mathbf{y}^I \rangle \leq v\} : \mathbf{y} \in \mathbb{C}^{T \times 1}, u, v \geq 0 \right\} \right|. \quad (40)$$

The first factor can be evaluated to be

$$\left| \left\{ \{i : -u \leq \langle \mathbf{x}_i^R, \mathbf{y}^R \rangle\} : \mathbf{y} \in \mathbb{C}^{T \times 1}, u, v \geq 0 \right\} \right| = \left| \left\{ \{i : -u \leq \langle \mathbf{x}_i^R, \mathbf{y}^R \rangle\} : \mathbf{y} \in \mathbb{C}^{T \times 1}, u \geq 0 \right\} \right| \quad (41)$$

$$= \left| \left\{ \{i : -u \leq \langle \mathbf{x}_i^R, \mathbf{d} \rangle\} : \mathbf{d} \in \mathbb{R}^{2T \times 1}, u \geq 0 \right\} \right| \quad (42)$$

$$= \left| \left\{ \{i : u \geq \langle \mathbf{x}_i^R, -\mathbf{d} \rangle\} : \mathbf{d} \in \mathbb{R}^{2T \times 1}, u \geq 0 \right\} \right| \quad (43)$$

$$= \left| \left\{ \{i : u \geq \langle \mathbf{x}_i^R, \mathbf{d} \rangle\} : \mathbf{d} \in \mathbb{R}^{2T \times 1}, u \geq 0 \right\} \right| \quad (44)$$

$$\leq \left| \left\{ \{i : u \geq \langle \mathbf{x}_i^R, \mathbf{d} \rangle\} : \mathbf{d} \in \mathbb{R}^{2T \times 1}, u \in \mathbb{R} \right\} \right| \quad (45)$$

$$\leq \bar{\chi}(\mathcal{C}^R). \quad (46)$$

Each remaining factor in (40) can similarly be bounded by  $\bar{\chi}(\mathcal{C}^R)$ . Therefore, we have  $\chi(\mathcal{C}) \leq [\bar{\chi}(\mathcal{C}^R)]^4$ . Applying (11) to this final inequality proves the proposition.

## APPENDIX C

### PROOF OF THEOREM 3

We need the following two lemmas. The following lemma has originally been stated in [13], but the provided proof had flaws. Here, we provide a corrected proof.

**Lemma 1.** *For any  $\mathbf{u}, \mathbf{v}, \mathbf{w} \in \mathcal{X}$ , we have*

$$\left| |\langle \mathbf{u}, \mathbf{v} \rangle|^2 - |\langle \mathbf{u}, \mathbf{w} \rangle|^2 \right| \leq \sqrt{1 - |\langle \mathbf{v}, \mathbf{w} \rangle|^2}. \quad (47)$$

*Proof.* Let  $\mathbf{G} \triangleq \mathbf{v}\mathbf{v}^\dagger - \mathbf{w}\mathbf{w}^\dagger$  and  $z \triangleq \langle \mathbf{v}, \mathbf{w} \rangle$ . It can be verified (after some straightforward calculations) that  $\mathbf{G}$  admits the spectral decomposition

$$\mathbf{G} = \sqrt{1 - |z|^2} [\mathbf{u}_1 \ \mathbf{u}_2] \begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix} [\mathbf{u}_1 \ \mathbf{u}_2]^\dagger \quad (48)$$

$$= \sqrt{1 - |z|^2} (\mathbf{u}_1 \mathbf{u}_1^\dagger - \mathbf{u}_2 \mathbf{u}_2^\dagger), \quad (49)$$

where

$$\mathbf{u}_1 = \alpha \mathbf{v} - \beta \mathbf{v}_0 \exp(j\angle z), \quad (50)$$

$$\mathbf{u}_2 = \beta \mathbf{v} + \alpha \mathbf{v}_0 \exp(j\angle z) \quad (51)$$

are orthonormal vectors with

$$\mathbf{v}_0 = \frac{\mathbf{w} - \mathbf{v}\mathbf{v}^\dagger\mathbf{w}}{\sqrt{1 - |z|^2}}, \quad (52)$$

$$(\alpha, \beta) = \left( \sqrt{\frac{1 + \sqrt{1 - |z|^2}}{2}}, \sqrt{\frac{1 - \sqrt{1 - |z|^2}}{2}} \right). \quad (53)$$

Therefore,

$$\left| |\langle \mathbf{u}, \mathbf{v} \rangle|^2 - |\langle \mathbf{u}, \mathbf{w} \rangle|^2 \right| = |\mathbf{u}^\dagger \mathbf{G} \mathbf{u}| \quad (54)$$

$$= \sqrt{1 - |z|^2} \left| |\langle \mathbf{u}, \mathbf{u}_1 \rangle|^2 - |\langle \mathbf{u}, \mathbf{u}_2 \rangle|^2 \right| \quad (55)$$

$$\leq \sqrt{1 - |z|^2} \left( |\langle \mathbf{u}, \mathbf{u}_1 \rangle|^2 + |\langle \mathbf{u}, \mathbf{u}_2 \rangle|^2 \right) \quad (56)$$

$$= \sqrt{1 - |z|^2} \|\mathbf{u}\|^2 \quad (57)$$

$$= \sqrt{1 - |z|^2}. \quad (58)$$

This concludes the proof.  $\square$

**Lemma 2.** *For every  $\epsilon > 0$  and any  $\mathbf{v}, \mathbf{w} \in \mathcal{X}$  with  $|\langle \mathbf{v}, \mathbf{w} \rangle|^2 \leq 1 - \epsilon$ , there are constants  $\tilde{\delta}, \tilde{C} > 0$  (that may depend on  $\epsilon$ ) such that  $\forall \delta \leq \tilde{\delta}, \exists \mathbf{z} \in \mathcal{D}_\delta$  with  $|\langle \mathbf{z}, \mathbf{w} \rangle| \geq |\langle \mathbf{v}, \mathbf{w} \rangle|$  and  $|\langle \mathbf{z}, \mathbf{v} \rangle| \geq 1 - \tilde{C}\delta$ .*

*Proof.* Consider a fixed  $\epsilon > 0$  and arbitrary vectors  $\mathbf{v}, \mathbf{w} \in \mathcal{X}$  with  $|\langle \mathbf{v}, \mathbf{w} \rangle|^2 \leq 1 - \epsilon$ . We first define some auxiliary variables. Let

$$A \triangleq \frac{2 \exp(j\angle \langle \mathbf{v}, \mathbf{w} \rangle)}{\epsilon}, \quad (59)$$

$$L \triangleq \|\mathbf{v} + A\sqrt{\delta}\mathbf{w}\|, \quad (60)$$

and

$$\mathbf{u} \triangleq \frac{\mathbf{v} + A\sqrt{\delta}\mathbf{w}}{L}. \quad (61)$$

Note that  $L^2 = 1 + \delta|A|^2 + 2|A||\langle \mathbf{v}, \mathbf{w} \rangle|$ , and thus

$$L = 1 + |A||\langle \mathbf{v}, \mathbf{w} \rangle|\sqrt{\delta} + O(\delta). \quad (62)$$

with

$$\frac{1}{L} = 1 - |A||\langle \mathbf{v}, \mathbf{w} \rangle|\sqrt{\delta} - O(\delta). \quad (63)$$

By definition, for every  $\delta > 0$ , there is a  $\mathbf{z} \in Y_\delta$  with

$$|\langle \mathbf{z}, \mathbf{u} \rangle|^2 \geq 1 - \delta. \quad (64)$$

We shall prove that such a choice of  $\mathbf{z}$  satisfies  $|\langle \mathbf{z}, \mathbf{w} \rangle| \geq |\langle \mathbf{v}, \mathbf{w} \rangle|$  and  $|\langle \mathbf{z}, \mathbf{v} \rangle| \geq 1 - O(\delta)$  for every sufficiently small  $\delta$ , and this will conclude the proof of the lemma. First, we show that  $|\langle \mathbf{z}, \mathbf{w} \rangle| \geq |\langle \mathbf{v}, \mathbf{w} \rangle|$ . Let

$$\tilde{\mathbf{z}} \triangleq \mathbf{z} \exp(-j\angle \langle \mathbf{z}, \mathbf{u} \rangle). \quad (65)$$

We have

$$\|\tilde{\mathbf{z}} - \mathbf{u}\|^2 = 2 - 2\Re\langle \tilde{\mathbf{z}}, \mathbf{u} \rangle \quad (66)$$

$$= 2 - 2|\langle \mathbf{z}, \mathbf{u} \rangle| \quad (67)$$

$$\leq 2 - 2\sqrt{1 - \delta} \quad (68)$$

$$< 2 - 2\sqrt{(1 - \delta)^2} \quad (69)$$

$$= 2\delta. \quad (70)$$

Therefore,  $\|\tilde{\mathbf{z}} - \mathbf{u}\| \leq \sqrt{2\delta}$ , and thus

$$\tilde{\mathbf{z}} = \mathbf{u} + \mathbf{t}\sqrt{2\delta} \quad (71)$$

for some  $\|\mathbf{t}\| \leq 1$ . We now have

$$|\langle \mathbf{z}, \mathbf{w} \rangle| = |\langle \tilde{\mathbf{z}}, \mathbf{w} \rangle| \quad (72)$$

$$= |\langle \mathbf{u}, \mathbf{w} \rangle + \langle \mathbf{t}, \mathbf{w} \rangle \sqrt{2\delta}| \quad (73)$$

$$= \left| \frac{\langle \mathbf{v}, \mathbf{w} \rangle + A\sqrt{\delta}}{L} + \langle \mathbf{t}, \mathbf{w} \rangle \sqrt{2\delta} \right| \quad (74)$$

$$\geq \frac{|\langle \mathbf{v}, \mathbf{w} \rangle + A\sqrt{\delta}|}{L} - |\sqrt{2\delta}| \quad (75)$$

$$= (|\langle \mathbf{v}, \mathbf{w} \rangle| + |A|\sqrt{\delta})(1 - |A||\langle \mathbf{v}, \mathbf{w} \rangle|\sqrt{\delta} - O(\delta)) - \sqrt{2\delta} \quad (76)$$

$$= |\langle \mathbf{v}, \mathbf{w} \rangle| + \left( \frac{2(1 - |\langle \mathbf{v}, \mathbf{w} \rangle|^2)}{\epsilon} - \sqrt{2} \right) \sqrt{\delta} - O(\delta) \quad (77)$$

$$\geq |\langle \mathbf{v}, \mathbf{w} \rangle| + (2 - \sqrt{2})\sqrt{\delta} - O(\delta), \quad (78)$$

where the first three equalities follow from (65), (71), and (61), respectively; (75) is a consequence of the reverse triangle inequality and the fact that  $|\langle \mathbf{t}, \mathbf{z} \rangle| \leq \|\mathbf{t}\|\|\mathbf{z}\| = 1$ ; (76) follows by observing that the phase of  $\langle \mathbf{v}, \mathbf{w} \rangle$  matches that of  $A\sqrt{\delta}$  and from the substitution

of the value of  $\frac{1}{L}$  in (63); (77) follows once we substitute the value of  $A$  and after some straightforward simplifications; and, finally, (78) follows since  $\frac{1-|\langle \mathbf{v}, \mathbf{w} \rangle|^2}{\epsilon} \geq 1$  by our initial assumption on  $\mathbf{v}$  and  $\mathbf{w}$ . The last inequality implies  $|\langle \mathbf{z}, \mathbf{v} \rangle| \geq |\langle \mathbf{v}, \mathbf{w} \rangle|$  for sufficiently small  $\delta$ .

We now show that  $|\langle \mathbf{z}, \mathbf{v} \rangle| \geq 1 - O(\delta)$ . First, note that according to (61), we have  $\mathbf{v} = L\mathbf{u} - A\sqrt{\delta}\mathbf{w}$ , and therefore,

$$|\langle \mathbf{v}, \mathbf{z} \rangle| = |L\langle \mathbf{u}, \mathbf{z} \rangle - A\sqrt{\delta}\langle \mathbf{w}, \mathbf{z} \rangle| \quad (79)$$

$$\geq L|\langle \mathbf{u}, \mathbf{z} \rangle| - |A|\sqrt{\delta}|\langle \mathbf{w}, \mathbf{z} \rangle|, \quad (80)$$

where the inequality follows from the reverse triangle inequality. Let us now consider the terms in the final lower bound one by one. The quantity  $L$  has already been evaluated in (62). We also have  $|\langle \mathbf{u}, \mathbf{z} \rangle| > 1 - \delta$  according to (64). We find an upper bound on  $|\langle \mathbf{w}, \mathbf{z} \rangle|$  using the same arguments as in (72) through (77). The only difference is that for (75), we use the triangle inequality instead of the reverse triangle inequality, and this leads to the bound  $|\langle \mathbf{w}, \mathbf{z} \rangle| \leq |\langle \mathbf{v}, \mathbf{w} \rangle| + O(\sqrt{\delta})$ . Substituting all these bounds and equalities to (80), we obtain

$$|\langle \mathbf{v}, \mathbf{z} \rangle| \geq (1 + |A||\langle \mathbf{v}, \mathbf{w} \rangle|\sqrt{\delta} + O(\delta))(1 - \delta) - |A|\sqrt{\delta}(|\langle \mathbf{v}, \mathbf{w} \rangle| + O(\sqrt{\delta})). \quad (81)$$

Upon expanding the parenthesis and simplifying, we have  $|\langle \mathbf{v}, \mathbf{z} \rangle| \geq 1 - O(\delta)$ , and this concludes the proof.  $\square$

We are now ready to prove the theorem. Let  $\mathbf{z} = \mathbf{F}(\mathbf{h})$ . We shall prove that for every sufficiently small  $\delta$ , there exists  $\mathbf{y} \in \mathcal{D}_\delta$  such that  $\gamma(\mathbf{y}, \mathbf{h}) \geq (1 - \delta)\gamma(\mathbf{z}, \mathbf{h})$ .

In a multicast network with two users, it is well-known [5] that

$$\gamma(\mathbf{z}, \mathbf{h}) = \max_{\mathbf{x} \in \mathcal{X}} \min_k |\langle \mathbf{x}, \mathbf{h}_k \rangle|^2 = \max_{\mathbf{X} \in \mathbb{C}^{t \times t}: \|\mathbf{X}\| \leq 1} \min_k \|\mathbf{X}\mathbf{h}_k\|^2. \quad (82)$$

In other words, if we consider beamforming as transmission over a rank-1 covariance matrix, the SNR provided by the best rank-1 covariance matrix is already equal to the SNR provided by a general-rank covariance matrix. Hence, for any  $\mathbf{X} \in \mathbb{C}^{t \times t}$  with  $\|\mathbf{X}\| \leq 1$ , we have  $\gamma(\mathbf{z}, \mathbf{h}) \geq \min_k \|\mathbf{X}\mathbf{h}_k\|^2$ . In particular, for  $\mathbf{X} = \frac{1}{\sqrt{T}}\mathbf{I}_T$ , we obtain

$$\gamma(\mathbf{z}, \mathbf{h}) \geq \frac{1}{T} \min_k \|\mathbf{h}_k\|^2. \quad (83)$$

Now, for any given  $k \in \{1, 2\}$ , let  $\vec{\mathbf{h}}_k = \mathbf{h}_k / \|\mathbf{h}_k\|$  and  $\alpha_k = |\langle \mathbf{z}, \vec{\mathbf{h}}_k \rangle|^2$ . We have  $\gamma(\mathbf{z}, \mathbf{h}) = \min_k (\alpha_k \|\mathbf{h}_k\|^2)$ . The lower bound in (83) then implies that either  $\alpha_1 \geq \frac{1}{T}$  or  $\alpha_2 \geq \frac{1}{T}$  (As otherwise, if  $\alpha_1 < \frac{1}{T}$  and  $\alpha_2 < \frac{1}{T}$ , we have  $\gamma(\mathbf{z}, \mathbf{h}) < \frac{1}{T} \min_k \|\mathbf{h}_k\|^2$  and this contradicts (83).). Suppose that  $\alpha_1 \geq \frac{1}{T}$  and  $\alpha_2 \geq \frac{1}{T}$ . By definition, there exists  $\mathbf{y}' \in \mathcal{D}_\delta$  with  $|\langle \mathbf{z}, \mathbf{y}' \rangle|^2 \geq 1 - \delta$ . Using Lemma 1, we can then obtain

$$|\langle \mathbf{y}', \vec{\mathbf{h}}_k \rangle|^2 \geq \alpha_k - \sqrt{1 - |\langle \mathbf{z}, \mathbf{y}' \rangle|^2} \quad (84)$$

$$\geq \alpha_k - \sqrt{\delta} \quad (85)$$

$$\geq \alpha_k (1 - T\sqrt{\delta}) \quad (86)$$

for any  $k \in \{1, 2\}$ , where the last inequality follows since  $\alpha_k \geq \frac{1}{T}$ . This implies

$$\gamma(\mathbf{y}', \mathbf{h}) = \min_k (|\langle \mathbf{y}', \vec{\mathbf{h}}_k \rangle|^2 \|\mathbf{h}_k\|^2) \quad (87)$$

$$\geq \min_k (\alpha_k \|\mathbf{h}_k\|^2) (1 - T\sqrt{\delta}) \quad (88)$$

$$= \gamma(\mathbf{z}, \mathbf{h}) (1 - T\sqrt{\delta}), \quad (89)$$

and thus proves the theorem for the case where  $\alpha_1, \alpha_2 \geq \frac{1}{T}$ .

Suppose now that  $\alpha_1 \geq \frac{1}{T}$  but  $\alpha_2 < \frac{1}{T}$ . In this case, consider the application of Lemma 2 with the choice of variables  $\mathbf{v} = \mathbf{z}$  and  $\mathbf{w} = \vec{\mathbf{h}}_2$ . In such a scenario, since  $|\langle \mathbf{v}, \mathbf{w} \rangle|^2 = |\langle \mathbf{z}, \vec{\mathbf{h}}_2 \rangle|^2 = \alpha_2 < \frac{1}{T}$ , Lemma 2 is applicable for the special case  $\epsilon = 1 - \frac{1}{T}$ , and moreover, in such an application, the constants  $\delta'$  and  $C'$  in the statement of Lemma 2 will depend only on  $T$ . Hence, according to Lemma 2, there are constants  $C_1, C_2 > 0$  (that now depend only on  $T$ ) such that  $\forall \delta < C_1, \exists \mathbf{y}'' \in \mathcal{D}_\delta, |\langle \mathbf{y}'', \mathbf{z} \rangle|^2 \geq 1 - C_2\delta$  and  $|\langle \mathbf{y}'', \vec{\mathbf{h}}_2 \rangle|^2 \geq |\langle \mathbf{z}, \vec{\mathbf{h}}_2 \rangle|^2 = \alpha_2$ . The inequality  $|\langle \mathbf{y}'', \mathbf{z} \rangle|^2 \geq 1 - C_2\delta$  implies (using the same arguments as in (84) through (86)) that

$$|\langle \mathbf{y}'', \vec{\mathbf{h}}_1 \rangle|^2 \geq \alpha_1 (1 - T\sqrt{C_2\delta}). \quad (90)$$

Combining this with the bound  $|\langle \mathbf{y}'', \vec{\mathbf{h}}_2 \rangle|^2 \geq \alpha_2$ , we obtain

$$\gamma(\mathbf{y}'', \mathbf{h}) \geq \gamma(\mathbf{z}, \mathbf{h}) (1 - T\sqrt{C_2\delta}), \quad (91)$$

which proves the theorem for the case  $\alpha_1 \geq \frac{1}{T}$  and  $\alpha_2 < \frac{1}{T}$ . The remaining case  $\alpha_1 < \frac{1}{T}$  and  $\alpha_2 \geq \frac{1}{T}$  can be handled in the same manner. This concludes the proof.

## REFERENCES

- [1] N. Jindal and Z.-Q. Luo, "Capacity limits of multiple antenna multicasting," *IEEE Intl. Symp. Info. Theory*, July 2006.
- [2] S. Y. Park, D. J. Love, and D. H. Kim, "Capacity limits of multi-antenna multicasting under correlated fading channels," *IEEE Trans. Commun.*, vol. 58, no. 7, pp. 2002–2013, July 2010.
- [3] S. Y. Park and D. J. Love, "Capacity limits of multiple antenna multicasting using antenna subset selection," *IEEE Trans. Signal Process.*, vol. 56, no. 6, pp. 2524–2534, June 2008.
- [4] N. D. Sidiropoulos, T. N. Davidson, and Z.-Q. Luo, "Transmit beamforming for physical layer multicasting," *IEEE Trans. Signal Process.*, vol. 54, no. 6, pp. 2239–2251, June 2006.
- [5] I. H. Kim, D. J. Love, and S. Y. Park, "Optimal and successive approaches to signal design for multiple antenna physical layer multicasting," *IEEE Trans. Commun.*, vol. 59, no. 8, pp. 2316–2327, Aug. 2011.
- [6] S. X. Wu, W.-K. Ma, and A. M.-C. So, "Physical-layer multicasting by stochastic transmit beamforming and Alamouti space-time coding," *IEEE Trans. Signal Process.*, vol. 61, no. 17, pp. 4230–4245, Sept. 2013.
- [7] A. Narula, M. J. Lopez, M. D. Trott, and G. W. Wornell, "Efficient use of side information in multiple antenna data transmission over fading channels," *IEEE J. Select. Areas Commun.*, vol. 16, no. 8, pp. 1423–1436, Oct. 1998.
- [8] D. J. Love, R. W. Heath, Jr., and T. Strohmer, "Grassmannian beamforming for multiple-input multiple-output wireless systems," *IEEE Trans. Inf. Theory*, vol. 49, no. 10, pp. 2735–2747, Oct. 2003.
- [9] K. K. Mukkavilli, A. Sabharwal, and E. Erkip, "On beamforming with finite-rate feedback for multiple antenna systems," *IEEE Trans. Inf. Theory*, vol. 49, no. 10, pp. 2562–2579, Oct. 2003.
- [10] J. C. Roh and B. D. Rao, "Transmit beamforming in multiple antenna systems with finite rate feedback: A VQ-based approach," *IEEE Trans. Inf. Theory*, vol. 52, no. 3, pp. 1101–1112, Mar. 2006.
- [11] E. Koyuncu and H. Jafarkhani, "Variable-length limited feedback beamforming in multiple-antenna fading channels," *IEEE Trans. Inf. Theory*, vol. 60, no. 11, pp. 7140–7165, Nov. 2014.
- [12] E. Chiu and V. K. N. Lau, "Precoding design for multi-antenna multicast broadcast services with limited feedback," *IEEE Systems J.*, vol. 4, no. 4, pp. 550–560, Dec. 2010.
- [13] X. Liu, E. Koyuncu, and H. Jafarkhani, "Multicast networks with variable-length limited feedback," *IEEE Trans. Wireless Commun.*, vol. 14, no. 1, pp. 252–264, Jan. 2015.
- [14] E. Koyuncu and H. Jafarkhani, "A systematic distributed quantizer design method with an application to MIMO broadcast channels," *IEEE Data Compression Conf.*, Mar. 2010.
- [15] A. E. Litvak, V. D. Milman, and G. Schechtman, "Averages of norms and quasi-norms," *Math Annalen*, vol. 312, no. 1, pp. 95–124, Sept. 1998.
- [16] E. F. Harding, "The number of partitions of a set of  $N$  points in  $k$  dimensions induced by hyperplanes," *Edinburgh Math. Society*, pp. 285–289, 1967.