Grassmannian Beamforming for MIMO Amplify-and-Forward Relaying

Behrouz Khoshnevis, Wei Yu, and Raviraj Adve

Abstract— We consider the problem of beamforming codebook design for limited feedback half-duplex multiple-input multiple-output (MIMO) amplify-and-forward (AF) relay system. In the first part of the paper, the direct link between the source and the destination is ignored. Assuming perfect channel state information (CSI), we show that the source and the relay should map their signals to the dominant right singular vectors of the source-relay and relay-destination channels. For the limited feedback scenario, we prove the appropriateness of Grassmannian codebooks as the source and relay beamforming codebooks based on the distributions of the optimal source and relay beamforming vectors. In the second part of the paper, the direct link is considered in the problem model. Assuming perfect CSI, we derive the optimization problem that identifies the optimal source beamforming vector and show that the solution to this problem is uniformly distributed on the unit sphere for independent and identically distributed (i.i.d) Rayleigh channels. For the limited feedback scenario, we justify the appropriateness of Grassmannian codebooks for quantizing the optimal source beamforming vector based on its distribution. Finally, a modified quantization scheme is presented, which introduces a negligible penalty in the system performance but significantly reduces the required number of feedback bits.

Index Terms— Amplify-and-forward relaying, Beamforming, Grassmannian codebooks, Multiple-input multiple-output systems.

I. INTRODUCTION

Multiple-input multiple-output (MIMO) technology provides a wireless system with a large number of degrees of freedom, which can be used for increasing the capacity and/or reliability of wireless channels. Relaying techniques, on the other hand, can extend the communication range and coverage by supporting shadowed users with the help of relay nodes and reduce the transmission power required to reach the users far from the base station. These benefits make MIMO relaying techniques a powerful candidate for implementation in the next generation wireless networks.

Considering a system with a single data stream and perfect channel knowledge at the destination, several methods can be used to achieve the potential benefits of the MIMO channel. Maximum ratio transmission and combining (MRT-MRC) technique [1] can achieve full diversity order while providing considerable array gain compared to space-time codes [2]. This gain is achieved at the expense of obtaining channel knowledge at the source, therefore, the destination needs to send the quantized channel information back to the source.

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; rsadve@comm.utoronto.ca).

While a general purpose minimum mean-square error (MMSE) quantizer can be used to quantize each channel matrix entry, doing so requires a large number of feedback bits and it does not preserve the structure of the optimal beamforming vector [3]. A more efficient approach is to share a codebook of beamforming vectors between the source and destination, and send back the label of the best beamforming vector to the source. For the case of flat Rayleigh fading channel, the codebook design problem has been shown to be related to the Grassmannian line packing problem [4, 5, 6].

In this paper, we generalize the idea of MRT-MRC to the MIMO relay channel with an amplify-and-forward (AF) relay station. A general information theoretic analysis of the MIMO relay channel has been presented in [7] and [8]. Although efficient signaling through the relay channel may require a full-duplex relay with specific processing capabilities (e.g. encoding/decoding), AF relays are still attractive due to their lower implementation complexity. Moreover, the full-duplex assumption cannot be realized by current technology, as the input and output signals need to be separated in time or frequency at the relay. For these reasons, this paper focuses on the half-duplex AF relay system.

The half-duplex MIMO AF relay channel has been considered in [9] and [10], where the authors present different solutions for maximizing the instantaneous capacity with respect to the weighting matrix at the relay. These papers assume no channel state information at the source (CSIT) and consider uniform power allocation over the source antennas. The work [11] assumes perfect CSIT and derives the optimal power allocation scheme for the source and relay by neglecting the direct link. Finally, the authors of [12] and [13] study the same problem including the direct link.

Our problem setup is different from these papers in two major aspects: 1) The objective of the aforementioned references is the maximization of the instantaneous capacity; our problem, however, can be categorized as a single-dimensional beamforming problem, where we optimize the source beamforming vector, destination combining vector, and relay weighting matrix to maximize the received signal-to-noise ratio. 2) These papers assume either no CSI or complete CSI. Our work, however, focuses on a "limited feedback" scenario, where the receiver end of a link sends the properly quantized channel state information back to the transmitter side.

It should be noted that we assume perfect channel state information at the receiver sides of the links (CSIR), i.e. the relay knows source-relay channel, and the destination knows source-destination and relay-destination channels. Although

the source-relay channel information can be made available to the destination, e.g. by forwarding the corresponding training sequence to destination [14], we do not assume such knowledge in this paper. An extension of the current work with this additional CSI assumption is presented in [15].

The structure of this paper is as follows. The analysis starts by first ignoring the direct link and deriving the corresponding optimal source and relay beamforming vectors. Based on the distributions of these vectors for Rayleigh fading channels, we justify use of Grassmannian codebooks in the corresponding quantization scheme.

In the second part of the paper, we include the direct link in the system model. As expected, one needs to know both source-relay and source-destination channels to determine the optimal source beamforming vector. We assume that such knowledge is available (for example at the relay). We then derive the optimization problem that characterizes the optimal source beamforming vector. Although this problem does not appear to have an analytic solution, we are able to show that for i.i.d. Rayleigh channels the solution to this problem is uniformly distributed on the unit sphere, based on which, the appropriateness of the Grassmannian quantizer can be shown analytically.

In the next step, the assumption of complete knowledge of the source-relay and source-destination channels is relaxed. We focus on a scheme where the destination quantizes the source-destination channel and sends it to the relay, and justify the use of the Grassmannian quantizer. Finally, a modified quantizer is presented, which requires fewer number of feedback bits and performs very close to the original quantizer.

The remainder of this paper is organized as follows. Section II provides a brief review of the connection between the Grassmannian line packing problem and MIMO beamforming codebook design. Sections III and IV present the optimal unquantized schemes and the corresponding quantization schemes, with and without the direct link. The simulation results are discussed in Section V. Finally, Section VI concludes the paper.

Notations: $\mathbb{R}$ and $\mathbb{C}$ denote the real and complex spaces. Bold upper case and lower case letters denote matrices and vectors respectively. $\mathbf{I}_n$ denotes the $n \times n$ identity matrix. $\mathcal{U}^m$ denotes the set of all unitary matrices in $\mathbb{C}^{m \times m}$. $|\cdot|$ and $\|\cdot\|$ denote the absolute value of a scalar and the Euclidean norm of a vector. $\|\cdot\|_F$ denotes the Frobenius norm of a matrix. $(\cdot)^T$ and $(\cdot)^H$ denote the transpose and Hermitian of a matrix. The notation $\mathbf{\Phi} = \texttt{diag}(\phi_1, \phi_2, \cdots, \phi_r) \in \mathbb{C}^{m \times n}$ with $r = \min\{m, n\}$ is used to denote a rectangular diagonal matrix with ϕ_i's on its main diagonal. $\mathcal{CN}(0, \mathbf{\Sigma})$ represents a circularly symmetric complex Gaussian distribution with zero mean and covariance matrix $\mathbf{\Sigma}$. Finally, $\mathrm{E}\{\cdot\}$ denotes the expectation operation.

II. GRASSMANNIAN LINE PACKING: A REVIEW

The connection between the Grassmannian line packing problem and beamforming codebook design for a Rayleigh fading channel has been independently observed in [5] and [6]. Consider the MIMO channel in Fig. 1, where $\sqrt{P}\mathbf{H} \in \mathbb{C}^{n \times m}$
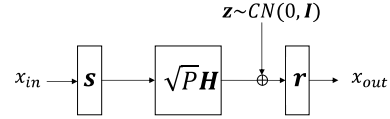

Fig. 1. Beamforming over MIMO channel.

models the flat fading channel matrix, $\mathbf{z} \sim \mathcal{CN}(0, \mathbf{I}_n)$ is the destination input noise vector, and m, n are the number of antennas at source and destination respectively. It can be shown that the source beamforming vector which maximizes the received SNR is the dominant right singular vector of $\mathbf{H}$, i.e. the right singular vector corresponding to the largest singular value of $\mathbf{H}$. The resulting (maximized) SNR is equal to $\gamma^\star = P\sigma_1^2$, where σ_1 is the largest singular value of $\mathbf{H}$. For the Rayleigh fading channel matrix $\mathbf{H}$, the singular vectors have been shown to be uniformly distributed on the unit sphere in $\mathbb{C}^m$ (see [5], [16]). Based on this fact, the authors of [5] showed that the beamforming codebook can be designed according to the criterion of the Grassmannian line packing problem, which is described below.

Consider the complex space $\mathbb{C}^m$. Define the distance of two unit vectors to be the sine of the angle between them:

$$d(\mathbf{w}_1, \mathbf{w}_2) = \sqrt{1 - |\mathbf{w}_1^H \mathbf{w}_2|^2}, \tag{1}$$

which is known as the chordal distance. For a codebook $\mathbf{C} = \{\mathbf{w}_1, \mathbf{w}_2, \cdots, \mathbf{w}_N\}$ with N distinct unit vectors as codewords, define $\delta(\mathbf{C})$ as the minimum distance of the codebook. For a fixed dimension m and codebook size N, the Grassmannian line packing problem [4] is that of finding a codebook $\mathbf{C}$ of size N, i.e. N codewords, with the largest minimum distance.

For the problem setup in Fig. 1, consider a beamforming codebook $\mathbf{C}(N, \delta)$ of size N and minimum distance δ. The destination chooses the vector in this codebook that maximizes the SNR and sends the label of this vector back to the source. Let $\tilde{\gamma}$ denote the resulting SNR: $\tilde{\gamma} = \max_{\mathbf{w} \in \mathbf{C}} P\|\mathbf{H}\mathbf{w}\|^2$. The authors of [5] used the distribution of optimal source beamforming vector to bound the average SNR loss as:

$$\mathrm{E}\{\gamma^\star\} - \mathrm{E}\{\tilde{\gamma}\} \le P\mathrm{E}\{\sigma_1^2\}\left(1 - N\left(\frac{\delta}{2}\right)^{2(m-1)}\left(1 - \frac{\delta^2}{4}\right)\right) \tag{2}$$

where m is the space dimension (number of source antennas). The upper bound in (2) is a decreasing function of δ, for any $m > 1$. Therefore, to minimize the upper bound of the SNR loss, we should maximize the minimum distance of the codebook, which is the same criterion used in the definition of the Grassmannian line packing problem.

III. MIMO AF RELAY CHANNEL WITHOUT THE DIRECT LINK

In this section, we consider the MIMO AF relay channel without the direct link and derive the optimal source beamforming vector, relay weighting matrix, and destination combining vectors in Section III.A. Next, we present the quantization scheme for the limited feedback case in Section III.B.
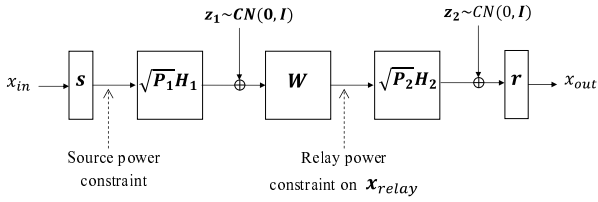
Fig. 2.   MIMO AF relay channel without the direct link.

Fig. 3.   (a) Optimal unquantized scheme without the direct link. (b) Quantized scheme without the direct link.

A. Optimal Unquantized Scheme

Consider the MIMO AF relay system in Fig. 2, where the direct link between source and destination is ignored. The source, the relay, and the destination are equipped with m, n and l antennas respectively. The matrices $\sqrt{P_1}\mathbf{H}_1 \in \mathbb{C}^{n\times m}$ and $\sqrt{P_2}\mathbf{H}_2 \in \mathbb{C}^{l\times n}$ model the flat fading channels of the source-relay and relay-destination links respectively. The coefficients P_1 and P_2 are referred to as the source-relay and relay-destination link SNR's. The source uses the vector $\mathbf{s}$ for beamforming. The relay multiplies its noisy received signal by the matrix $\mathbf{W} \in \mathbb{C}^{n\times n}$ and sends it to the destination. The destination recovers its symbol using the combining vector $\mathbf{r}$. We assume power constraints at the source and the relay outputs to be equal to 1, without loss of generality.

The problem is to find the optimal $\mathbf{s}$, $\mathbf{W}$ and $\mathbf{r}$, to maximize the SNR at the destination output, subject to power constraints at the source and at the relay. For this problem setup, a reasonable solution is "matching", as described below. The source should map its symbol to the dominant right singular vector of $\mathbf{H}_1$; the relay should absorb maximum signal power by matching to the effective channel $\mathbf{H}_1\mathbf{s}$, scale the resulting (noisy) signal to meet its power constraint and transmit it through the dominant right singular vector of $\mathbf{H}_2$. Finally, the destination should match to the relay-destination link by using the dominant left singular vector of $\mathbf{H}_2$ as the combining vector. This matching solution is depicted in Fig. 3a, in which

$$\mathbf{H}_1 = \mathbf{A}\mathbf{\Phi}\mathbf{B}^H, \quad \mathbf{H}_2 = \mathbf{F}\mathbf{\Psi}\mathbf{G}^H, \qquad (3)$$

are the singular value decompositions[1] (SVD) of $\mathbf{H}_1$ and $\mathbf{H}_2$, and

$$\mathbf{A}=[\mathbf{a}_1|\mathbf{a}_2|\cdots|\mathbf{a}_n] \in \mathcal{U}^n, \quad \mathbf{F}=[\mathbf{f}_1|\mathbf{f}_2|\cdots|\mathbf{f}_l] \in \mathcal{U}^l,$$
$$\mathbf{B}=[\mathbf{b}_1|\mathbf{b}_2|\cdots|\mathbf{b}_m] \in \mathcal{U}^m, \quad \mathbf{G}=[\mathbf{g}_1|\mathbf{g}_2|\cdots|\mathbf{g}_n] \in \mathcal{U}^n,$$
$$\mathbf{\Phi}=\mathtt{diag}\{\phi_1,\phi_2,\cdots,\phi_{r_1}\}, \quad \mathbf{\Psi}=\mathtt{diag}\{\psi_1,\psi_2,\cdots,\psi_{r_2}\},$$

where $r_1 = \min\{n,m\}$, $r_2 = \min\{l,n\}$. Although matching seems to be the natural solution to this problem, showing that the optimal $\mathbf{W}$ is a rank one matrix and that matching is optimal is not trivial. This is mainly due to the noise amplification at the relay, which generates colored noise at the destination input.

The optimality of matching at the relay is shown in [9], where the authors derive the optimal relay matrix with the objective of rate maximization. By defining $\mathbf{H}_1\mathbf{s}$ and $\mathbf{H}_2^H\mathbf{r}$ at the effective source-relay and relay-destination channels, the rate maximization is equivalent to the maximization of the

[1]In all SVDs throughput this paper, the singular values are arranged in descending order.
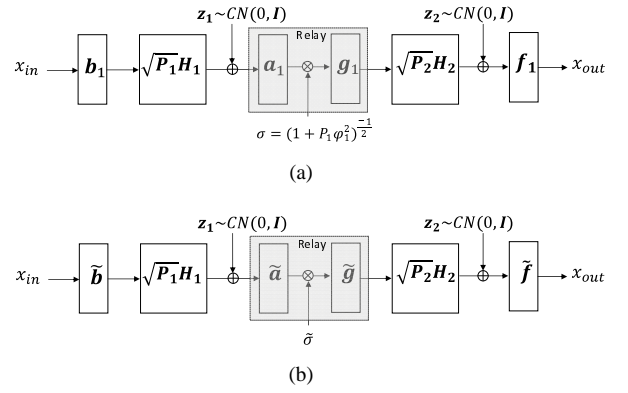
received SNR, and one can apply the results of [9] to show that the optimal relay matrix should match to these effective channels from left and right. In this paper, we present an alternative derivation of the optimal relay matrix using more direct convexity arguments.

The relay and destination output signals in Fig. 2 are:

$$x_{out} = \sqrt{P_1 P_2}\mathbf{r}^H\mathbf{H}_2\mathbf{W}\mathbf{H}_1\mathbf{s}x_{in} + \sqrt{P_2}\mathbf{r}^H\mathbf{H}_2\mathbf{W}\mathbf{z}_1 + \mathbf{r}^H\mathbf{z}_2,$$
$$\mathbf{x}_{relay} = \sqrt{P_1}\mathbf{W}\mathbf{H}_1\mathbf{s}x_{in} + \mathbf{W}\mathbf{z}_1,$$

where $\mathbf{z}_1 \sim \mathcal{CN}(0,\mathbf{I})$ and $\mathbf{z}_2 \sim \mathcal{CN}(0,\mathbf{I})$ are the complex Gaussian noise vectors at the relay and the destination. The source power constraint is satisfied by letting $\mathrm{E}\{|x_{in}|^2\} = 1$ and $\|\mathbf{s}\| = 1$. Also, the relay power constraint, which limits the power of the amplified signal and noise, can be expressed as: $\mathrm{E}\{\|\mathbf{x}_{relay}\|^2\} = P_1\|\mathbf{W}\mathbf{H}_1\mathbf{s}\|^2 + \|\mathbf{W}\|_F^2 = 1$. Finally, the received SNR can be written as:

$$\gamma = \frac{P_1 P_2 \left|\mathbf{r}^H\mathbf{H}_2\mathbf{W}\mathbf{H}_1\mathbf{s}\right|^2}{P_2\left\|\mathbf{W}^H\mathbf{H}_2^H\mathbf{r}\right\|^2 + \|\mathbf{r}\|^2},$$

where we can assume $\|\mathbf{r}\| = 1$, without loss of generality. The optimization problem can be summarized as:

$$\max \quad \frac{P_1 P_2 \left|\mathbf{r}^H\mathbf{H}_2\mathbf{W}\mathbf{H}_1\mathbf{s}\right|^2}{P_2\left\|\mathbf{W}^H\mathbf{H}_2^H\mathbf{r}\right\|^2 + 1} \qquad (4)$$

$$\text{s.t.} \quad \begin{cases} \|\mathbf{s}\| = \|\mathbf{r}\| = 1, \quad P_1\|\mathbf{W}\mathbf{H}_1\mathbf{s}\|^2 + \|\mathbf{W}\|_F^2 = 1 \\ \mathbf{W} \in \mathbb{C}^{n\times n}, \ \mathbf{s} \in \mathbb{C}^m, \ \mathbf{r} \in \mathbb{C}^l. \end{cases}$$

Theorem 1: The optimal values of the source beamforming vector, the destination combining vector, and the relay weighting matrix for the problem in (4) are given by: $\mathbf{s}^\star=\mathbf{b}_1$, $\mathbf{r}^\star=\mathbf{f}_1$, $\mathbf{W}^\star=\sigma\mathbf{g}_1\mathbf{a}_1^H$, where we have used the SVD equations in (3), and $\sigma=\left(1+P_1\phi_1^2\right)^{-\frac{1}{2}}$. Note that $\mathbf{W}^\star$ is a rank one matrix.

Proof: The optimization is accomplished in two steps.

Step1) Maximization with respect to $\mathbf{W}$: Fix $\mathbf{s}$ and $\mathbf{r}$, and define $\mathbf{h}_1=\sqrt{P_1}\mathbf{H}_1\mathbf{s}$ and $\mathbf{h}_2=\sqrt{P_2}\mathbf{H}_2^H\mathbf{r}$. Let $c_1=\|\mathbf{h}_1\|$ and $c_2=\|\mathbf{h}_2\|$. Consider $\mathbf{W}=\mathbf{U}\mathbf{\Sigma}\mathbf{V}^H$ as the SVD of $\mathbf{W}$, where $\mathbf{U},\mathbf{V}\in\mathcal{U}^n$ and $\mathbf{\Sigma}=\mathtt{diag}\{\sigma_1,\sigma_2,\cdots,\sigma_n\}$. The calculations provided below perform the optimization with respect to $\mathbf{U}$, $\mathbf{V}$ and $\mathbf{\Sigma}$.

Define $\mathbf{x} = \mathbf{V}^H\mathbf{h}_1$ and $\mathbf{y} = \mathbf{U}^H\mathbf{h}_2$, which impose the constraints $\|\mathbf{x}\| = \|\mathbf{h}_1\| = c_1$ and $\|\mathbf{y}\| = \|\mathbf{h}_2\| = c_2$ on $\mathbf{x} =$

$[x_1, x_2, \cdots, x_n]^T$, $\mathbf{y} = [y_1, y_2, \cdots, y_n]^T \in \mathbb{C}^n$. The problem in (4) can now be rephrased[2] as a maximization with respect to $\mathbf{x}$, $\mathbf{y}$ and $\boldsymbol{\Sigma}$:

$$\max \quad \frac{|\mathbf{y}^H \boldsymbol{\Sigma} \mathbf{x}|^2}{\|\boldsymbol{\Sigma} \mathbf{y}\|^2 + 1} \tag{5}$$

$$\text{s.t.} \quad \begin{cases} \|\mathbf{x}\| = c_1, \ \|\mathbf{y}\| = c_2, \ \sum_{i=1}^n \sigma_i^2 |x_i|^2 + \sum_{i=1}^n \sigma_i^2 = 1 \\ x_i, \ y_i \in \mathbb{C}, \ \sigma_i \geq 0, \ i = 1, 2, \cdots, n \end{cases}$$

In problem (5), the constraint on $\mathbf{y}$ does not depend on $\mathbf{x}$ and $\boldsymbol{\Sigma}$. Therefore, we can easily derive the optimal $\mathbf{y}$ in terms of $\mathbf{x}$ and $\boldsymbol{\Sigma}$:[3]

$$\mathbf{y} = c \left(\boldsymbol{\Sigma}^2 + \frac{1}{c_2^2} \mathbf{I} \right)^{-1} \boldsymbol{\Sigma} \mathbf{x}, \tag{6}$$

where the scalar c is chosen to satisfy $\|\mathbf{y}\| = c_2$. By substituting (6) in the objective of (5), we get:

$$\gamma = \mathbf{x}^H \boldsymbol{\Sigma} \left(\boldsymbol{\Sigma}^2 + \frac{1}{c_2^2} \mathbf{I} \right)^{-1} \boldsymbol{\Sigma} \mathbf{x}. \tag{7}$$

For the next step, we find an upper bound for the SNR expression in (7) and present the optimal values of $\mathbf{x}$ and $\boldsymbol{\Sigma}$ that achieve this upper bound. Considering (7), our maximization problem is equivalent to:

$$\max \quad \sum_{i=1}^n |x_i|^2 \frac{\sigma_i^2}{\sigma_i^2 + \frac{1}{c_2^2}} \tag{8}$$

$$\text{s.t.} \quad \begin{cases} \|\mathbf{x}\| = c_1, \ \sum_{i=1}^n \sigma_i^2 |x_i|^2 + \sum_{i=1}^n \sigma_i^2 = 1 \\ x_i \in \mathbb{C}, \ \sigma_i \geq 0, \ i = 1, 2, \cdots, n \end{cases}$$

Define $\beta_i = \frac{|x_i|^2}{c_1^2} = \frac{|x_i|^2}{\|\mathbf{x}\|^2}$. Clearly, $0 \leq \beta_i \leq 1$ and $\sum_{i=1}^n \beta_i = 1$. Now, consider the objective function in (8):

$$\begin{aligned} \gamma &= \sum_i |x_i|^2 \frac{\sigma_i^2}{\sigma_i^2 + 1/c_2^2} = c_1^2 \sum_i \frac{|x_i|^2}{c_1^2} \frac{\sigma_i^2}{\sigma_i^2 + 1/c_2^2} \\ &= c_1^2 \sum_i \beta_i \frac{\sigma_i^2}{\sigma_i^2 + 1/c_2^2} \leq c_1^2 \frac{\sum_i \beta_i \sigma_i^2}{\sum_i \beta_i \sigma_i^2 + 1/c_2^2} \quad (9) \\ &= c_1^2 \frac{\sum_i \sigma_i^2 |x_i|^2}{\sum_i \sigma_i^2 |x_i|^2 + c_1^2/c_2^2} = c_1^2 \frac{\zeta}{\zeta + c_1^2/c_2^2}, \quad (10) \end{aligned}$$

where $\zeta \stackrel{def}{=} \sum_i \sigma_i^2 |x_i|^2$. The inequality in (9) is a result of the concavity of the function $\frac{t}{t + 1/c_2^2}$ for $t \geq 0$. Now, from the second constraint of the problem (8), we have:

$$1 - \sum_i \sigma_i^2 = \sum_i \sigma_i^2 |x_i|^2 \leq \sum_i \sigma_i^2 \sum_i |x_i|^2 = c_1^2 \sum_i \sigma_i^2.$$

Therefore, $\sum_i \sigma_i^2 \geq \frac{1}{1 + c_1^2}$ and by applying the same constraint, we can bound ζ:

$$\zeta = \sum_i \sigma_i^2 |x_i|^2 = 1 - \sum_i \sigma_i^2 \leq \frac{c_1^2}{1 + c_1^2}. \tag{11}$$

---

[2]The power constraint of the relay is computed as follows:

$P_1 \|\mathbf{W} \mathbf{H}_1 \mathbf{s}\|^2 + \|\mathbf{W}\|_F^2 = \|\mathbf{U} \boldsymbol{\Sigma} \mathbf{V}^H \mathbf{h}_1\|^2 + \|\mathbf{W}\|_F^2 = \|\boldsymbol{\Sigma} \mathbf{x}\|^2 + \sum \sigma_i^2 = \sum \sigma_i^2 |x_i|^2 + \sum \sigma_i^2$.

[3]We have:

$$\max_{\|\mathbf{y}\| = c_2} \frac{|\mathbf{y}^H \boldsymbol{\Sigma} \mathbf{x}|^2}{\|\boldsymbol{\Sigma} \mathbf{y}\|^2 + 1} = \max_{\mathbf{y}} \frac{\mathbf{y}^H \boldsymbol{\Sigma} \mathbf{x} (\boldsymbol{\Sigma} \mathbf{x})^H \mathbf{y}}{\mathbf{y}^H \left(\boldsymbol{\Sigma}^2 + \frac{1}{c_2^2} \mathbf{I}_n \right) \mathbf{y}}.$$

Equations (6) and (7) are derived using the following result from linear algebra:

For any vector $\mathbf{a}$ and positive definite matrix $\mathbf{B}$, we have $\frac{\mathbf{y}^H \mathbf{a} \mathbf{a}^H \mathbf{y}}{\mathbf{y}^H \mathbf{B} \mathbf{y}} \leq \mathbf{a}^H \mathbf{B}^{-1} \mathbf{a}$, with equality if $\mathbf{y} = c \mathbf{B}^{-1} \mathbf{a}$, for arbitrary nonzero scalar c.

Finally, by combining (10) and (11), and noting that (10) is increasing in ζ, we have the following upper bound for the SNR:

$$\gamma \leq \frac{c_1^2 c_2^2}{1 + c_1^2 + c_2^2}. \tag{12}$$

By reconsidering the problem in (5), it is easy to check that the following choices of $\mathbf{x}$, $\boldsymbol{\Sigma}$ and $\mathbf{y}$ satisfy the constraints and achieve the upper bound in (12).

$$\mathbf{x} = [c_1, 0, \cdots, 0]^T, \ \mathbf{y} = [c_2, 0, \cdots, 0]^T, \ \boldsymbol{\Sigma} = \text{diag}\{\sigma, 0, \cdots, 0\}, \tag{13}$$

where $\sigma = \left(1 + c_1^2 \right)^{-\frac{1}{2}}$. Recalling the definitions of $\mathbf{x}$, $\mathbf{y}$, c_1 and c_2, the optimal values in (13) can be achieved by:

$$\begin{aligned} \mathbf{V} &= [\hat{\mathbf{h}}_1 | \mathbf{v}_1 | \cdots | \mathbf{v}_{n-1}], \quad \mathbf{U} = [\hat{\mathbf{h}}_2 | \mathbf{u}_1 | \cdots | \mathbf{u}_{n-1}], \\ \boldsymbol{\Sigma} &= \text{diag}\{\sigma, 0, \cdots, 0\} \end{aligned} \tag{14}$$

where $\hat{\mathbf{h}}_1 = \frac{\mathbf{h}_1}{\|\mathbf{h}_1\|}$, $\hat{\mathbf{h}}_2 = \frac{\mathbf{h}_2}{\|\mathbf{h}_2\|}$ and $\sigma = (1 + \|\mathbf{h}_1\|^2)^{-\frac{1}{2}}$. Here $\{\mathbf{v}_1, \cdots, \mathbf{v}_{n-1}\}$ and $\{\mathbf{u}_1, \cdots, \mathbf{u}_{n-1}\}$ are arbitrary orthonormal basis for the null-spaces of the $\mathbf{h}_1$ and $\mathbf{h}_2$ respectively. To summarize, having $\mathbf{s}$ and $\mathbf{r}$ fixed, the optimal structure of $\mathbf{W} = \mathbf{U} \boldsymbol{\Sigma} \mathbf{V}^H$ and the corresponding SNR value are:

$$\mathbf{W} = \sigma \hat{\mathbf{h}}_2 \hat{\mathbf{h}}_1^H \tag{15}$$

$$\gamma = \frac{\|\mathbf{h}_1\|^2 \|\mathbf{h}_2\|^2}{1 + \|\mathbf{h}_1\|^2 + \|\mathbf{h}_2\|^2}, \tag{16}$$

where $\sigma = (1 + \|\mathbf{h}_1\|^2)^{-\frac{1}{2}}$, $\mathbf{h}_1 = \sqrt{P_1} \mathbf{H}_1 \mathbf{s}$, and $\mathbf{h}_2 = \sqrt{P_2} \mathbf{H}_2^H \mathbf{r}$. This concludes the maximization with respect to $\mathbf{W}$.

Step 2) Maximization with respect to $\mathbf{s}$ and $\mathbf{r}$:
From (16) we see that γ is increasing both in $\|\mathbf{h}_1\|$ and $\|\mathbf{h}_2\|$. Therefore, for maximizing the SNR, we should maximize $\|\mathbf{h}_1\|$ and $\|\mathbf{h}_2\|$, subject to $\|\mathbf{s}\| = \|\mathbf{r}\| = 1$. Considering the definitions of $\mathbf{h}_1$ and $\mathbf{h}_2$, the optimal value is achieved by letting $\mathbf{s}$ be the dominant right singular vector of $\mathbf{H}_1$ and $\mathbf{r}$ be the dominant left singular vector of $\mathbf{H}_2$. ∎

Substituting the optimal solution, found in Theorem 1, in equation (16) reveals the optimal SNR:

$$\gamma^\star = \frac{\gamma_1^\star \gamma_2^\star}{1 + \gamma_1^\star + \gamma_2^\star}, \tag{17}$$

where

$$\begin{aligned} \gamma_1^\star &= \max_{\|\mathbf{s}\| = 1} P_1 \|\mathbf{H}_1 \mathbf{s}\|^2 = P_1 \phi_1^2, \\ \gamma_2^\star &= \max_{\|\mathbf{r}\| = 1} P_2 \|\mathbf{H}_2^H \mathbf{r}\|^2 = P_2 \psi_1^2. \end{aligned} \tag{18}$$

The optimal structure, outlined in Theorem I, can be achieved if the relay informs the source of $\mathbf{b}_1$ (dominant right singular vector of $\mathbf{H}_1$) and the destination informs the relay of $\mathbf{g}_1$ (dominant right singular vector of $\mathbf{H}_2$). In the next section, we characterize the codebooks that should be used for quantizing these optimal beamforming vectors.

B. Quantization Scheme

Fig. 3b shows the quantization scheme. We assume that the source beamforming vector $\tilde{\mathbf{b}}$ belongs to a codebook $\mathbf{C}_1(N_1, \delta_1)$ shared between the source and the relay. Similarly,

the relay beamforming vector $\tilde{\mathbf{g}}$ belongs to a possibly different codebook $\mathbf{C}_2(N_2, \delta_2)$, which is shared between the relay and the destination. The relay and the destination use $\tilde{\mathbf{a}}$ and $\tilde{\mathbf{f}}$ for receive combining respectively. All transmit/receive vectors $\tilde{\mathbf{a}}$, $\tilde{\mathbf{b}}$, $\tilde{\mathbf{f}}$ and $\tilde{\mathbf{g}}$ are assumed to be unit norm, and $\sigma = (1 + P_1|\tilde{\mathbf{a}}^H\mathbf{H}_1\tilde{\mathbf{b}}|^2)^{-1/2}$ in order to satisfy the relay power constraint. The received SNR of the quantized scheme can be easily shown to be equal to:

$$\gamma = \frac{\gamma_1\gamma_2}{1 + \gamma_1 + \gamma_2}, \qquad (19)$$

where $\gamma_1 = P_1\left|\tilde{\mathbf{a}}^H\mathbf{H}_1\tilde{\mathbf{b}}\right|^2$ and $\gamma_2 = P_2\left|\tilde{\mathbf{f}}^H\mathbf{H}_2\tilde{\mathbf{g}}\right|^2$ are the received SNR's of the source-relay and relay-destination links. As γ is increasing both in γ_1 and γ_2, we should maximize these quantities to maximize γ. This is accomplished by letting $\tilde{\mathbf{a}}$ and $\tilde{\mathbf{f}}$ to be matched to $\mathbf{H}_1\tilde{\mathbf{b}}$ and $\mathbf{H}_2\tilde{\mathbf{g}}$, and choosing $\tilde{\mathbf{b}}$ and $\tilde{\mathbf{g}}$ based on: $\tilde{\mathbf{b}} = \arg\max_{\mathbf{w}\in\mathbf{C}_1} P_1\|\mathbf{H}_1\mathbf{w}\|^2$ and $\tilde{\mathbf{g}} = \arg\max_{\mathbf{w}\in\mathbf{C}_2} P_2\|\mathbf{H}_2\mathbf{w}\|^2$. The corresponding received SNR values are

$$\tilde{\gamma}_1 = \max_{\mathbf{w}\in\mathbf{C}_1} P_1\|\mathbf{H}_1\mathbf{w}\|^2, \quad \tilde{\gamma}_2 = \max_{\mathbf{w}\in\mathbf{C}_2} P_2\|\mathbf{H}_2\mathbf{w}\|^2, \qquad (20)$$

and the maximum received SNR of the quantized scheme $\tilde{\gamma}$ can be computed by substituting these quantities in (19):

$$\tilde{\gamma} = \frac{\tilde{\gamma}_1\tilde{\gamma}_2}{1 + \tilde{\gamma}_1 + \tilde{\gamma}_2}. \qquad (21)$$

In Appendix II.A, we use the distributions of the optimal beamforming vectors $\mathbf{b}_1$ and $\mathbf{g}_1$ for Rayleigh channels to compute the following upper bound for the total loss in SNR caused by quantization.

$$\mathrm{E}\{\gamma^\star\} - \mathrm{E}\{\tilde{\gamma}\} \leq 2mnP_1\left(1 - N_1\left(\frac{\delta_1}{2}\right)^{2(m-1)}\left(1 - \frac{\delta_1}{2}\right)\right) + 2nlP_2\left(1 - N_2\left(\frac{\delta_2}{2}\right)^{2(n-1)}\left(1 - \frac{\delta_2}{2}\right)\right). \qquad (22)$$

This upper bound is decreasing in δ_1 and δ_2 for any $m > 1$ and $n > 1$. Therefore, to minimize this upper bound, we should maximize the minimum distances δ_1 and δ_2. This proves the efficiency of the Grassmannian codebooks for quantizing the optimal beamforming vectors.

IV. MIMO AF RELAY CHANNEL WITH THE DIRECT LINK

In this section, the direct link is included in the system model (Fig. 4). The optimal unquantized scheme is derived in Section IV.A and the quantization scheme is presented in Section IV.B. Finally, in Section IV.C we introduce a modified quantized scheme, which significantly reduces the number of feedback bits with a negligible degradation in the system performance.

A. Optimal Unquantized Scheme

Consider the half-duplex MIMO relay channel in Fig. 4. At the first time slot, the relay is silent and the destination receives its symbol. At the second time slot, the source is silent and the relay amplifies and forwards its signal (received in the first
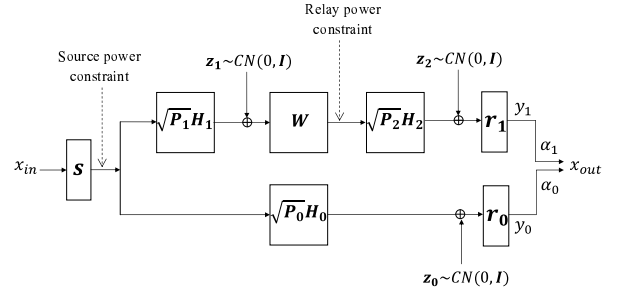

Fig. 4. MIMO AF relay channel with the direct link.

time slot) to the destination. The destination has access to two received symbols y_0 and y_1 separated in time:

$$y_0 = \sqrt{P_0}\mathbf{r}_0^H\mathbf{H}_0\mathbf{s}x + \mathbf{r}_0^H\mathbf{z}_0,$$
$$y_1 = \sqrt{P_1P_2}\mathbf{r}_1^H\mathbf{H}_2\mathbf{W}\mathbf{H}_1\mathbf{s}x + \mathbf{r}_1^H\left(\sqrt{P_2}\mathbf{H}_2\mathbf{W}\mathbf{z}_1 + \mathbf{z}_2\right).$$

The destination combines y_0 and y_1 to compute the output symbol $x_{out} = \alpha_0 y_0 + \alpha_1 y_1$. To maximize the total received SNR, α_0 and α_1 should be MMSE (or scaled MRC) coefficients,[4] which result in the total received SNR:

$$\gamma = \gamma_0 + \gamma_r, \qquad (23)$$

where γ_0 and γ_r are the received SNR values of the direct link and the source-relay-destination link. Therefore, the total SNR is maximized if γ_0 and γ_r are maximized. The only common parameter in maximizing these two quantities is the source beamforming vector $\mathbf{s}$. By fixing $\mathbf{s}$ and following the same steps as in Section III, the optimal values of other parameters can be easily derived, as shown in Fig. 5a. The corresponding received SNR's of the direct link and relay link are:

$$\gamma_0 = P_0\|\mathbf{H}_0\mathbf{s}\|^2, \quad \gamma_r = \frac{\gamma_1\gamma_2^\star}{1 + \gamma_1 + \gamma_2^\star}, \qquad (24)$$

where $\gamma_1 = P_1\|\mathbf{H}_1\mathbf{s}\|^2$ and $\gamma_2^\star = P_2\|\mathbf{H}_2\mathbf{g}_1\|^2 = P_2\psi_1^2$. By combining (23) and (24) the total received SNR is:

$$\gamma = \frac{P_1\|\mathbf{H}_1\mathbf{s}\|^2\gamma_2^\star}{1 + P_1\|\mathbf{H}_1\mathbf{s}\|^2 + \gamma_2^\star} + P_0\|\mathbf{H}_0\mathbf{s}\|^2.$$

Therefore, the optimal $\mathbf{s}$ can be expressed as:

$$\mathbf{s}^\star = \arg\max_{\|\mathbf{s}\|=1} \frac{\|\mathbf{H}_1\mathbf{s}\|^2}{\|\mathbf{H}_1\mathbf{s}\|^2 + \lambda} + \mu\|\mathbf{H}_0\mathbf{s}\|^2, \qquad (25)$$

where $\lambda = \frac{1+\gamma_2^\star}{P_1}$ and $\mu = \frac{P_0}{\gamma_2^\star}$. The corresponding total received SNR is:

$$\gamma^\star = \frac{\gamma_1^\star\gamma_2^\star}{1 + \gamma_1^\star + \gamma_2^\star} + \gamma_0^\star, \qquad (26)$$

where $\gamma_0^\star = P_0\|\mathbf{H}_0\mathbf{s}^\star\|^2$ and $\gamma_1^\star = P_1\|\mathbf{H}_1\mathbf{s}^\star\|^2$.

The objective function of the problem in (25) can potentially have multiple local maximum points. Moreover, the global maximum point is not unique[5]. This problem does not appear to have an analytic solution and as a result we use a numerical

---

[4]More exactly, $\alpha_0 = \frac{d_0^*}{\sigma_0^2}$ and $\alpha_1 = \frac{d_1^*}{\sigma_1^2}$, where $d_0 = \sqrt{P_0}\mathbf{r}_0^H\mathbf{H}_0\mathbf{s}$, $d_1 = \sqrt{P_1P_2}\mathbf{r}_1^H\mathbf{H}_2\mathbf{W}\mathbf{H}_1\mathbf{s}$, $\sigma_0^2 = \|\mathbf{r}_0\|^2$, and $\sigma_1^2 = P_2\left\|\mathbf{W}^H\mathbf{H}_2^H\mathbf{r}_1\right\|^2 + \|\mathbf{r}_1\|^2$.
[5]If $\mathbf{s}$ is a global maximum point, so is $e^{j\theta}\mathbf{s}$, for any $\theta \in \mathbb{R}$.
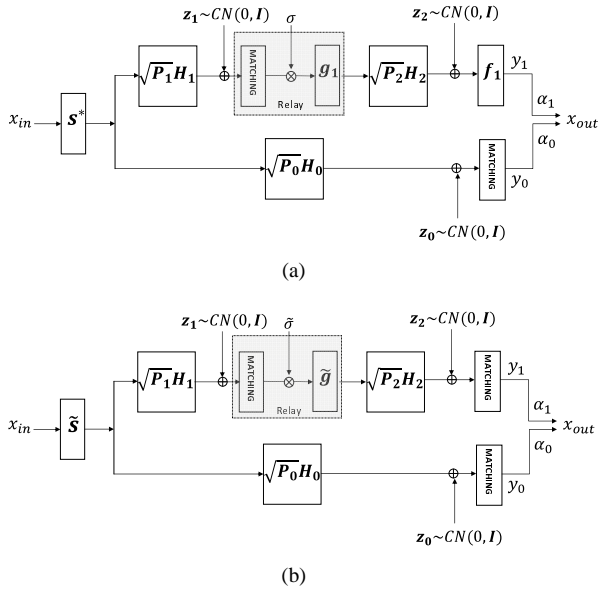
Fig. 5. (a) Optimal unquantized scheme for MIMO AF relay channel with the direct link. In the first time slot, the relay and the destination match to $\mathbf{H}_1\mathbf{s}^\star$ and $\mathbf{H}_0\mathbf{s}^\star$ respectively. (b) Quantized scheme for MIMO AF relay channel with direct link. In the first time slot, the relay and the destination match to $\mathbf{H}_1\tilde{\mathbf{s}}$ and $\mathbf{H}_0\tilde{\mathbf{s}}$ respectively. In the second time slot, the destination matches to $\mathbf{H}_2\tilde{\mathbf{g}}$.

approach to perform this optimization, which will be described in Section V. Despite the fact that we do not have a closed form expression for the solution of problem (25), we are still able to identify the distribution of the solution for Rayleigh fading channels. The main result of this section is the following theorem.

Theorem 2: For independent Rayleigh channel matrices $\mathbf{H}_0$ and $\mathbf{H}_1$, the optimal source beamforming vector $\mathbf{s}^\star$ that maximizes the total received SNR (or equivalently the objective function in (25)) is uniformly distributed on the unit sphere in $\mathbb{C}^m$, where m is the number of source antennas.

Proof: See Appendix I. ∎

The result in Theorem 2 is used in Appendix II.B to derive an SNR loss upper bound, similar to (2) and (22), which justifies use of the Grassmannian codebook for quantizing the optimal source beamforming vector $\mathbf{s}^\star$.

B. Quantization Scheme

Having identified the optimal scheme, we continue by considering the quantization scheme in Fig. 5b. The source-destination, source-relay, relay-destination and total received SNR values are given by:

$$\gamma_0 = P_0\|\mathbf{H}_0\tilde{\mathbf{s}}\|^2, \quad \gamma_1 = P_1\|\mathbf{H}_1\tilde{\mathbf{s}}\|^2, \quad \gamma_2 = P_2\|\mathbf{H}_2\tilde{\mathbf{g}}\|^2,$$
$$\gamma = \frac{\gamma_1\gamma_2}{1+\gamma_1+\gamma_2} + \gamma_0. \tag{27}$$

We need to maximize γ in (27) with respect to the source and relay beamforming vectors $\tilde{\mathbf{s}}$ and $\tilde{\mathbf{g}}$, which belong to certain codebooks $\mathbf{C}_1(N_1, \delta_1)$ and $\mathbf{C}_2(N_2, \delta_2)$ as in Section III.B. Clearly, $\tilde{\mathbf{g}}$ should be chosen to maximize γ_2:

$$\tilde{\mathbf{g}} = \arg\max_{\mathbf{w}\in\mathbf{C}_2} P_2\|\mathbf{H}_2\mathbf{w}\|^2, \tag{28}$$

and the corresponding relay-destination received SNR is: $\tilde{\gamma}_2 = \max_{\mathbf{w}\in\mathbf{C}_2} P_2\|\mathbf{H}_2\mathbf{w}\|^2$. For choosing $\tilde{\mathbf{s}}$, we need to know both $\mathbf{H}_0$ and $\mathbf{H}_1$. In reality, however, such a knowledge is not available, and the destination and relay should somehow exchange their information about $\mathbf{H}_0$ and $\mathbf{H}_1$. In this paper, we focus on a system, where the destination quantizes $\mathbf{H}_0$ and sends it to the relay[6]. It should be noted that, the only way that $\mathbf{H}_0$ contributes to the problem in (27) is through the term $\|\mathbf{H}_0\mathbf{w}\|^2$, which can be expanded as follows: $\|\mathbf{H}_0\mathbf{w}\|^2 = \sum_{i=1}^{R_0} \nu_i^2|\mathbf{e}_i^H\mathbf{w}|^2$, where ν_i's and $\mathbf{e}_i$'s are the singular values and right singular vectors of $\mathbf{H}_0$ and $R_0 = \text{rank}(\mathbf{H}_0)$. Therefore, the relay only needs to know the singular values and the right singular vectors of the direct link channel. Since our focus in this paper is on the vector quantization feedback schemes, we assume that the relay knows the singular values completely but has only access to the quantized versions of the singular vectors. For quantizing the singular vectors, the destination and the relay share a codebook $\mathbf{C}_0(N_0, \delta_0)$, which is possibly different from $\mathbf{C}_2$ (used for determining $\tilde{\mathbf{g}}$). We assume that the destination quantizes each vector $\mathbf{e}_i$ to a vector $\tilde{\mathbf{e}}_i \in \mathbf{C}_0$ that is closest to $\mathbf{e}_i$:[7]

$$\tilde{\mathbf{e}}_i = \arg\min_{\mathbf{w}\in\mathbf{C}_0} d(\mathbf{w}, \mathbf{e}_i). \tag{29}$$

Having ν_i's and $\tilde{\mathbf{e}}_i$'s at the relay, the problem of finding the source beamforming vector $\tilde{\mathbf{s}}$ can be formulated as:

$$\tilde{\mathbf{s}} = \arg\max_{\mathbf{w}\in\mathbf{C}_1} \frac{\|\mathbf{H}_1\mathbf{w}\|^2}{\|\mathbf{H}_1\mathbf{w}\|^2 + \tilde{\lambda}} + \tilde{\mu}\sum_{i=1}^{R_0} \nu_i^2|\tilde{\mathbf{e}}_i^H\mathbf{w}|^2, \tag{30}$$

where $\tilde{\lambda} = \frac{1+\tilde{\gamma}_2}{P_1}$, $\tilde{\mu} = \frac{P_0}{\tilde{\gamma}_2}$, and $\tilde{\gamma}_2 = \max_{\mathbf{w}\in\mathbf{C}_2} P_2\|\mathbf{H}_2\mathbf{w}\|^2$. The total received SNR $\tilde{\gamma}$ can be computed by substituting (28) and (30) in (27). In Appendix II.B, we use the distribution of $\mathbf{s}^\star$, given in Theorem 2, to prove the following upper bound on the average loss in SNR due to quantization:

$$\begin{aligned}
\mathbb{E}\{\gamma^\star\} - \mathbb{E}\{\tilde{\gamma}\} &\leq 4mlP_0\left(1-N_0\left(\frac{\delta_0}{2}\right)^{2(m-1)}\left(1-\frac{\delta_0}{2}\right)\right) \\
&+ 2\left(mlP_0+mnP_1\right)\left(1-N_1\left(\frac{\delta_1}{2}\right)^{2(m-1)}\left(1-\frac{\delta_1}{2}\right)\right) \\
&+ 2nlP_2\left(1-N_2\left(\frac{\delta_2}{2}\right)^{2(n-1)}\left(1-\frac{\delta_2}{2}\right)\right).
\end{aligned} \tag{31}$$

The upper bound in (31) is decreasing in $\delta_0 = \delta(\mathbf{C}_0)$ for any $m > 1$. This justifies use of the Grassmannian codebook to quantize the singular vectors of $\mathbf{H}_0$, since it has the maximum minimum distance δ_0. The same conclusion holds for $\mathbf{C}_1$ and $\mathbf{C}_2$, since the upper bound in (31) is decreasing in δ_1 and δ_2 for any $m, n > 1$. To summarize the results, all three codebooks $\mathbf{C}_0$, $\mathbf{C}_1$ and $\mathbf{C}_2$ need to be Grassmannian codebooks to minimize the upper bound of the loss in the total

---

[6]In this case, the relay becomes responsible for determining the source beamforming vector. Equivalently, we could assume that the relay quantizes $\mathbf{H}_1$ and sends it to the destination, in which case the destination would determine the source beamforming vector.

[7]Throughput this paper, the notion of "closeness" is based on the "distance" defined by the chordal distance function in (1).

received SNR. We refer to the scheme, identified by (30), as the "properly quantized scheme".

In the following, we outline the steps in determining the beamforming vectors of the "properly quantized scheme" (Fig. 5b).

1) The destination uses a Grassmannian codebook $\mathbf{C}_2$, shared between the destination and the relay, to quantize $\mathbf{g}$, the dominant right singular vector of the relay-destination channel $\mathbf{H}_2$. The label of the quantized vector is sent to the relay. The relay uses this vector for its beamforming in the second time slot. The destination also sends the SNR value $\tilde{\gamma}_2$ to the relay. This will be used in step 3.

2) The destination quantizes the right singular vectors of the source-destination channel using a separate Grassmannian codebook $\mathbf{C}_0$, which is also shared between the destination and the relay. The labels of the quantized vectors and the singular values ν_i's are sent to the relay.

3) The relay forms the objective function in (30) and maximizes it over the Grassmannian codebook $\mathbf{C}_1$, which is shared between the source and the relay. The relay sends the label of the maximizing vector to the source. The source uses this vector for its beamforming in the first time slot.

Before concluding Section IV, we introduce a modified scheme which performs very close to the "properly quantized scheme" and requires fewer number of feedback bits.

C. Modified Quantized Scheme

Consider the problem of determining the source beamforming vector for the quantization scheme in Fig. 5b, i.e. (30). There are two links between the source and the destination: the direct source-destination link and the source-relay-destination link, which we refer to as the relay link.

If the direct link is much weaker than the relay link and can be ignored, our problem reduces to the problem in Section III and the relay does not need to know anything about the direct link channel $\mathbf{H}_0$. On the other hand, if the relay link is very weak and can be ignored, the only thing that relay needs to know about $\mathbf{H}_0$ is its dominant right singular vector. Therefore, in both of these extreme cases we do not need to have any knowledge of $\mathbf{H}_0$ other than its dominant right singular vector. Based on this intuition, we propose a new scheme, referred to as the "modified quantized scheme", in which the destination only quantizes the dominant right singular vector of $\mathbf{H}_0$ and sends the corresponding codeword index (and the largest singular value ν_1) to the relay. The relay then determines the source beamforming vector by forming the following problem.

$$\tilde{\mathbf{s}}_{modified} = \arg \max_{\mathbf{w} \in \mathbf{C}_1} \frac{\|\mathbf{H}_1\mathbf{w}\|^2}{\|\mathbf{H}_1\mathbf{w}\|^2 + \tilde{\lambda}} + \tilde{\mu}\nu_1^2|\tilde{\mathbf{e}}_1^H\mathbf{w}|^2, \quad (32)$$

where $\tilde{\lambda}$ and $\tilde{\mu}$ have the same definitions as in (30).

The "modified quantized scheme" requires much fewer number of feedback bits, since it only quantizes one singular vector (see step 2 for the "properly quantized scheme"). Our
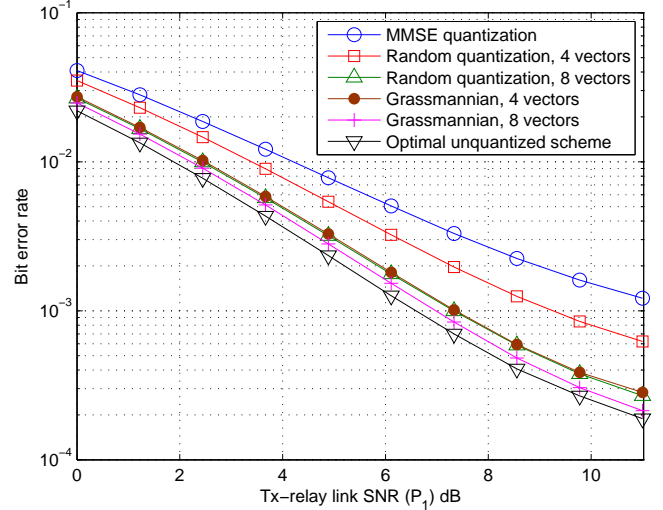

Fig. 6. Performance of the Grassmannian, MMSE, and random quantizers without the direct link. The relay-destination link SNR is fixed at 8dB.

simulation and analytical results show that the "modified quantized scheme" performs very close to the "properly quantized scheme", as we will see in Section V.

V. SIMULATION RESULTS

The simulation results are divided into two subsections. In Section V.A the direct link between the source and the destination is ignored (Fig. 2). In Section V.B, the simulation results are presented for the case where the direct link is present in the model (Fig. 4).

The general setup for the simulations is as follows. The input symbols belong to a BPSK constellation with unit power. The entries of the channel matrices are generated independently according to $\mathcal{CN}(0,1)$. To model quasi-static fading channels, the simulation time is divided to $20,000$ coherence intervals, each consisting of 200 symbols. The channels are assumed to be constant over each coherence interval and to be independent from one interval to the other.

A. MIMO AF Relay Channel without the Direct Link

In this section, all of the terminals are assumed to have two antennas ($m=n=l=2$). The relay-destination link SNR is fixed at $P_2 = 8$dB and the BER values have been recorded for different values of the source-relay link SNR P_1.

For quantization purposes, the source and the relay share a codebook $\mathbf{C}_1$ of size N_1. Similarly, the relay and destination share a codebook $\mathbf{C}_2$ of size N_2. Fig. 6 compares the performance of the "optimal unquantized scheme" (Fig. 3a) with the performance of the Grassmannian codebooks $\mathbf{C}_1$ and $\mathbf{C}_2$ of sizes $N_1=N_2=4$ or 8. The Grassmannian codebooks are adopted from [5]. The total number of the feedback bits used by the Grassmannian quantizer is $\log_2 N_1 + \log_2 N_2$ which equals 4 or 6 bits for $N_1=N_2=4$ or 8 respectively. As Fig. 6 shows, we can get very close to the optimal scheme with only a few number of bits per each coherence interval.

Fig. 6 also compares the performance of the Grassmannian quantizer with other quantization schemes. For the MMSE quantization scheme, we have used two bits to quantize each channel entry. For $m=n=l=2$, this results in $2(mn+nl)=16$ bits. For the random quantizer, we use a set of randomly selected vectors on the unit sphere as the quantization code-words. The performance of the random scheme has been averaged over ten such codebooks. As Fig. 6 shows, the Grassmannian codebooks show considerable gain as compared to the random quantizer. However, this gain decreases as the codebook sizes are increased from 4 to 8. The main advantage of the random codebooks is that they are easier to generate as compared to the Grassmannian codebooks.

B. MIMO AF Relay Channel with the Direct Link

In this section, all the stations are assumed to have three antennas ($m=n=l=3$). Fig. 7 shows the BER values for quantized and unquantized schemes as a function of direct link SNR P_0, when the source-relay and relay-destination link SNR's are fixed at $P_1 = P_2 = 2$dB. For the optimal scheme, we use the gradient descent method for determining the source beamforming vector from (25). The constraint $\|\mathbf{s}\|=1$ is eliminated by the change of variable $\mathbf{s}=\frac{\mathbf{u}}{\|\mathbf{u}\|}$.

The "modified unquantized scheme" in Fig. 7 has the same structure as the "optimal unquantized scheme" with the difference that the relay only considers the dominant singular value and singular vector of $\mathbf{H}_0$ in formulating the problem of determining the source beamforming vector. This problem is exactly the same as the problem (25), used by the optimal scheme, except that $\|\mathbf{H}_0\mathbf{s}\|^2$ is replaced by $\nu_1^2|\mathbf{e}_1^H\mathbf{s}|^2$, where ν_1 and $\mathbf{e}_1$ are the dominant singular value and right singular vector of $\mathbf{H}_0$. We can analytically show that the average SNR loss of the "modified unquantized scheme" compared with the "optimal unquantized scheme" is at most $10\log_{10}(1 + \mathrm{E}\{\nu_2^2\}/\mathrm{E}\{\nu_1^2\})$, where ν_2 is the second largest singular value of $\mathbf{H}_0$. For $m=l=3$, this bound is equal to 1.24dB. Although this bound guarantees that modified scheme would perform well, it is not necessarily tight, as the simulation results in Fig. 7 show that the actual SNR loss is very small. The "modified unquantized scheme" is the basis for the "modified quantized scheme" in Section IV.C.

The "properly quantized scheme", as discussed in Section IV.B, consists of three codebooks $\mathbf{C}_0$, $\mathbf{C}_1$ and $\mathbf{C}_2$ of sizes N_0, N_1 and N_2. Fig. 7 shows the performance of the "properly quantized scheme" with Grassmannian codebooks of sizes $N_0=N_1=N_2=8$ or 16, adopted from [17].

The figure also shows the performance of the Grassmannian codebooks with "modified quantized scheme" (see Section IV.C). This scheme shows a negligible performance degra-dation with respect to the "properly quantized scheme", but requires fewer number of feedback bits. Following the three steps for the "properly quantized scheme" in Section IV.B, one can easily check that we need a total of $(1 + R_0)b + \log_2(N_0^{R_0}N_1N_2)$ feedback bits for the "properly quantized scheme" and $2b + \log_2(N_0N_1N_2)$ bits for the "modified quantized scheme". Here b is the number of bits that a (hypo-thetical) scalar quantizer uses for quantizing a scalar value.
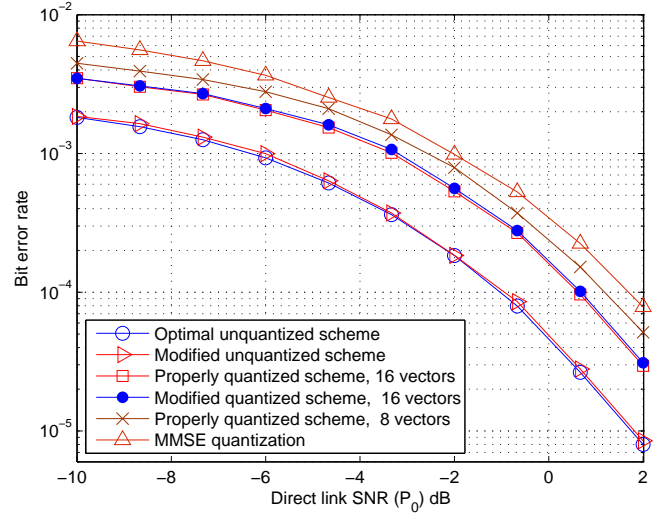

Fig. 7. Performance of the quantized and unquantized schemes with the direct link. The source-relay and relay-destination link SNR's are fixed at $P_1 = P_2 = 2$dB.
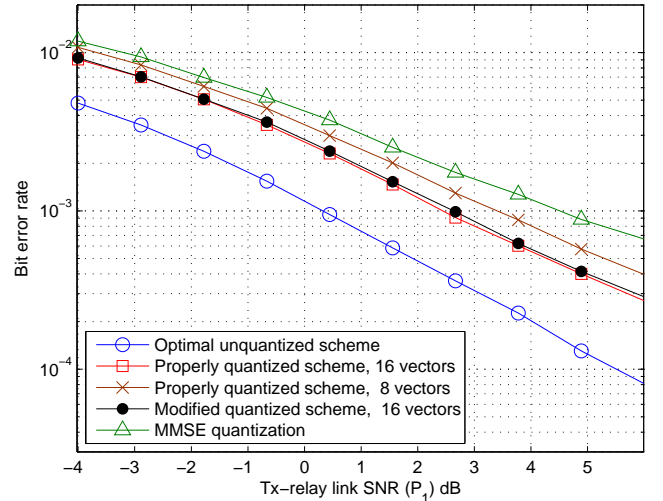

Fig. 8. Performance of the quantized and unquantized schemes with the direct link. The direct link and relay-destination link SNR are fixed at $P_0 = -4$dB and $P_2 = 2$dB.

Table I compares these values for $N=N_0=N_1=N_2=8, 16$, $m=n=l=3$, and full rank channel matrix $\mathbf{H}_0$. Fig. 7 also shows the performance of the MMSE quantizer. This quan-tizer requires $2(mn+ml+ln)$ bits for quantizing the channel matrices and b bits for quantizing $\tilde{\gamma}_2$.

Fig. 8 compares the performance of the same schemes of Fig. 7 in a different scenario. For this figure, the direct link and relay-destination link SNR's are fixed at $P_0 = -4$dB and $P_2 = 2$dB. The BER values have been recorded for different values of the source-relay link SNR P_1. Once again, we see that the performance of the "modified quantized scheme" is very close to the "properly quantized scheme".

TABLE I
COMPARISON OF THE NUMBER OF THE FEEDBACK BITS FOR DIFFERENT
QUANTIZATION SCHEMES; $m = n = l = 3$;
$N_0 = N_1 = N_2 = N = 8, 16$.

Scheme	Number of feedback bits	
	$N = 8$	$N = 16$
Properly quantized	$15 + 4b$	$20 + 4b$
Modified quantized	$9 + 2b$	$12 + 2b$
MMSE	$54 + b$	

VI. CONCLUSION

In this paper, we derived the optimal source and relay beamforming vectors to maximize the total received SNR of MIMO AF relay channel both with and without the direct link. For the limited feedback scenario, we proposed a quantization scheme based on the Grassmannian codebooks and analytically proved its efficiency. A modified quantized scheme was also presented, which performs very close to the original quantization scheme and requires considerably fewer number of feedback bits. Finally, the analytical results were verified by comparing the performances of the unquantized and quantized schemes under different scenarios.

APPENDIX I
THE DISTRIBUTION OF THE OPTIMAL BEAMFORMING VECTOR $\mathbf{s}^\star$

In this appendix, we show that there exists a solution $\mathbf{s}^\star$ to the problem (25) that is uniformly distributed on the unit sphere in $\mathbb{C}^m$, where m is the number of source antennas. The problem (25) is repeated here:

$$\mathbf{s}^\star = \arg \max_{\|\mathbf{s}\|=1} \frac{\|\mathbf{H}_1\mathbf{s}\|^2}{\|\mathbf{H}_1\mathbf{s}\|^2 + \lambda} + \mu\|\mathbf{H}_0\mathbf{s}\|^2, \qquad \text{(I.1)}$$

Consider $\mathbf{H}_0 = \mathbf{U}_0\mathbf{\Sigma}_0\mathbf{V}_0^H$ and $\mathbf{H}_1 = \mathbf{U}_1\mathbf{\Sigma}_1\mathbf{V}_1^H$ as the SVD of $\mathbf{H}_0$ and $\mathbf{H}_1$. Clearly, $\|\mathbf{H}_0\mathbf{s}\| = \|\mathbf{\Sigma}_0\mathbf{V}_0^H\mathbf{s}\|$ and $\|\mathbf{H}_1\mathbf{s}\| = \|\mathbf{\Sigma}_1\mathbf{V}_1^H\mathbf{s}\|$, since $\mathbf{U}_0$ and $\mathbf{U}_1$ are unitary matrices.

It is easy to check that $\mathbf{s}^\star = \mathbf{V}_0\eta(\mathbf{\Sigma}_0, \mathbf{\Sigma}_1, \mathbf{V}_1^H\mathbf{V}_0)$ is a solution to (I.1), where the function $\eta(\cdot, \cdot, \cdot)$ is defined to be a solution to the following problem:

$$\eta(\mathbf{\Sigma}_0, \mathbf{\Sigma}_1, \mathbf{V}_1^H\mathbf{V}_0) \overset{def}{=}$$
$$\arg \max_{\|\mathbf{t}\|=1} \frac{\|\mathbf{\Sigma}_1\mathbf{V}_1^H\mathbf{V}_0\mathbf{t}\|^2}{\|\mathbf{\Sigma}_1\mathbf{V}_1^H\mathbf{V}_0\mathbf{t}\|^2 + \lambda} + \mu\|\mathbf{\Sigma}_0\mathbf{t}\|^2. \text{ (I.2)}$$

If we fix $\mathbf{\Sigma}_0$ and $\mathbf{\Sigma}_1$, the solution $\mathbf{s}^\star$, identified above, can be expressed as a function of $\mathbf{V}_0$ and $\mathbf{V}_1$:

$$\mathbf{s}^\star = \zeta_{\mathbf{\Sigma}_0, \mathbf{\Sigma}_1}(\mathbf{V}_0, \mathbf{V}_1) \overset{def}{=} \mathbf{V}_0\eta(\mathbf{\Sigma}_0, \mathbf{\Sigma}_1, \mathbf{V}_1^H\mathbf{V}_0). \qquad \text{(I.3)}$$

Now, for any unitary matrix $\mathbf{Q}$, we have the following from (I.3):

$$\zeta_{\mathbf{\Sigma}_0, \mathbf{\Sigma}_1}(\mathbf{Q}\mathbf{V}_0, \mathbf{Q}\mathbf{V}_1) = \mathbf{Q}\zeta_{\mathbf{\Sigma}_0, \mathbf{\Sigma}_1}(\mathbf{V}_0, \mathbf{V}_1) = \mathbf{Q}\mathbf{s}^\star.$$

For a Rayleigh channel matrix $\mathbf{H}_0$, we know that the matrix $\mathbf{V}_0$ is independent of $\mathbf{\Sigma}_0$ and its distribution does not change by pre-multiplication by a unitary matrix $\mathbf{Q}$. The same argument holds for $\mathbf{H}_1$, $\mathbf{V}_1$ and $\mathbf{\Sigma}_1$. Therefore, conditioned on

$\mathbf{\Sigma}_0$ and $\mathbf{\Sigma}_1$, the matrix $\mathbf{Q}\mathbf{V}_0$ has the same distribution as $\mathbf{V}_0$, and similarly $\mathbf{Q}\mathbf{V}_1$ has the same distribution as $\mathbf{V}_1$. Since the source-destination and source-relay channels are assumed to be independent, $\mathbf{V}_0$ and $\mathbf{V}_1$ are also independent, and therefore the joint distribution of $(\mathbf{V}_0, \mathbf{V}_1)$ is also the same as the joint distribution of $(\mathbf{Q}\mathbf{V}_0, \mathbf{Q}\mathbf{V}_1)$. Hence, any arbitrary function of these pairs will have the same distribution. By applying this to the function $\zeta_{\mathbf{\Sigma}_0, \mathbf{\Sigma}_1}(\cdot)$, we conclude that $\mathbf{s}^\star = \zeta_{\mathbf{\Sigma}_0, \mathbf{\Sigma}_1}(\mathbf{V}_0, \mathbf{V}_1)$ and $\mathbf{Q}\mathbf{s}^\star = \zeta_{\mathbf{\Sigma}_0, \mathbf{\Sigma}_1}(\mathbf{Q}\mathbf{V}_0, \mathbf{Q}\mathbf{V}_1)$ have the same distribution. Since this is true for any unitary matrix $\mathbf{Q}$, we conclude that $\mathbf{s}^\star$ is uniformly distributed on the complex unit sphere, conditioned on $\mathbf{\Sigma}_0$ and $\mathbf{\Sigma}_1$.

Note that if the conditional distribution of $\mathbf{s}^\star$ is uniform, its unconditional distribution is also uniform. Moreover, the random vector $\mathbf{s}^\star$ is independent of the random matrices $\mathbf{\Sigma}_0$ and $\mathbf{\Sigma}_1$, since its conditional and unconditional distributions are the same.

APPENDIX II
PROOF OF SNR LOSS UPPER BOUNDS

In this appendix, we prove the SNR loss upper bounds of (22) and (31) using the following lemmas. The proofs of these lemmas will not be given for the sake of brevity.

Lemma 1: For nonnegative variables x_1, x_2, y_1 and y_2, we have:

$$\left| \frac{x_1 y_1}{1 + x_1 + y_1} - \frac{x_2 y_2}{1 + x_2 + y_2} \right| \le |x_1 - x_2| + |y_1 - y_2|.$$

Lemma 2: For the matrix $\mathbf{H} \in \mathbb{C}^{p \times q}$ with independent $\mathcal{CN}(0, 1)$ entries, we have: $\mathrm{E}\left\{\sum_i \sigma_i^2\right\} = pq$, where σ_i's are the singular values of $\mathbf{H}$.

Lemma 3: For any unit vectors $\mathbf{u}$, $\mathbf{v}$, and $\mathbf{w}$, we have:

$$\left| |\mathbf{u}^H\mathbf{v}|^2 - |\mathbf{v}^H\mathbf{w}|^2 \right| \le 2d(\mathbf{u}, \mathbf{w}),$$

where the distance function $d(\cdot, \cdot)$ is defined in (1).

Lemma 4: Consider the codebook $\mathbf{C} = \{\mathbf{w}_1, \mathbf{w}_2, \cdots, \mathbf{w}_N\}$ and the matrix $\mathbf{H}$ with σ_i's as its singular values. For any unit vector $\mathbf{s}$, define $\mathbf{s}_{\mathbf{C}} \in \mathbf{C}$ as the closest vector in codebook $\mathbf{C}$ to $\mathbf{s}$ and let $d_{\mathbf{C}}(\mathbf{s}) \overset{def}{=} d(\mathbf{s}, \mathbf{s}_{\mathbf{C}})$, where $d(\cdot, \cdot)$ is the distance function defined in (1). Then, we have:

$$\left| \|\mathbf{H}\mathbf{s}\|^2 - \|\mathbf{H}\mathbf{s}_{\mathbf{C}}\|^2 \right| \le 2\left(\sum_i \sigma_i^2\right) d_{\mathbf{C}}(\mathbf{s}),$$

Lemma 5: Consider the codebook $\mathbf{C}(N, \delta)$ and the function $d_{\mathbf{C}}(\cdot)$ defined in Lemma 4. For the random vector $\mathbf{s} \in \mathbb{C}^m$ uniformly distributed on the unit sphere, we have:

$$\mathrm{E}\{d_{\mathbf{C}}(\mathbf{s})\} \le 1 - N\left(\frac{\delta}{2}\right)^{2(m-1)}\left(1 - \frac{\delta}{2}\right).$$

A. Proof of the Upper Bound in (22)

The optimal unquantized scheme SNR and the quantized scheme SNR are given in (17) and (21), which are repeated here:

$$\gamma^\star = \frac{\gamma_1^\star\gamma_2^\star}{1 + \gamma_1^\star + \gamma_2^\star}, \quad \tilde{\gamma} = \frac{\tilde{\gamma}_1\tilde{\gamma}_2}{1 + \tilde{\gamma}_1 + \tilde{\gamma}_2}, \qquad \text{(II.1)}$$

where $\gamma_1^\star$, $\gamma_2^\star$, $\tilde{\gamma}_1$ and $\tilde{\gamma}_2$ are defined in (18) and (20). Clearly $\gamma_1^\star \ge \tilde{\gamma}_1$ and $\gamma_2^\star \ge \tilde{\gamma}_2$, and therefore, $\gamma^\star \ge \tilde{\gamma}$. Our goal is

to bound $\gamma^\star - \tilde{\gamma}$. For this purpose, we need the following definitions:

$$\gamma_1' \overset{def}{=} P_1 \|\mathbf{H}_1 \mathbf{b}'\|^2, \quad \gamma_2' \overset{def}{=} P_2 \|\mathbf{H}_2 \mathbf{g}'\|^2, \quad \gamma' \overset{def}{=} \frac{\gamma_1' \gamma_2'}{1 + \gamma_1' + \gamma_2'},$$

where $\mathbf{b}'$ is the closest vector in the codebook $\mathbf{C}_1$ to $\mathbf{b}_1$, and $\mathbf{g}'$ is the closest vector in the codebook $\mathbf{C}_2$ to $\mathbf{g}_1$. By considering the definitions of $\tilde{\gamma}_1$ and $\tilde{\gamma}_2$ in (20) and the fact that $\mathbf{b}' \in \mathbf{C}_1$ and $\mathbf{g}' \in \mathbf{C}_2$, it is clear that $\tilde{\gamma}_1 \geq \gamma_1'$ and $\tilde{\gamma}_2 \geq \gamma_2'$, and therefore, $\tilde{\gamma} \geq \gamma'$. Hence, we can write:

$$\gamma^\star - \tilde{\gamma} \leq \gamma^\star - \gamma' = \frac{\gamma_1^\star \gamma_2^\star}{1 + \gamma_1^\star + \gamma_2^\star} - \frac{\gamma_1' \gamma_2'}{1 + \gamma_1' + \gamma_2'}$$
$$\overset{(a)}{\leq} (\gamma_2^\star - \gamma_2') + (\gamma_1^\star - \gamma_1'), \qquad (\text{II.2})$$

where (a) follows from Lemma 1. The terms on the right side of (II.2) can be bounded as follows.

Noting the definitions of $\gamma_1^\star$, γ_1', we have:

$$\gamma_1^\star - \gamma_1' = P_1 \left(\|\mathbf{H}_1 \mathbf{b}_1\|^2 - \|\mathbf{H}_1 \mathbf{b}'\|^2 \right) \overset{(b)}{\leq} 2P_1 \left(\sum_i \phi_i^2 \right) d_{\mathbf{C}_1}(\mathbf{b}_1),$$

where (b) follows from Lemma 4, and ϕ_i's are singular values of $\mathbf{H}_1$. The term $\gamma_2^\star - \gamma_2'$ can be similarly bounded. Combining these bounds with (II.2), we get the following upper bound:

$$\gamma^\star - \tilde{\gamma} \leq 2P_1 \left(\sum_i \phi_i^2 \right) d_{\mathbf{C}_1}(\mathbf{b}_1) + 2P_2 \left(\sum_i \psi_i^2 \right) d_{\mathbf{C}_2}(\mathbf{g}_1),$$
$$(\text{II.3})$$

where ψ_i's are singular values of $\mathbf{H}_2$. Noting that the singular vectors $\mathbf{b}_1$ and $\mathbf{g}_1$ are uniformly distributed on the unit spheres (of the corresponding dimensions) and are independent of the singular values, we can apply Lemma 2 and 5 to (II.3) to achieve the upper bound in (22).

B. Proof of the Upper Bound in (31)

Define:

$$\gamma(\mathbf{s}_1, \mathbf{s}_2) \overset{def}{=} \frac{\gamma_1(\mathbf{s}_1) \gamma_2(\mathbf{s}_2)}{1 + \gamma_1(\mathbf{s}_1) + \gamma_2(\mathbf{s}_2)} + \gamma_0(\mathbf{s}_1), \qquad (\text{II.4})$$

where $\gamma_i(\mathbf{s}) \overset{def}{=} P_i \|\mathbf{H}_i \mathbf{s}\|^2$, for $i = 0, 1, 2$. With this definition, the SNR of the optimal unquantized scheme $\gamma^\star$ and the SNR of the "properly quantized scheme" $\tilde{\gamma}$ can be expressed as:

$$\gamma^\star = \gamma(\mathbf{s}^\star, \mathbf{g}_1) = \frac{\gamma_1^\star \gamma_2^\star}{1 + \gamma_1^\star + \gamma_2^\star} + \gamma_0^\star, \qquad (\text{II.5})$$
$$\tilde{\gamma} = \gamma(\tilde{\mathbf{s}}, \tilde{\mathbf{g}}). \qquad (\text{II.6})$$

In (II.5), $\mathbf{g}_1$ is the dominant right singular vector of $\mathbf{H}_2$, and $\mathbf{s}^\star$ is defined in (25). Also $\gamma_0^\star = \gamma_0(\mathbf{s}^\star)$, $\gamma_1^\star = \gamma_1(\mathbf{s}^\star)$, and $\gamma_2^\star = \gamma_2(\mathbf{g}_1)$. In (II.6), $\tilde{\mathbf{g}} = \arg\max_{\mathbf{w} \in \mathbf{C}_2} \gamma_2(\mathbf{w})$, as defined in (28), and $\tilde{\mathbf{s}}$, defined in (30), can be expressed as:

$$\tilde{\mathbf{s}} = \arg\max_{\mathbf{w} \in \mathbf{C}_1} \chi(\mathbf{w}, \tilde{\mathbf{g}}), \qquad (\text{II.7})$$

where

$$\chi(\mathbf{s}_1, \mathbf{s}_2) \overset{def}{=} \frac{\gamma_1(\mathbf{s}_1) \gamma_2(\mathbf{s}_2)}{1 + \gamma_1(\mathbf{s}_1) + \gamma_2(\mathbf{s}_2)} + P_0 \sum_i \nu_i^2 \left| \tilde{\mathbf{e}}_i^H \mathbf{s}_1 \right|^2. \quad (\text{II.8})$$

Here ν_i's are the singular values of $\mathbf{H}_0$ and $\tilde{\mathbf{e}}_i = \arg\min_{\mathbf{w} \in \mathbf{C}_0} d(\mathbf{w}, \mathbf{e}_i)$, where $\mathbf{e}_i$'s are the right singular values of $\mathbf{H}_0$.

Our goal is to bound the SNR loss $\gamma^\star - \tilde{\gamma}$. For this purpose, we need to define:

$$\mathbf{s}' = \arg\max_{\mathbf{w} \in \mathbf{C}_1} \gamma(\mathbf{w}, \tilde{\mathbf{g}}) , \quad \gamma' = \gamma(\mathbf{s}', \tilde{\mathbf{g}}). \qquad (\text{II.9})$$

We can write:

$$\gamma^\star - \tilde{\gamma} = (\gamma^\star - \gamma') + (\gamma' - \tilde{\gamma}). \qquad (\text{II.10})$$

Based on the definitions given above, each of these terms can be bounded as follows.

Step 1) Bounding $\gamma^\star - \gamma'$:
Define:

$$\gamma'' \overset{def}{=} \frac{\gamma_1'' \gamma_2''}{1 + \gamma_1'' + \gamma_2''} + \gamma_0'',$$
$$\gamma_0'' \overset{def}{=} \gamma_0(\mathbf{s}''), \quad \gamma_1'' \overset{def}{=} \gamma_1(\mathbf{s}''), \quad \gamma_2'' \overset{def}{=} \gamma_2(\mathbf{g}''), \quad (\text{II.11})$$

where $\mathbf{s}'' \in \mathbf{C}_1$ is the closest vector in the codebook $\mathbf{C}_1$ to $\mathbf{s}^\star$, and $\mathbf{g}'' \in \mathbf{C}_2$ is the closest vector in the codebook $\mathbf{C}_2$ to $\mathbf{g}_1$. Noting the definitions of $\tilde{\mathbf{g}}$ and $\mathbf{s}'$, it is clear that $\gamma' \geq \gamma''$ and we can write:

$$\gamma^\star - \gamma' \leq \gamma^\star - \gamma'' \leq \left| \frac{\gamma_1^\star \gamma_2^\star}{1 + \gamma_1^\star + \gamma_2^\star} - \frac{\gamma_1'' \gamma_2''}{1 + \gamma_1'' + \gamma_2''} \right| + |\gamma_0^\star - \gamma_0''|$$
$$\overset{(a)}{\leq} |\gamma_2^\star - \gamma_2''| + |\gamma_1^\star - \gamma_1''| + |\gamma_0^\star - \gamma_0''|$$
$$\overset{(b)}{=} P_2 \left| \|\mathbf{H}_2 \mathbf{g}_1\|^2 - \|\mathbf{H}_2 \mathbf{g}''\|^2 \right| + P_1 \left| \|\mathbf{H}_1 \mathbf{s}^\star\|^2 - \|\mathbf{H}_1 \mathbf{s}''\|^2 \right|$$
$$+ P_0 \left| \|\mathbf{H}_0 \mathbf{s}^\star\|^2 - \|\mathbf{H}_0 \mathbf{s}''\|^2 \right|$$
$$\overset{(c)}{\leq} 2P_2 \left(\sum_i \psi_i^2 \right) d_{\mathbf{C}_2}(\mathbf{g}_1) + 2P_1 \left(\sum_i \phi_i^2 \right) d_{\mathbf{C}_1}(\mathbf{s}^\star)$$
$$+ 2P_0 \left(\sum_i \nu_i^2 \right) d_{\mathbf{C}_1}(\mathbf{s}^\star),$$
$$(\text{II.12})$$

where we have used Lemma 1 for (a). In (b), $\{\gamma_i^\star\}_{i=0}^2$ and $\{\gamma_i''\}_{i=0}^2$ have been replaced by their definitions. Finally, (c) results from Lemma 4.

Step 2) Bounding $\gamma' - \tilde{\gamma}$:
For this step, we will need the following lemma.

Lemma 6: For any unit vector $\mathbf{s}$, we have:

$$|\gamma(\mathbf{s}, \tilde{\mathbf{g}}) - \chi(\mathbf{s}, \tilde{\mathbf{g}})| \leq 2P_0 \sum_i \nu_i^2 d_{\mathbf{C}_0}(\mathbf{e}_i).$$

Proof: Noting the definition of $\gamma(\cdot, \cdot)$ in (II.4),

$$\gamma(\mathbf{s}, \tilde{\mathbf{g}}) = \frac{\gamma_1(\mathbf{s}) \gamma_2(\tilde{\mathbf{g}})}{1 + \gamma_1(\mathbf{s}) + \gamma_2(\tilde{\mathbf{g}})} + P_0 \sum_i \nu_i^2 \left| \mathbf{e}_i^H \mathbf{s} \right|^2.$$

Therefore,

$$|\gamma(\mathbf{s}, \tilde{\mathbf{g}}) - \chi(\mathbf{s}, \tilde{\mathbf{g}})| = P_0 \left| \sum_i \nu_i^2 \left(\left| \mathbf{e}_i^H \mathbf{s} \right|^2 - \left| \tilde{\mathbf{e}}_i^H \mathbf{s} \right|^2 \right) \right|$$
$$\overset{(a)}{\leq} 2P_0 \sum_i \nu_i^2 d(\mathbf{e}_i, \tilde{\mathbf{e}}_i),$$

where in (a), we have used Lemma 3. Noting the definition of $\tilde{\mathbf{e}}_i$'s, we have $d(\mathbf{e}_i, \tilde{\mathbf{e}}_i) = d_{\mathbf{C}_0}(\mathbf{e}_i)$ and the proof is complete. ∎

Let $\theta = 2P_0 \sum_i \nu_i^2 d_{\mathbf{C}_0}(\mathbf{e}_i)$, then we have:

$$\gamma' = \gamma(\mathbf{s}', \tilde{\mathbf{g}}) \overset{(a)}{\leq} \chi(\mathbf{s}', \tilde{\mathbf{g}}) + \theta \overset{(b)}{\leq} \chi(\tilde{\mathbf{s}}, \tilde{\mathbf{g}}) + \theta \overset{(c)}{\leq} \gamma(\tilde{\mathbf{s}}, \tilde{\mathbf{g}}) + 2\theta$$
$$= \tilde{\gamma} + 2\theta, \text{ (II.13)}$$

where in (a) and (c) we have used Lemma 6, and (b) results from (II.7) and the fact that $\mathbf{s}' \in \mathbf{C}_1$. By combining (II.10), (II.12) and (II.13) we get the following upper bound:

$$\gamma^\star - \tilde{\gamma} \leq 4P_0 \sum_i \nu_i^2 d_{\mathbf{C}_0}(\mathbf{e}_i) + 2P_2 \left(\sum_i \psi_i^2 \right) d_{\mathbf{C}_2}(\mathbf{g}_1)$$
$$+ 2 \left(P_1 \left(\sum_i \phi_i^2 \right) + P_0 \left(\sum_i \nu_i^2 \right) \right) d_{\mathbf{C}_1}(\mathbf{s}^\star). \text{ (II.14)}$$

From Appendix I, $\mathbf{s}^\star$ is uniformly distributed on the unite sphere and is independent of the singular values ϕ_i's and ν_i's. The same argument holds for the singular vectors $\mathbf{g}_1$ and $\mathbf{e}_i$'s and the corresponding singular values ψ_i's and ν_i's. By considering these facts and taking the expectation of both sides of (II.14) and using Lemma 2 and 5, we get the upper bound in (31).

REFERENCES

[1] T. K. Y. Lo, "Maximum ratio transmission," *IEEE Trans. Commun.*, vol. 47, no. 10, pp. 1458-1461, Oct. 1999.

[2] S. M. Alamouti, "A simple transmit diversity technique for wireless communications," *IEEE J. Select. Areas Commun.*, vol. 16, no. 8, pp. 1451-1458, Oct. 1998.

[3] D. J. Love, R. W. Heath Jr., W. Santipach, and M. L. Honig, "What is the value of limited feedback for MIMO channels?," *IEEE Commun. Mag.*, vol. 42, no. 10, pp. 54-59, Oct. 2004.

[4] J. H. Conway, R. H. Hardin, and N. J. A. Sloane, "Packing lines, planes, etc.: Packings in Grassmannian spaces," *Exper. Math.*, vol. 5, no. 2, pp. 139-159, 1996.

[5] D. J. Love and R. W. Heath Jr., "Grassmannian beamforming for multiple-input multiple-output wireless systems," *IEEE Trans. Inform. Theory*, vol. 49, no. 10, pp. 2735-2747, Oct. 2003.

[6] K. K. Mukkavilli, A. Sabharwal, E. Erkip, and B. Aazhang, "On beamforming with finite rate feedback in multiple-antenna systems," *IEEE Trans. Inform. Theory*, vol. 49, no. 10, pp. 2562-2579, Oct. 2003.

[7] B. Wang, J. Zhang, and A. Host-Madsen, "On the capacity of MIMO relay channels," *IEEE Trans. Inform. Theory*, vol. 51, no. 1, pp. 29-43, Jan. 2005.

[8] C. K. Lo, S. Vishwanath, and R. W. Heath Jr., "Rate bounds for MIMO relay channels using precoding," in *Proc. IEEE GLOBECOM*, vol. 3, pp. 1172-1176, St. Louis, MO, Nov. 2005.

[9] X. Tang and Y. Hua, "Optimal design of non-regenerative MIMO wireless relays," *IEEE Trans. Wireless Commun.*, vol. 6, no. 4, pp. 1398-1407, Apr. 2007.

[10] O. Munoz, J. Vidal, and A. Agustin, "Non-regenerative MIMO relaying with channel state information," in *Proc. ICASSP05*, vol. 3, Mar. 2005.

[11] I. Hammerström and A. Wittneben, "Power allocation schemes for amplify-and-forward MIMO-OFDM relay links," *IEEE Trans. Wireless Commun.*, vol. 6, no. 8, pp. 2798-2802, Aug. 2007.

[12] Z. Fang, Y. Hua, and J. C. Koshy "Jonit source and relay optimization for a non-regenerative MIMO relay," *IEEE Workshop on Sensor Arrays and Multi-channel Processing*, Waltham, MA, U.S.A., Jul. 2006.

[13] N. Varanese, O. Simeone, Y. Bar-Ness, and U. Spagnolini, "Achievable rates of multi-hop and cooperative MIMO amplify-and-forward relay systems with full CSI," in *Proc. Signal Processing Advances in Wireless Communications SPAWC '06*, Jul. 2006.

[14] N. Ahmed, M. Khojastepour, A. Sabharwal, and B. Aazhang, "Outage minimization with limited feedback for the fading relay channel," *IEEE Trans. Commun.*, vol. 54, no. 4, pp. 659-669, Apr. 2006.

[15] B. Khoshnevis, W. Yu, and R. Adve, "Grassmannian beamforming for MIMO amplify-and-forward relaying", *Conference on Information Sciences and Systems (CISS)*, Princeton, NJ, U.S.A., Mar. 2008.

[16] T. L. Marzetta and B. M. Hochwald, "Capacity of a mobile multiple-antenna communication link in Rayleigh flat fading," *IEEE Trans. Inform. Theory*, vol. 45, no. 1, pp. 139-157, Jan. 1999.

[17] D. J. Love. Grassmannian Subspace Packing. [Online] Available: http://cobweb.ecn.purdue.edu/~djlove/grass.htm
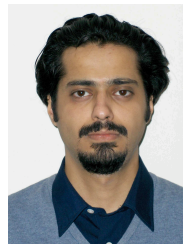
Behrouz Khoshnevis received B.Sc. and M.Sc. degrees from Sharif University of Technology, Tehran, Iran, in 2003 and 2005 respectively, both in Electrical Engineering. From 2003 to 2006, he was with the Advanced Communications Research Institute, Sharif University of Technology, Tehran, Iran. He is currently pursuing the Ph.D. degree in Electrical and Computer Engineering at the University of Toronto, Toronto, Ontario, Canada. Behrouz Khoshnevis is a member of IEEE Communications Society.
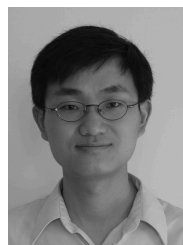
Wei Yu (S'97-M'02) received the B.A.Sc. degree in Computer Engineering and Mathematics from the University of Waterloo, Waterloo, Ontario, Canada in 1997 and M.S. and Ph.D. degrees in Electrical Engineering from Stanford University, Stanford, CA, in 1998 and 2002 respectively. Since 2002, he has been an Assistant Professor with the Electrical and Computer Engineering Department at the University of Toronto, Toronto, Ontario, Canada, where he holds a Canada Research Chair. His main research interests include multiuser information theory, optimization, wireless communications and broadband access networks.

Prof. Wei Yu was an Editor for *IEEE Transactions on Wireless Communications* from 2004 to 2007. He was a Guest Editor of *IEEE Journal on Selected Areas in Communications* for a special issue on "Nonlinear Optimization of Communications Systems" in 2006, and a Guest Editor of *EURASIP Journal on Applied Signal Processing* for a special issue on "Advanced Signal Processing for Digital Subscriber Lines" in 2005. He received Early Researcher Award from Ontario in 2006, and the Early Career Teaching Award from the Faculty of Applied Science and Engineering, University of Toronto, in 2007.

Raviraj Adve (S'88, M'97, SM'03) received his B. Tech. degree in Electrical Engineering from IIT, Bombay in 1990 and his Ph.D. from Syracuse University in 1996, where his dissertation received a Doctoral Prize. Between 1997 and August 2000, he worked for Research Associates for Defense Conversion Inc. on contract with the Air Force Research Laboratory at Rome, NY. He joined the faculty at the University of Toronto in August 2000. Dr. Adve's research interests include developing practical signal processing algorithms for wireless communications and radar systems. He is currently focused on linear precoding in the multiuser downlink, cooperative communications and target detection using distributed sensing networks.