Interference Mitigation via Relaying
S. Arvin Ayoughi, Wei Yu

Abstract
This paper studies the effectiveness of relaying for interference mitigation in an interference-limited communication scenario. We are motivated by the observation that in a cellular network, a relay node placed at the cell-edge observes intercell interference that is correlated with intended receiver’s interference. A relaying link can effectively allow their antennas to be pooled together for both signal enhancement and interference mitigation. We model this scenario by a multiple-input multiple-output (MIMO) Gaussian relay channel with a digital relay-destination link of finite capacity, with correlated noises across relay and destination antennas. Assuming compress-and-forward strategy with Gaussian input distribution and quantization noise, we characterize the achievable rate using a coordinate ascent algorithm for jointly optimizing transmit and quantization covariance matrices. For fixed input distribution, the globally optimum quantization noise covariance matrix can be found in closed-form using a transformation for the relay’s observation that simultaneously diagonalizes two conditional covariance matrices by *congruence. For fixed quantization, the globally optimum transmit covariance matrix can be found via convex optimization. This paper further shows that such an optimized achievable rate is to within a constant additive gap of capacity of the channel, and that the optimal structure of the quantization noise covariance enables a characterization of the slope of the achievable rate as a function of the relay-destination link capacity. Finally, this paper studies the improvement in spatial degrees of freedom (DoF) by MIMO relaying in presence of noise correlation and reveals a relation between the DoF gain and the aforementioned slope via a connection to the deterministic relay channel capacity.

Index Terms
Gaussian MIMO relay channel, noise correlation, compress-and-forward, reverse water-filling, approximate capacity.

I. INTRODUCTION

Interference is a key limiting factor in many communication scenarios. A wireless cellular network with densely deployed basestations (BS’s), for example, is typically interference limited. The provision of high data rate at the cell edge, where the signal is the weakest and the interference is the strongest, is a major challenge in cellular network physical-layer design.

This paper explores the use of relays for interference mitigation. The idea is that by placing a multiple-antenna relay at the cell edge (see Fig. 1), a user device in close proximity of a relay would be able to establish an out-of-band relay link and benefit from relay’s observation in decoding the downlink signal. In essence, the user and the relay would be able to pool their antennas together—the extra spatial dimensions allow for not only signal enhancement but more importantly also interference mitigation, because the interferences at the user side and the relay are highly correlated due to their physical proximity.

The benefit of antenna pooling depends crucially on the quality of the relay link. The goal of this paper is to quantify the benefit of relaying for a cellular downlink as a function of the relay link capacity. Toward this end, this paper studies a multiple-input multiple-output (MIMO) Gaussian relay channel with digital relay-destination link, where, due to common intercell interference, the noise processes across the relay and destination antennas are correlated. This is a simple yet fundamental model of cooperative communications for which information theoretical analysis can yield significant insight. In particular, this paper adopts a compress-and-forward relaying strategy, which is appropriate given the physical proximity of the relay and the destination in the scenario of interest [1], [2]. In this case, the relay provides the destination with a compressed version of its observations, the accuracy of which is determined by the available capacity of relay-to-destination link. The relay uses Wyner-Ziv coding to exploit the side information available at the destination in quantizing its observations. The goal of this paper is to analyze the optimal transmission and quantization strategies for the MIMO relay channel with digital relay link in the presence of correlated interference.

A. Main Results

This paper makes the following contributions toward the goal of understanding how to best take advantage of the MIMO relay for both signal enhancement and interference mitigation:

- We propose an iterative algorithm for optimizing the covariance matrices of Gaussian input signal and Gaussian quantization noise for the MIMO relay channel. We show that the optimization of input covariance matrix for fixed quantization noise distribution is a concave optimization, and the optimization of quantization noise covariance matrix for fixed input distribution can be solved via a simultaneous diagonalization transformation.

This work was supported by Huawei Technologies Canada Co., Ltd., and by National Science and Engineering Research Council of Canada. The materials in this paper have been presented in part at IEEE International Conference on Communication (ICC), London, UK, June 2015, and in part at the 14th Canadian Workshop on Information Theory (CWIT), St. John’s, NL, Canada, July 2015. 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 M5S 3G4, Canada (e-mails: sa.ayoughi@mail.utoronto.ca, weiyu@comm.utoronto.ca)
The allocation of relaying bits across the spatial dimensions should follow a reverse water-filling solution, where more bits are used to quantize spatial dimensions with higher conditional signal-to-interference-and-noise ratio (CSINR).

We characterize the slope of compress-and-forward rate curve versus relaying link capacity for optimized quantization noise covariance at a fixed input distribution. The slope is related to the generalized eigenvalues of certain conditional covariance matrices, and asymptotically approaches the maximum value of 1, if the MIMO relay channel contains a so-called asymptotically deterministic component.

We characterize the improvement in spatial degrees of freedom (DoF) by compress-and-forward relaying as a function of the numbers of antennas in the network. A distributed interference zero-forcing scheme at the relay is shown to be DoF optimal.

The optimized compress-and-forward strategy is shown to achieve the capacity of the MIMO relay channel with noise correlation to within a constant additive gap.

B. Related Works

The relay channel is a classic model that has been widely studied in the literature. Specifically related to this work, relaying in the presence of noise correlation for single-input single-output (SISO) Gaussian channel is studied in [3], where the negative correlation between relay and destination noises is shown to improve the achievable rate of compress-and-forward relaying scheme. For the Gaussian MIMO relay channel with independent noises, a coordinate ascent procedure for maximizing the compress-and-forward achievable rate over input and quantization noise covariance matrices is proposed in [4]. For optimizing quantization at the relay under fixed input distribution, [4] uses the conditional Karhunen-Loève Transform (CKLT) of the relay’s observed vector given destination’s observation, followed by a reverse water-filling allocation of the relay link rates.

This paper focuses on the relay channel with correlated noises for which the solution is more complex. We show that the right transformation is the simultaneous diagonalization by *congruence of two conditional covariance matrices. Although our solution to the optimization problem can eventually be shown to be the same as that of [8], the direct diagonalization approach in this paper is simpler and provides insight into the optimized MIMO relaying strategy for interference mitigation and signal enhancement.

For the SISO Gaussian relay channel with noise correlation, [9] shows that compress-and-forward achieves the capacity to within a constant additive gap. For the MIMO Gaussian relay channel with independent noise vectors, capacity approximation using the partial decode-and-forward scheme is provided in [10]. In this paper, for the MIMO Gaussian relay channel with noise correlation, we use the simultaneous diagonalization transform to show that compress-and-forward achieves the capacity to within a constant additive gap that is tighter than the gap of extension of [9] to the MIMO case.

This paper studies the improvement in spatial DoF due to the compress-and-forward relaying. The DoF-optimal transformation at the relay prior to quantization involves distributed zero-forcing of interference. Distributed zero-forcing of interference has been considered for various classes of SISO relay networks with analog relaying links, e.g., [11]–[14]. This paper deals with a MIMO relay channel with digital relay-destination link in which distributed zero-forcing of interference reveals the asymptotically deterministic components of the MIMO relay channel. The determinism here refers to the condition that the observation of the relay is a deterministic function of the input channel and the observation of the destination. As shown in [15], compress-and-forward achieves the cut-set bound in this case. This paper illustrates that this type of determinism occurs in the MIMO relay channel in the asymptotically high SNR ans INR regime. This is the fundamental reason that compress-and-forward relaying can be effective in improving the overall throughput in a MIMO relay channel [16].

C. Notation

In this paper we denote matrices by uppercase letters, e.g., $H$, vectors of random variables by uppercase bold letters, e.g., $X$, and its realization by lowercase bold letters, e.g., $x$. Also, $0_{r \times 1}$ stands for the $r$-dimensional zero vector and $I_r$ is the $r \times r$ identity matrix. Conjugate transpose, Moore-Penrose pseudoinverse, trace, determinant, rank, the $i$th largest eigenvalue, and the row span of a matrix are denoted by $(\cdot)^\dagger$, $(\cdot)^{-1}$, $\text{tr}(\cdot)$, $|\cdot|$, $\text{rank}(\cdot)$, $\lambda_i(\cdot)$, and $\text{rowspan}(\cdot)$, respectively. The expectation operator is shown by $\mathbb{E}(\cdot)$. The set of complex numbers is denoted by $\mathbb{C}$. Finally, we define $(x)^+ = \max(0, x)$.

1Here, “*” refers to conjugate transpose of the transformation matrix. Although, in this paper we use $(\cdot)^\dagger$ to denote conjugate transpose, we preferred to not to change the terminology of [6].
Fig. 1. A cellular network with a BS and two cell-edge devices that pool antennas through a finite-capacity digital link. The relay and the destination devices experience correlated noise due to common interference signals.

D. Organization of the Paper

Section II introduces the system model, relaying scheme, and the problem formulations. The proposed iterative algorithm for optimizing the compress-and-forward scheme is presented in Section III. Section IV provides interpretations for the optimal structure of quantization noise and characterizes the marginal improvement in the overall achievable rate with respect to relay-destination link capacity. Section V provides a DoF analysis of the MIMO relay channel. Section VI interprets the analysis via a connection to deterministic relay channel. Section VII shows that compress-and-forward achieves the capacity of the MIMO Gaussian relay channel with noise correlation to within a constant additive gap. Finally, simulation results are presented in Section VIII, and Section IX concludes the paper.

II. MIMO GAUSSIAN RELAY CHANNEL WITH NOISE CORRELATION

A. System Model

Consider the cellular downlink transmission from a BS to a cell-edge user located in close proximity of a relay, as shown in Fig. 1. Due to common interference from nearby BS’s, the noises across the relay and the destination antennas are highly correlated. An out-of-band relay link is established between the relay and the user for interference mitigation and signal enhancement.

Mathematically, this communication scenario is modeled as a Gaussian MIMO relay channel with correlated interference and with a digital relay-destination link of capacity $C_0$ bits. The source, relay, and destination are equipped with $s$, $r$, and $d$ antennas, respectively. Let $t$ be the total number of antennas from all the interfering BS’s. The received signals at the relay and destination can be written as

$$Y_r = H_{sr}X + N_r,$$

$$Y_d = H_{sd}X + N_d,$$

where

$$N_r = H_{tr}X_t + N_1,$$

$$N_d = H_{td}X_t + N_2$$

are the correlated noises. Here, $H_{sr} \in \mathbb{C}^{r \times s}$ and $H_{sd} \in \mathbb{C}^{d \times s}$ are the source-relay and source-destination channel matrices respectively; $H_{tr} \in \mathbb{C}^{r \times t}$ and $H_{td} \in \mathbb{C}^{d \times t}$ are the interferers-to-relay and interferers-to-destination channel matrices respectively; $N_1 \sim \mathcal{CN}(0_{r \times 1}, \sigma^2_1I_r)$ and $N_2 \sim \mathcal{CN}(0_{d \times 1}, \sigma^2_2I_d)$ are additive and independent background noises at the relay and the destination respectively; $X \in \mathbb{C}^{s \times 1}$ is the transmit signal vector from the source under power constraint $E(X^\dagger X) \leq P$; finally $X_t \in \mathbb{C}^{t \times 1}$ is the interference signal vector that is assumed to be Gaussian with $X_t \sim \mathcal{CN}(0_{t \times 1}, S_{X_t})$, independent of everything else, and is treated as a part of noise. In this paper, we assume that the entries of the channel matrices are drawn independently from a continuous probability distribution, so that each of them as well as their concatenations are full rank almost surely, in order to carry out DoF analysis.

B. Capacity Upper and Lower Bounds

We state several well-known results on the capacity of the relay channel.
Theorem 1 ([1, Theorem 4]) The capacity of the relay channel is upper bounded by

$$C \leq \max_{p(x), \mathbb{E}[X \mid X] \leq P} \min \{ I(X; Y_r, Y_d), I(X; Y_d) + C_0 \}.$$  

(5)

The evaluation of the cut-set upper bound is a convex optimization problem. Here, the optimal $p(x)$ is a multivariate Gaussian, i.e., $X \sim \mathcal{CN}(0_{n \times 1}, S_X)$ for some positive semidefinite $S_X$. The two mutual information terms in cut-set bound are both concave functions of $S_X$.

Theorem 2 ([1, Theorem 6]) The capacity of the relay channel is lower bounded by

$$C \geq \max_{p(x)p(y_r | y_r), \mathbb{E}[X \mid X] \leq P} \min \{ I(X; \hat{Y}_r, Y_d), I(X; Y_d) + C_0 - I(Y_r; \hat{Y}_r | Y_d, X) \}. \tag{6}$$

The above lower bound is based on the compress-and-forward relaying scheme, but the evaluation of this achievable rate is not always straightforward. For the Gaussian relay channel, although it can be shown that Gaussian quantization at the relay is optimal for Gaussian signaling at the source [8] and vice versa, the jointly optimal input distribution and quantization test channel of the compress-and-forward lower bound is not yet known. See [17] for an example where jointly Gaussian distribution is suboptimal. For tractability, this paper restricts attention to jointly Gaussian transmission $X \sim \mathcal{CN}(0_{n \times 1}, S_X)$ and Gaussian quantization modeled as

$$\hat{Y}_r = Y_r + Q,$$ \tag{7}

where the quantization noise $Q \sim \mathcal{CN}(0_{n \times 1}, S_Q)$ is independent of all variables. Even then, the optimization of the achievable rate over $(S_X, S_Q)$ is still not straightforward. This optimization is a main subject of this paper.

In the above compress-and-forward scheme, the destination first decodes the quantized version of the relay’s observation uniquely, then decodes the source message. Reliable unique decoding of the quantization codeword requires the compression rate at the relay not to exceed the relay link capacity. In an alternative decoding scheme, the destination decodes the message by non-unique decoding of the quantization codeword at the relay, thus allowing the compression rate at the relay to potentially exceed the relay link capacity. Although it can be shown that such flexibility does not result in higher achievable rate, the resulting rate expression has the advantage that it resembles the cut-set bound, which is a useful feature that enables a constant gap characterization of the relay channel capacity. We state this alternative rate expression below.

Theorem 3 ([18, Theorem 16.4]) The capacity of the relay channel is lower bounded by

$$C \geq \max_{p(x)p(y_r | y_r), \mathbb{E}[X \mid X] \leq P} \min \left\{ I(X; \hat{Y}_r, Y_d), I(X; Y_d) + C_0 - I(Y_r; \hat{Y}_r | Y_d, X) \right\}. \tag{8}$$

Furthermore, optimization problems (6) and (8) have the same global maximum, and their optimal distributions $p(x)p(y_r | y_r)$ are equal.

C. Problem Formulation

This paper addresses four aspects of MIMO compress-and-forward relaying in the presence of noise correlation.

1) Optimization of Input and Quantization Covariance Matrices: This paper proposes a method to numerically optimize the achievable rate as in (6). Assuming joint Gaussian input distribution and quantization test channel, the achievable rate (6) can be expressed as

$$R_{CF}(C_0) = \max_{S_X, S_Q} \frac{f_o(S_X, S_Q)}{f_c(S_X, S_Q) \leq C_0 \ \text{tr}(S_X) \leq P \ S_X \geq 0, \ S_Q \geq 0}, \tag{9}$$

where the objective function is

$$f_o(S_X, S_Q) = I(X; \hat{Y}_r, Y_d) = \log \frac{HS_XH^\dagger + S_{\text{int}} + \sigma^2 I_{(r+d)} + S_Q \begin{bmatrix} 0_{r \times r} & 0_{r \times d} \\ 0_{d \times r} & 0_{d \times d} \end{bmatrix}}{S_{\text{int}} + \sigma^2 I_{(r+d)} + \begin{bmatrix} S_Q & 0_{r \times d} \\ 0_{d \times r} & 0_{d \times d} \end{bmatrix}}. \tag{10}$$
and the constraint is

$$f_c(S_X, S_Q) = I(Y_r, \hat{Y}_r | Y_d) = \log \frac{|HS_XH^\dagger + S_{\text{int}} + \sigma^2I_{(r+d)} + \begin{bmatrix} S_Q & 0_{r \times d} \\ 0_{d \times r} & 0_{d \times d} \end{bmatrix}|}{|H_{sd}S_XH_{sd}^\dagger + S_{\text{int}}^{(2,2)} + \sigma^2I_d|},$$

(11)

where $H = [H_{sr} \ H_{sd}]^\dagger$, and

$$S_{\text{int}} = \begin{bmatrix} S_{\text{int}}^{(1,1)} \\ S_{\text{int}}^{(1,2)} \\ S_{\text{int}}^{(2,1)} \\ S_{\text{int}}^{(2,2)} \end{bmatrix} = \begin{bmatrix} H_{tr}S_XH_{tr}^\dagger & H_{tr}S_XH_{td}^\dagger \\ H_{td}S_XH_{tr}^\dagger & H_{td}S_XH_{td}^\dagger \end{bmatrix}$$

(12)

is the interference covariance matrix.

2) Characterization of the Slope of Achievable Rate with Respect to $C_0$: This paper further evaluates the effectiveness of compress-and-forward in improving the overall rate in a relay channel as measured by the slope

$$\frac{dR_{CF}(C_0)}{dC_0}.$$ (13)

Here, $R_{CF}(C_0)$ is evaluated at a fixed $S_X$. By the upper bound the maximum value of this slope is 1. This paper provides conditions under which the slope is asymptotically close to the maximum value of 1.

3) DoF Improvement by Compress-and-Forward: This paper also evaluates the effectiveness of compress-and-forward in the large $C_0$ and high SNR regime by studying the improvement in spatial DoF. In particular, define

$$\rho \triangleq \frac{1}{\sigma^2},$$ (14)

and let the relay-destination capacity scale with $\rho$ as

$$C_0(\rho) = \alpha \log(\rho) + o(\log(\rho)).$$ (15)

The DoF improvement due to relaying is defined as

$$\Delta \text{DoF} \triangleq \lim_{\rho \to \infty} \frac{R_{CF}(C_0(\rho)) - R_{CF}(0)}{\log(\rho)}.$$ (16)

Interestingly, this paper shows that the conditions, on the numbers of antennas in the system $(s, d, r, t)$, under which relaying brings DoF improvement is identical to the conditions for asymptotically having the maximum slope of the achievable rate with respect to $C_0$.

4) Capacity of the Relay Channel to Within Constant Gap: Finally, this paper shows the constant-gap optimality of the compress-and-forward strategy for the MIMO relay channel with noise correlation. We bound the gap between the achievable rate and the cut-set bound by a constant that only depends on the numbers of antennas. In the absence of an exact capacity characterization, the constant gap characterization provides an optimality guarantee, which is relevant in the high SNR regime.

III. OPTIMIZATION OF COMPRESS-AND-FORWARD

The joint optimization of transmission at the source and quantization at the relay is crucial for characterizing the relay channel performance under noise correlation. This section provides an optimization method and illustrates the structure of optimal $S_X$ and $S_Q$ at the stationary point of the overall optimization problem.

A. Iterative Optimization of Lagrangian

The joint optimization of the input and quantization covariance matrices is not a convex optimization problem, as both the objective function and the constraint of are concave in $S_X$ and convex in $S_Q$. This paper takes an approach of forming the Lagrangian

$$\mathcal{L}(S_X, S_Q, \mu) = f_o(S_X, S_Q) - \mu(f_c(S_X, S_Q) - C_0),$$ (17)

then finding a stationary point of the Lagrangian for a fixed Lagrange multiplier $\mu$ by solving

$$\max_{S_X \succeq 0, \ S_Q \succeq 0, \ tr(S_X) \leq P} \mathcal{L}(S_X, S_Q, \mu),$$ (18)

and finally searching for the optimal $\mu$ that results in

$$f_c(S_X^*, S_Q^*) = C_0.$$ (19)
in an outer loop. It can be easily shown that the optimal \( \mu^* \in (0, 1) \), because if \( \mu \geq 1 \), i.e., if the relay link capacity constraint penalizes the objective at more than a 1:1 ratio, then the optimal \( S_Q^* \) would be infinite, resulting in \( f_c(S_X, S_Q^*) = 0 \). Thus, the outer loop for finding \( \mu^* \) is a one-dimensional root-finding problem that can be solved by bisection (assuming that \( f_c(S_X^*(\mu), S_Q^*(\mu)) \) is decreasing in \( \mu \)).

We propose an iterative coordinate ascent approach for solving (18). For a fixed \( \mu \in (0, 1) \), each of the individual optimizations of \( S_X \) and \( S_Q \) can be solved to global optimality (and the global optimum is essentially unique). The iterative optimization process generates a nondecreasing sequence of the Lagrangian objective, hence it converges.

The details of optimizing \( S_Q \) for a fixed \( S_X \) and optimizing \( S_X \) for a fixed \( S_Q \) are provided in the subsequent sections. The overall iterative approach is summarized as Algorithm 1. The following theorem states the convergence result formally.

**Theorem 4** Assuming that the optimal \( S_X \) for a fixed \( S_Q \) is unique and the optimal \( S_Q \) for a fixed \( S_X \) is unique, the inner iterative optimization procedure in Algorithm 1 converges to a stationary point of the Lagrangian maximization problem (18). Further, the optimal \( \mu \) is one that satisfies (19). Such a \( \mu \) leads to a KKT point of the joint transmit and quantization noise covariance optimization problem (9).

**Proof:** For a fixed \( \mu \), coordinate ascent on the Lagrangian is monotonically increasing, hence convergent. The uniqueness in the optimization of \( S_X \) for a fixed \( S_Q \) and in the optimization of \( S_Q \) for a fixed \( S_X \) ensures that coordinate ascent converges to a stationary point. This together with a \( \mu \) that satisfies (19) gives a KKT point of (9).

There are special cases, for example when \( s = 1 \) or \( r = 1 \), where the above procedure would produce a globally optimal solution. However, in general only convergence to a stationary point is assured.

**B. Optimization of \( S_X \) for a Fixed \( S_Q \)**

Although the optimization problem (9) is not concave in \( S_X \) for fixed \( S_Q \), we observe that the maximization of its Lagrangian (17) at a given \( S_Q^* \) is concave.

**Theorem 5** The maximization of Lagrangian of (9) over \( S_X \) for fixed \( S_Q \) in (20), i.e.,

\[
\max_{S_X \geq 0, \text{tr}(S_X) \leq \rho} \mathcal{L}(S_X, S_Q, \mu),
\]

is a concave optimization problem.

**Proof:** For fixed \( S_Q \) the Lagrangian can be written as a function of \( S_X \) as

\[
\mathcal{L} = I(X; \tilde{Y}_r, Y_d) - \mu I(Y_r; Y_d | Y_d) + \mu C_0
\]

\[
= (1-\mu) I(X; \tilde{Y}_r, Y_d) + \mu I(X; Y_d) + \text{const.}
\]

\[
= (1-\mu) \log |HS_XH^\dagger + S_{\text{int}} + \sigma^2 I_{(r+d)} + \begin{bmatrix} S_Q & 0_{r \times d} \\ 0_{d \times r} & 0_{d \times d} \end{bmatrix}| + \mu \log |H_{sd}S_XH_{sd}^\dagger + S_{\text{int}}^{(2,2)} + \sigma^2 I_d| + \text{const.}
\]

which is a concave logdet function for \( \mu \in (0, 1) \).

The above functional form of the Lagrangian provides intuition about the optimal choice of \( S_X \). The Lagrangian is a convex combination of two terms. The first term corresponds to the channel from \( X \) to the combined relay and destination receiver \((\tilde{Y}_r, Y_d)\), while the second term corresponds to the channel from \( X \) to the destination \( Y_d \) alone. For larger values of \( C_0 \) (or equivalently small values of \( \mu \)), the optimal \( S_X \) should be close to the water-filling covariance matrix against the combined vector channel \( H \). For small values of \( C_0 \), the optimal \( S_X \) should be close to the water-filling covariance matrix against the source-destination channel \( H_{sd} \) alone. In the general case, maximizing the convex combination captures the tradeoff between the two solutions.
C. Optimization of $S_Q$ for a Fixed $S_X$

We now provide a closed-form solution for the optimal $S_Q$ that maximizes the Lagrangian \((18)\) for a given $\hat{S}_X$, i.e., solving

$$\max_{S_Q \succeq 0} L(\hat{S}_X, S_Q, \mu).$$

(21)

For the optimization of $S_Q$ when $\hat{S}_X$ is kept fixed, the objective and constraint functions \((10)-(11)\) can be rewritten as

$$f_o = \log |S_{Y,|Y_d} + S_Q| - \log |S_{Y,|Y_d, X} + S_Q| + \text{const.},$$

(22)

$$f_c = \log |S_{Y,|Y_d} + S_Q| - \log |S_Q| + \text{const.},$$

(23)

and Lagrangian in \((21)\) as

$$L(\hat{S}_X, S_Q, \mu) = (1 - \mu) \log |S_{Y,|Y_d} + S_Q| + \mu \log |S_Q| - \log |S_{Y,|Y_d, X} + S_Q| + \text{const.},$$

(24)

where the conditional covariances are obtained using the generalized Schur complement formula \((19)\)

$$S_{Y,|Y_d} = H_s r S_X H_s^T + S_{\text{int}}^{(1,1)} + \sigma^2 I_r - (H_s r S_X H_s^T + S_{\text{int}}^{(1,1)})(H_{sd} S_X H_{sd}^T + S_{\text{int}}^{(2,2)} + \sigma^2 I_d)^{-1}(H_{sd} S_X H_{sd}^T + S_{\text{int}}^{(2,1)}),$$

(25)

and

$$S_{Y,|Y_d, X} = H_s r S_X H_s^T + S_{\text{int}}^{(1,1)} + \sigma^2 I_r - \left[ H_s r S_X H_{sd}^T + S_{\text{int}}^{(2,2)} H_{sd} S_X \right]^{-1} \left[ H_{sd} S_X H_{sd}^T + S_{\text{int}}^{(2,1)} \right].$$

(26)

The main step in obtaining the global optimum of \((24)\) in closed form is the following simultaneous diagonalization by *congruence of $S_{Y,|Y_d}$ and $S_{Y,|Y_d, X}$ based on \([6]\) Corollary 7.6.5).

**Lemma 1** There exists a non-singular matrix $C_r \in \mathbb{C}^{r \times r}$ such that $C_r^t S_{Y,|Y_d, X} C_r = I_r$ and $C_r^t S_{Y,|Y_d} C_r = \Lambda$, where $\Lambda$ is a diagonal matrix. The diagonal elements $\lambda_i$ in decreasing order are called the generalized eigenvalues. Moreover, $\lambda_i \geq 1$ for $i = 1, \ldots, r$.

**Proof:** Both $S_{Y,|Y_d}$ and $S_{Y,|Y_d, X}$ are positive definite matrices. Let $S_{Y,|Y_d, X}^{-1} = R^T R$ be the Cholesky decomposition. Now, consider the eigendecomposition $R S_{Y,|Y_d, X} R^T = V \Lambda V^T$. Then, $C_r R^T$ satisfies $C_r^t S_{Y,|Y_d, X} C_r = I_r$ and $C_r^t S_{Y,|Y_d} C_r = \Lambda$ simultaneously. Moreover, $S_{Y,|Y_d} \succeq S_{Y,|Y_d, X}$ implies $\Lambda \succeq I_r$. \(\Box\)

The above result allows the MIMO relay channel to be reduced to the parallel scalar channels. We can now use the approach of \([3], [20]\) to solve the subsequent scalar quantization noise optimization problem.

**Theorem 6** Let $C_r$ be the matrix that simultaneously diagonalizes $S_{Y,|Y_d, X}$ and $S_{Y,|Y_d}$, and $\lambda_i$’s be the generalized eigenvalues. The optimal quantization scheme that maximizes Lagrangian in \((27)\) for fixed $S_X$ involves first multiplying $Y_r$ by $C_r$, then independently quantizing the $i^{th}$ elements of $C_r Y_r$ at rate

$$c_i^r = \left[ \log(\lambda_i - 1) - \log \frac{\mu}{1 - \mu} \right]^+. $$

(27)

The corresponding normalized quantization noise variance for describing the $i^{th}$ element is

$$\Sigma_{Q_i}^{*, i} = \begin{cases} \frac{1}{1 - \frac{\mu}{\lambda_i}} & \mu < 1 - \frac{1}{\lambda_i} \\ +\infty & \mu \geq 1 - \frac{1}{\lambda_i} \end{cases}. $$

(28)

The global optimum of \((27)\) is achieved by $S_Q^{*} = C_r^{-1} \Sigma_Q^{*, i} C_r^{-1}$.

**Proof:** For $\mu \in (0, 1)$, the Lagrangian \((24)\) can be written as

$$\mathcal{L} \overset{(a)}{=} (1 - \mu) \log |\lambda S_{Q}^{-1} + I_r| - \log |\hat{S}_Q^{-1} + I_r| + \text{const.}$$

$$\overset{(b)}{=} (1 - \mu) \log |\lambda S_{Q}^{-1} + I_r| - \log |\Sigma_Q^{-1} + I_r| + \text{const.}$$

where $(a)$ follows from the change of variable $\hat{S}_Q = C_r^t S_Q C_r$, with $C_r$ as in Lemma \([1]\) and $(b)$ follows from \([20]\) Lemma 5) where $\Sigma_Q$ comes from the eigen-decomposition $\hat{S}_Q = U \Sigma_Q U^T$. Observe that the equality in $(b)$ is obtained with $U = I_r$. Thus, it is without loss of optimality to restrict $\hat{S}_Q$ to be diagonal.
Consider the change of variable [4]

$$c_i = \log \left(1 + \frac{\lambda_i}{\Sigma_{Q_i}}\right), \ i = 1, \ldots, r,$$  

where $\Sigma_{Q_i}$’s are the diagonal entries of $\Sigma_Q$. An interpretation of $c_i$ is that it is the portion of the available $C_0$ allocated for quantization of the $i^{th}$ element of $C_r Y_r$. Using (29), the Lagrangian can be written as

$$L = \sum_{i=1}^r ((1 - \mu) c_i - \log (2^{c_i} + \lambda_i - 1)) + \text{const.}$$  

(30)

It can be readily checked that (30) is concave in $c_i$ for $\lambda_i \geq 1$. It is easy to see that the optimal $c_i^*$ is (27) and, therefore, $\Sigma_{Q_i}^{ii,*}$ is given by (28). The optimal $S_Q$ is $S_Q^* = C_r^{-1} \Sigma_{Q} C_r^{-1}$. \hfill \Box

The above solution illustrates that the optimal compression involves a simultaneous diagonalization of the vector channel together with an optimal allocation of relay link capacity according to (27), which allocates the available $C_0$ for quantizing elements of $C_r Y_r$ using reverse water-filling on $\log (\lambda_i - 1)$’s. An example of such a reverse water-filling rate allocation is shown in Fig. 2. Here, $\lambda_i - 1$ can be thought of as the $i^{th}$ element’s CSINR given $Y_d$. To illustrate this interpretation of CSINR using an example, consider the case of $r = 1$, where we have

$$\lambda_1 - 1 = \frac{s_{Y_r,Y_d}}{s_{Y_r,Y_d,X}} - \frac{s_{Y_r,X}}{s_{Y_r,Y_d,X}}.$$  

(31)

It can be seen that $\lambda_1 - 1$ is the ratio of the conditional variance of signal to the conditional variance of interference plus noise given $Y_d$. In the next section, we explain how the magnitude of generalized eigenvalues determines the performance of MIMO compress-and-forward relaying.

A special case of this optimization problem is considered in [4], where the noises at $Y_r$ and $Y_d$ are independent and $S_{Y_r,Y_d,X}$ is an identity matrix. In this case $S_{Y_r,Y_d,X}$ and $S_{Y_r,Y_d}$ can be diagonalized simultaneously by the CKLT of
Y_r given Y_d, which is a unitary transformation [5]. For the more general correlated noise case, the above simultaneous diagonalization is needed to transform the matrix optimization problem to scalar optimization.

We note that in [8] a diagonalization approach, known as CCA, is taken to solve the same problem, but from a source coding perspective. Instead of diagonalizing S_Y_r Y_d X and S_Y_r Y_d, the approach of [8] diagonalizes S_X Y_d and S_Y_r Y_d using a singular value decomposition of the matrix S_X Y_d K_{XY} S_Y_r Y_d, where K_{XY_r} is a certain matrix of regression coefficients. It can be shown that the resulting diagonalization makes S^{(1,2)}_{XY_r Y_d} diagonal as well. Subsequently, the diagonal elements from the diagonalization of S^{(1,2)}_{XY_r Y_d} are used to find the optimal solution to the overall problem. We observe that the CCA approach in [8] can be interpreted as an indirect simultaneous diagonalization of S_Y_r Y_d X and S_Y_r Y_d, and, therefore, is equivalent to the transformation presented here. However, the direct simultaneous diagonalization of S_Y_r Y_d X and S_Y_r Y_d is simpler and gives more structural insight about the optimal MIMO compress-and-forward strategy in Section IV.

IV. INTERPRETING THE GENERALIZED EIGENVALUES

It is mentioned in the last section that \( \lambda_i - 1 \) can be interpreted as the CSINR at the \( i \)-th element of \( C_r Y_r \). The reverse water-filling solution with respect to CSINRs reveals significant insight into the effectiveness of MIMO compress-and-forward relaying for improving the overall rate, as explained below.

A. The Slope of \( \bar{R}_{CF}(C_0) \)

For the MIMO relay channel under study, the following theorem relates the slope of \( \bar{R}_{CF}(C_0) \) to the magnitude of the generalized eigenvalues \( \lambda_i \)’s.

**Theorem 7** In the optimal allocation of the relaying capacity \( C_0 \) to elements of \( C_r Y_r \) for quantization, given in (27), let \( \bar{C}_{0,i} \) be the largest \( C_0 \) for which the optimal quantization rate of the \( i \)-th element is zero, i.e., \( c_i > 0 \) iff \( C_0 > \bar{C}_{0,i} \). We have

\[
\bar{C}_{0,i} = \sum_{j=1}^{i-1} \log \lambda_j - \frac{1}{\lambda_i - 1}, \quad i = 2, \ldots, r,
\]

and

\[
\frac{d \bar{R}_{CF}(C_0)}{d C_0} \bigg|_{C_0 = \bar{C}_{0,i}} = 1 - \frac{1}{\lambda_i},
\]

where \( \bar{R}_{CF}(C_0) \) is (9) evaluated at given \( S_X \).

**Proof:** By Theorem 6, the optimization problem (21) can be transformed into a convex problem, therefore, strong duality holds. Moreover, using solution (27) the resulting \( \bar{R}_{CF}(C_0) \) is differentiable over \( C_0 > 0 \). Therefore, the optimal Lagrange multiplier in the KKT solution for a given \( C_0 \) characterizes the slope of \( \bar{R}_{CF}(C_0) \) at \( C_0 \) (21), i.e.,

\[
\frac{d \bar{R}_{CF}(C_0)}{d C_0} = \mu^*(C_0).
\]

When \( C_0 \) is such that \( c_i^* = 0 \) and \( c_i^* + \cdots + c_{i-1}^* = C_0 \) for \( 2 \leq i \leq r \), using the KKT solution (27), the Lagrange multiplier is given by

\[
\mu^*(C_0) = \frac{1}{1 + \left( \frac{2 C_0}{\prod_{i=1}^{r} (\lambda_i - 1)} \right)^{\frac{1}{r}}}. \tag{35}
\]

Now, \( \bar{C}_{0,i} \) is the largest value of \( C_0 \) for which the optimal quantization rate of the \( i \)-th element is zero, \( c_i^* = 0 \). Using (27) again, we have

\[
\log \frac{1 - \mu^*(\bar{C}_{0,i})}{\mu^*(C_0)} + \log(\lambda_i - 1) = 0.
\]

This together with (34) and (35) yields (32) and (33).

Fig. 7 illustrates this result. The slope of the compress-and-forward achievable rate at various points of \( C_0 \) is determined by the generalized eigenvalues of the conditional covariance matrices. A larger \( \lambda_i \) implies a slope closer to the maximum value of 1.
B. Zero CSINR: Reversely Degraded Components

When \( \lambda_i(S_{Y_r|Y_d}S_{Y_r|Y_d}^{-1}) = 1 \), the corresponding element in \( C_rY_r \) is conditionally independent of \( X \) given \( Y_d \). Since such an element does not convey any further information to destination, the optimized compress-and-forward does not assign any portion of \( C_0 \) for describing it.

**Definition 1** The \( i^{th} \) element of \( C_rY_r \) is called reversely degraded, if \( \lambda_i(S_{Y_r|Y_d}S_{Y_r|Y_d}^{-1}) = 1 \).

When the relay node is equipped with large number of antennas, the number of elements of \( C_rY_r \) to be quantized, i.e., elements that are not reversely degraded, can be much smaller than its dimension. The following proposition bounds the number of reversely degraded components in the relay channel. It shows that the optimal scheme quantizes no more than \( \min(r, s^t) \) elements. This result is used for the constant gap characterization of the capacity in Section VII as well.

**Theorem 8** In the MIMO relay channel under study, the number of reversely degraded components is at least \( (r - s^t)^+ \), where \( s^t \) is the number of independent data streams transmitted by the source, i.e., \( \text{rank}(S_X) = s^t \). Hence, the optimal quantization scheme in (27) describes at most \( \min(r, s^t) \) elements of \( C_rY_r \).

**Proof:** By Lemma [1], \( C_r (S_{Y_r|Y_d} - S_{Y_r|Y_d}^r)C_r^t = \Lambda - I_r \). Therefore, the number of reversely degraded components of the channel is equal to \( r - \text{rank}(C_r (S_{Y_r|Y_d} - S_{Y_r|Y_d})C_r^t)) = r - \text{rank}(S_{Y_r|Y_d} - S_{Y_r|Y_d}) \). Now, using the generalized Schur complement formula, we have

\[
S_{Y_r|Y_d} - S_{Y_r|Y_d} = S_{Y_r|Y_d}^{-1}S_{Y_r|Y_d}^{-1}.
\]

(37)

The result follows by noting that \( \text{rank}(S_{Y_r|Y_d} - S_{Y_r|Y_d}) \leq \min(r, s^t) \). \( \square \)

C. Infinite CSINR: Asymptotically Deterministic Components

Relaying is particularly effective in certain situations where the relay observation \( Y_r \) is a deterministic function of \( (Y_d, X) \). This happens for example in the asymptotic case for the SISO channel, where we have \( (s, d, r, t) = (1, 1, 1, 1) \) and

\[
Y_r = \left( h_{sr} - \frac{h_{sr}h_{sd}}{h_{td}} \right) X + \frac{h_{tr}}{h_{td}} Y_d + N_1 - \frac{h_{tr}}{h_{td}} N_2.
\]

(38)
for \( h_{td} > 0 \). When the noise power is zero, we have \( Y_r = f(X, Y_d) \). It can be shown that compress-and-forward for relay channels with this type of determinism can achieve the cut-set upper bound \([15]\). When this happens, every relay bit is worth one bit improvement of overall transmission rate (except in some degenerate cases). In other words, the slope of \( R_{CF}(C_0) \) is equal to 1, which can be verified for the above example by noting that \( s_{Y_r|Y_d} \) and \( s_{Y_r|Y_d} \) are noiseless parts of relay and destination observations, respectively. Hence, for Theorem 11 that the design of the combining matrix at the relay is crucial for achieving the optimal DoF. If one does not combine the

\[
\sigma^{1+
\frac{h_{td}}{2}} \left| P + \frac{1-h_{td}}{2} \right|^2 \sigma^2.
\]

Thus, \( \lim_{\sigma^2 \to 0} \lambda_1 = \frac{s_{Y_r|Y_d} \lambda_1}{s_{Y_d|Y_d}} = \infty \) (assuming \( h_{sr}, h_{td} \neq h_{td} h_{tr} \)). Hence, \( \lim_{\sigma^2 \to 0} \frac{dR_{CF}(C_0)}{dC_0} \bigg|_{C_0=0} = 1 \) by Theorem 7.

The following result reveals the number of asymptotic deterministic components of this type for the MIMO relay channel under consideration.

**Definition 2** The \( i^{th} \) element of \( C_r Y_r \) is called asymptotically deterministic, if \( \lim_{\sigma^2 \to 0} \lambda_i(S_{Y_r|Y_d} S_{Y_r|Y_d}^{-1}) = \infty \).

**Theorem 9** In the MIMO relay channel under study, the number of asymptotically deterministic components is

\[
r' - r'' = \min \left( \frac{r, s^t}{r + d - t}, \left( s^t + t - d \right)^+ \right),
\]

almost surely, where

\[
r' \triangleq \text{rank}(S_{Y_r|Y_d}) = \min \left( r, \left( s^t + t - d \right)^+ \right),
\]

\[
r'' \triangleq \text{rank}(S_{Y_r|Y_d}) = \min \left( r, \left( t - d \right)^+ \right).
\]

Here,

\[
Y_r = H_{sr} X + H_{tr} X_t
\]

\[
Y_d = H_{sd} X + H_{td} X_t
\]

are noiseless parts of relay and destination observations, respectively. Hence, for \( i = 1, \ldots, r' - r'' \), the slope \([33]\) approaches 1 as \( \sigma^2 \to 0 \).

**Proof:** See Appendix [B]

A key implication of the above result is the following. For MIMO relay channel under consideration, relaying with small \( C_0 \) is most effective if \( \frac{dR_{CF}(C_0)}{dC_0} \bigg|_{C_0=0} \) is at the maximum value of 1. This happens when \( \lambda_1 \to \infty \), i.e., when the MIMO relay channel has at least one asymptotic deterministic component, or equivalently, by Theorem 9 when \( r' > r'' \). By (39), this holds when \( r + d > t \) and \( d < s^t + t \). We state this result below and will return to the interpretation of this condition in Section VI.

**Corollary 10** For the MIMO relay channel under study, \( \lim_{\sigma^2 \to 0} \frac{dR_{CF}(C_0)}{dC_0} \bigg|_{C_0=0} = 1 \) iff \( r + d > t \) and \( d < s^t + t \).

**V. DoF Improvement by Relaying**

We gain further insight into the benefit of relaying by analyzing the DoF of the MIMO relay channel. Consider the DoF of the channel under study without the relay, and with a relay that has infinite link capacity, i.e., respectively,

\[
\text{DoF}_D \triangleq \lim_{\rho \to \infty} \frac{R_{CF}(0)}{\log \rho} = \min \left( s, (d - t)^+ \right),
\]

\[
\text{DoF}_R \triangleq \lim_{\rho \to \infty} \frac{R_{CF}(\infty)}{\log \rho} = \min \left( s, (r + d - t)^+ \right).
\]

At infinite relay link capacity, the DoF gain due to relaying can be seen to be

\[
\text{DoF}_R - \text{DoF}_D = \min \left( r, s, (r + d - t)^+, (s + t - d)^+ \right).
\]

Curiously, the above expression is identical to the number of asymptotically deterministic components derived in the previous section. We will return to this connection in the next section.

For the finite relay link capacity case, we now aim to characterize the overall DoF gain due to relaying. First, we note that the design of the combining matrix at the relay is crucial for achieving the optimal DoF. If one does not combine the observation at the relay and instead implements naive i.i.d. quantization on a per-antenna basis, it would result in a DoF loss in general.

**Theorem 11** For the relay channel under study, the achievable DoF improvement by compress-and-forward with relay’s quantization scheme \([2]\) restricted to \( Q \sim CN(0_{r \times 1}, q_I) \) is

\[
\Delta \text{DoF}_{i.i.d.} = (\text{DoF}_R - \text{DoF}_D) \min \left( 1, \frac{\alpha}{\rho^2} \right)
\]

(47)
almost surely, where \( r' = \text{rank}(S_{Y_r|Y_d}) = \min(r, s + t - d) \) and \( \alpha \) is the DoF of the relaying link capacity \( C_0 \) as in [15].

Proof: See Appendix [C]

Note that when \( d < t \), the coefficient of \( \alpha \) in (47) is

\[
\frac{(\text{DoF}_R - \text{DoF}_D)}{r'} = \frac{\min(s, r + d - t)}{\min(r, s + t - d)} < 1,
\]

whereas, by the next theorem, the coefficient of \( \alpha \) in the optimal DoF gain is 1. This loss in DoF is due to the non-optimized choice of quantization scheme.

The optimal DoF can be achieved by using a combining matrix \( \hat{C}_r \) of dimension \( \min(r, s, (r + d - t)^+) \times r \) followed by i.i.d. quantization at the relay based on the principle of aligning the observed interference at the relay with the row space of the observed interference at the destination. The optimality of this scheme follows by the cut-set bound (5). Details of the proof are deferred to Appendix [D] We call this combining strategy distributed zero-forcing, because in zero-forcing of interference with \( r + d \) antennas pooled together, the combining matrix \( C_r \) is the submatrix corresponding to \( r \) antennas of the relay.

Theorem 12 For the relay channel under study, the optimal DoF improvement by relaying is

\[
\Delta \text{DoF}^* = \min(\text{DoF}_R - \text{DoF}_D, \alpha)
\]

almost surely.

Proof: See Appendix [D]

VI. CONNECTION BETWEEN DOF IMPROVEMENT AND ASYMPTOTICALLY DETERMINISTIC COMPONENTS

Combining the results of the previous two sections, we see that the DoF improvement at infinite relay link rate is equal to the number of asymptotically deterministic components in the MIMO relay channel. Thus, interestingly, the DoF improvement at \( C_0 = \infty \) is related to the slope of the achievable rate at small \( C_0 \). In this section, we first provide an information theoretical proof of this equivalence, then further explore this connection via the DoF optimal distributed zero-forcing combiner design at the relay.

Theorem 13 For the MIMO relay channel under study, the maximum DoF improvement achieved with an infinite capacity relay link is equal to the number of asymptotically deterministic components, i.e.,

\[
\text{DoF}_R - \text{DoF}_D = r' - r''.
\]

The DoF gain, or equivalently the number of asymptotically deterministic components, is greater than zero if and only if \( d < s + t \) and \( r + d > t \).

Proof: Fix \( S_X \) to be full-rank. By Theorem[9] the number of asymptotically deterministic components is \( r' - r'' \), the difference of ranks of \( S_{Y_r|Y_d} \) and \( S_{Y_r|Y_d,X} \). To relate this to \( \text{DoF}_R - \text{DoF}_D \), we expand the conditional mutual information below in two different ways

\[
I(X; Y_r | Y_d) = I(X; Y_r, Y_d) - I(X; Y_d),
\]

\[
I(X; Y_r | Y_d) = h(Y_r | Y_d) - h(Y_r | Y_d, X).
\]

The DoF of the first equality is \( \text{DoF}_R - \text{DoF}_D \), while the DoF of the second is \( r' - r'' \).

We now further explore this connection via the concept of determinism in the relay channel. As already mentioned, for a relay channel with \( Y_r = f(Y_d, X) \), compress-and-forward improves the rate in a 1:1 ratio of \( C_0 \) as long as \( I(X; Y_r, Y_d) > I(X; Y_d) \) for some \( p(x) \) [15]. A relay channel with \( Y_r = f(Y_d) \), however, is reversely degraded and its capacity, given by \( \max_{p(x)} I(X; Y_d) \), does not depend on \( C_0 \) [11].

For the MIMO relay channel considered here, when \( d \geq t + s \), the channel [42]-[43] is reversely degraded, because

\[
Y_r = [H_{sr} H_{tr}] \left[ \begin{bmatrix} H_{sd}^d \\ \end{bmatrix} H_{td} \right] \left[ H_{td} \right]^{-1} \left[ \begin{bmatrix} H_{td}^d \\ \end{bmatrix} \right] Y_d.
\]

In this case, the slope of \( \tilde{R}_{CF}(C_0) \) does not approach one.
When \( d < t + s \), the channel \([42, 43]\) can contain deterministic components that are not reversely degraded. It turns out that the DoF-optimal distributed zero-forcing scheme mentioned in Section\( \text{V} \) can reveal such deterministic components. More specifically, we first form \( \bar{Y}_r' \), defined to be the component of \( \bar{Y}_r \) that is independent of \( \bar{Y}_d \), i.e.,

\[
\bar{Y}_r' = \bar{Y}_r - S_{\bar{Y}_r, \bar{Y}_d}^{(1,2)} S_{\bar{Y}_d}^{-1} \bar{Y}_d \\
= \begin{bmatrix} H_{sr} H_{tr} \end{bmatrix} \bar{P} \begin{bmatrix} X \\ X_t \end{bmatrix} \triangleq \begin{bmatrix} H_{sr} H_{tr} \end{bmatrix} \bar{X} \\
= \begin{bmatrix} X \\ X_t \end{bmatrix},
\]

(54)

where

\[
\bar{P} = I_{s+t} - \begin{bmatrix} S_X H_{sd}^H \\ S_X H_{sd} H_{td} \end{bmatrix} S_{\bar{Y}_d}^{-1} \begin{bmatrix} H_{sd} H_{td} \end{bmatrix}.
\]

(55)

is a projection matrix. Now, consider the effective channel

\[
\begin{bmatrix} \bar{Y}_r' \\ \bar{Y}_d \end{bmatrix} = \begin{bmatrix} H_{sr} H_{tr} \\ H_{sd} H_{td} \end{bmatrix} \begin{bmatrix} X \\ X_t \end{bmatrix}.
\]

(56)

If \( r + d > t \), then we can select a combining matrix \( \bar{C}_r \in \mathbb{C}^{(r+d-t) \times r} \) such that

\[
[\bar{C}_r \ \bar{A}] \begin{bmatrix} H_{tr} \\ H_{td} \end{bmatrix} = 0,
\]

(57)

in which case

\[
\bar{C}_r \bar{Y}_r' = [\bar{C}_r \ \bar{A}] \begin{bmatrix} H_{sr} \\ H_{sd} \end{bmatrix} X = \bar{A} \bar{Y}_d
\]

(58)

is the component of \( \bar{Y}_r \) that is a function of \( \bar{Y}_d \), \( X \), but not a function of \( \bar{Y}_d \) alone. Such deterministic component gives rise to the slope of \( R_{CF}(C_0) \) at \( C_0 = 0 \) to be 1. It exists if and only if both \( r + d > t \) and \( d < s + t \).

We now see that using distributed zero-forcing combining at the relay not only gives rise to DoF-optimal relaying strategy, but also reveals the deterministic components of the MIMO relay channel. It is for this reason that the condition for having DoF improvement by \( C_0 = \infty \) is the same as the condition under which relay has deterministic components and has \( \lim_{s \to 0} \frac{d R_{CF}(C_0)}{d C_0} = 1 \) at small \( C_0 \).

From DoF point of view, one can interpret the above conditions on the numbers of antennas in the following way. To improve DoF by adding \( r \) more antennas to the destination, the destination must not already have sufficient antennas to be able to completely mitigate interference and to resolve the intended signal, hence \( d < s + t \) is required. Further, the relay and the destination together must have enough antennas to mitigate all of the interference, so \( r + d > t \) is needed. Thus, we need \( r + d > t \) and \( d < s + t \), as otherwise adding \( r \) more antennas to the receiver cannot improve the spatial DoF.

### VII. Constant Gap Characterization of Capacity

Throughout this paper, we have focused on the compress-and-forward relaying strategy. In this final section of the paper, we show that compress-and-forward can achieve the capacity of the Gaussian MIMO relay channel with noise correlation to within a constant gap.

**Theorem 14** For the MIMO relay channel under study, the achievable rate using optimized input and quantization covariance matrices given by Algorithm\[1\] is to within \( \min(r, s) \) bits of the capacity.

**Proof:** See Appendix\[E\]

### VIII. Numerical Examples

In this section, we provide numerical examples of MIMO relaying in a wireless cellular environment. We simulate a picocell network with a BS transmitting at a maximum power of 1Watt over 10MHz to a user located 100m away. A relay node, located 10m away from the user, pools antennas with user's receiver over a digital link of capacity \( C_0 \). Here, \( C_0 \) is normalized by the 10MHz of bandwidth and is expressed in b/s/Hz. This scenario is shown in Fig.\[1\] The background noise power spectral density is \(-170dBm/Hz\). A channel model with path loss exponent of 3.76 and 8dB shadowing is used.
A. Relaying for Interference Mitigation and Signal Enhancement

Fig. 4 illustrates the effect of antenna pooling for enhancing the intended signal when there is no interference, so noises at relay and destination are independent. It shows the rate improvement by relaying in a channel with \((s, d, r, t) = (3, 2, 2, 0)\). Without the relay, DoF is limited by the number of antennas at the destination, which is \(\text{DoF}_D = 2\). Adding \(r = 2\) extra antennas to the destination improves DoF by \(\Delta \text{DoF} = 1\). Also, this channel has 1 asymptotically deterministic component, and as shown in Fig. 4, at \(C_0 = 10\) b/s/Hz, the improvement in throughput by the optimized compress-and-forward is around \(9.2\) b/s/Hz, achieving an almost 1:1 improvement in the overall rate for each relaying bit. Here, antenna pooling improves the overall throughput considerably.

Fig. 4 also demonstrates the importance of optimizing the input covariance matrix. We plot the achievable rates for two suboptimal choices of input covariance; the water-filling covariance of the source-to-destination and source-to-relay-and-destination point-to-point channels. For these fixed input covariance matrices, we optimize the quantization noise covariance. Depending on the value of \(C_0\), each of them can be strictly suboptimal. For small values of \(C_0\), steering the beam toward destination is close to optimal. However, as \(C_0\) increases, this \(S_X\) fails to achieve benefits of antenna pooling. For large values of \(C_0\), steering the beam toward both relay and destination is close to optimal, but a gap exits at small \(C_0\).

Fig. 5 presents the results for a channel with \((s, d, r, t) = (2, 3, 3, 0)\). Without the relay, DoF is constrained by the number of BS antennas, not the number of antennas at the destination. Therefore, relaying does not increase the overall DoF. Also, as predicted by Theorem 13, the improvement in overall throughput is not notable. In this case, with \(C_0 = 6\) b/s/Hz, the throughput improvement is only around \(1.7\) b/s/Hz.

Fig. 6 considers the same setup as Fig. 5 except that now four interfering single-antenna BS’s are added and we have \((s, d, r, t) = (2, 3, 3, 4)\). BS’s are placed on a hexagonal grid, with minimum BS-to-BS distance of 200m. Due to interference,
the throughput at $C_0 = 0$ is considerably lower, but the optimized use of the relay link is able to improve the throughput significantly. This is because, due to the common intercell interference, now the noises at the relay and destination are correlated. By exploiting such noise correlation in using the relay link, the destination can effectively pool the $d = 3$ existing antennas of its own to improve DoF. The maximum possible improvement is around 16.7 b/s/Hz, which is achieved at $C_0 = 25$ b/s/Hz. Here, the DoF improvement is $\Delta\text{DoF} = 2$ and the channel has 2 asymptotically deterministic components. At around $C_0 = 10$ b/s/Hz, the improvement in throughput is around 9.7 b/s/Hz. It is worth noting that the overall throughput at large $C_0$ in Fig. 6 is close to the achievable rate of the $C_0 = 0$ scenario in Figs. 4 and 5 illustrating the almost complete interference rejection capability of relaying.

Figs. 5 and 6 also demonstrate the importance of optimizing the quantization noise covariance matrix. The achievable rate for a simple suboptimal choice of $S_Q = qI_r$ is included, where $q$ is set to satisfy the relaying rate constraint with equality, for the optimal $S_X$ obtained from Algorithm 1. This simple choice of $S_Q$ results in a strictly suboptimal performance as shown in Figs. 5 and 6. By Theorem 11, it wastes the relaying link’s DoF in scenario of Fig. 6.

B. Effect of Numbers of Antennas on the Slope

We illustrate the effect of number of antennas at the relay ($r$) and at the destination ($d$) on the slope of compress-and-forward rate versus relaying link capacity $\bar{R}_{CF}(C_0)$. In this example, the source is equipped with five antenna and there are 18 interfering BS’s each equipped with one antenna, i.e., $(s, t) = (5, 18)$. The interfering BS’s are placed on a hexagonal grid, with minimum BS-to-BS distance of 200m. Fig. 7 shows the value of $1 - \frac{1}{\lambda_i}$ for $i = 1, \ldots, 6$, which is the slope of $\bar{R}_{CF}(C_0)$ at $\bar{C}_i$ as in (33), averaged over 100 realizations of the channel and fixed transmit covariance $S_X$. 

![Fig. 5. Throughput improvement versus relaying capacity $C_0$ for the case where the source has two antennas ($s = 2$), and the relay and destination are each equipped with three antennas ($r = 3$ and $d = 3$) with no inter-cell interference ($t = 0$).](image-url)
Since the source has $s = 5$ antennas, by Theorem 8 the 6th element of $C_r Y_r$ is always reversely degraded and its CSINR is zero, i.e., $\lambda_6 = 1$. The optimal quantization does not quantize this element, because it cannot further improve the throughput. In results of simulations in Fig. 7(f), the slope of rate curve at $C_0$ is almost zero.

For $i = 1, \ldots, 5$, by (39) in Theorem 9 when $r$ and $d$ are such that both $r + d > 17 + i$ and $d < 24 - i$, the $i$th element of $C_r Y_r$ is asymptotically deterministic. Hence, the value of $1 - \frac{1}{\lambda_i}$ approaches 1 as the power of noise goes to zero. When $d \geq 24 - i$ the $i$th element of $C_r Y_r$ is asymptotically reversely degraded, and $1 - \frac{1}{\lambda_i}$ does not approach 1. Also, when $r + d \leq 17 + i$, the 18-dimensional interference signal makes both $S_{Y_r | Y_d X}$ and $S_{Y_r | Y_d}$ full rank, and keeps the slope $1 - \frac{1}{\lambda_i}$ away from 1. These are indeed observed by numerical simulations in Figs. 7(a)-(e).

IX. CONCLUSION

This paper considers MIMO relaying for interference mitigation and signal enhancement in a wireless communication network. Joint optimization of transmit and quantization noise covariance matrices in compress-and-forward scheme enables the receiver to efficiently utilize extra spatial dimensions from a relay node. This paper shows that this scheme achieves the capacity of the channel to within a constant gap.

Optimizing the relay’s quantization noise covariance matrix is crucial for efficient interference mitigation. This is the key step in characterizing the capacity of the MIMO relay channel to within the constant gap. It further reveals the asymptotically deterministic and reversely degraded components of the channel. When the channel has an asymptotically deterministic component, the slope of the optimized compress-and-forward rate $R_{CF}(C_0)$ at small $C_0$ asymptotically approaches its maximum of 1 at high SNR and INR. Finally, when the number of destination’s antennas is less than the dimension of the interference,
optimizing relay’s quantization enables achieving maximum possible DoF of the channel through distributed zero-forcing of interference.

Antenna pooling is most efficient when the number of antennas at the source $s$, at the destination $d$, at the relay $r$, and the dimension of interference $t$ satisfy both $r + d > t + i - 1$ and $d < s + t - i + 1$ for $i = 1, \ldots, 5$, and is close to zero for $i = 6$.

**APPENDIX A**

The following lemma characterizes the scaling of eigenvalues of the two conditional covariance matrices at the relay with respect to the background noise power, as it goes to zero. It is used for counting the number of asymptotic deterministic components and for the DoF analysis in Appendices B to D.

**Lemma 2** At the limit of high SNR and INR, i.e., as $\sigma^2 \to 0$, we have

$$
\lambda_i(S_{\bar{Y}_r | \bar{Y}_d}) = \lambda_i(S_{\bar{Y}_r | \bar{Y}_d}) + a_i^2 + O(\sigma^4),
$$

$$
\lambda_i(S_{\bar{Y}_r | \bar{Y}_d, x}) = \lambda_i(S_{\bar{Y}_r | \bar{Y}_d, x}) + a_i^2 + O(\sigma^4),
$$

for positive $a_i$’s and $a_i^2$’s. Here, $\bar{Y}_r$ and $\bar{Y}_d$ are as defined in (42)-(43).
Proof: Define \( L \triangleq \begin{bmatrix} H_{sd}S_{X}^\frac{1}{2} & H_{td}S_{X}^\frac{1}{2} \\ H_{td}S_{X}^\frac{1}{2} & H_{tr}S_{X}^\frac{1}{2} \end{bmatrix} \) and \( M \triangleq \begin{bmatrix} H_{sr}S_{X}^\frac{1}{2} & H_{tr}S_{X}^\frac{1}{2} \end{bmatrix} \). We have
\[
S_{Y,\tau\mid Y_d} = S_{\bar{Y},\tau} + \sigma^2I_r - ML^\dagger (LL^\dagger + \sigma^2I_d)^{-1} LM^\dagger \\
= S_{\bar{Y},\tau} + \sigma^2I_r - MVD^\dagger (DD^\dagger + \sigma^2I_d)^{-1} DV^\dagger M^\dagger \\
= S_{Y,\tau\mid Y_d} + \sigma^2I_r - ML^\dagger \left( \sum_{n=0}^{\infty} (-\sigma^2)^n (DD^\dagger)^{-(n+1)} \right) DV^\dagger M^\dagger \\
\triangleq S_{Y,\tau\mid Y_d} + \sum_{n=1}^{\infty} (-1)^{n+1} \sigma^{2n} S_n, \tag{61}
\]
where, in \((a)\) we used Schur complement formula, in \((b)\) we use SVD \( L = UDV^\dagger \), and \((c)\) follows from Taylor expansion of each diagonal element.

Similar to the above, by taking SVD of \( H_{td}S_{X}^\frac{1}{2} \) one can write
\[
S_{Y,\tau\mid Y_d, X} = S_{\bar{Y},\tau\mid Y_d, X} + \sum_{n=1}^{\infty} (-1)^{n+1} \sigma^{2n} S'_n. \tag{62}
\]
All \( S_n\)'s and \( S'_n\)'s are positive semidefinite, and both \( S_1\) and \( S'_1\) are full rank.

Since \( S_{Y,\tau\mid Y_d} \) and \( S_{Y,\tau\mid Y_d, X} \) have convergent series in \( \sigma^2 \) with Hermitian \( S_n\)'s and \( S'_n\)'s, the eigenvalues \( \lambda_i(S_{Y,\tau\mid Y_d}) \) and \( \lambda_i(S_{Y,\tau\mid Y_d, X}) \) have convergent series in \( \sigma^2 \) as well \[22, \text{Chapter 1}\]
\[
\lambda_i(S_{Y,\tau\mid Y_d}) = \sum_{n=0}^{\infty} a_{i,n} \sigma^{2n}, \tag{63}
\]
and
\[
\lambda_i(S_{Y,\tau\mid Y_d, X}) = \sum_{n=0}^{\infty} a'_{i,n} \sigma^{2n}. \tag{64}
\]

Using Weyl’s inequality \[6, \text{Theorem 4.3.1}\] over the summations in \((61)\) and \((62)\) inductively, we have
\[
\lambda_i(S_{Y,\tau\mid Y_d}) + \sum_{n=1}^{\infty} (-1)^{n+1} \sigma^{2n} \lambda_r(S_n) \leq \lambda_i(S_{Y,\tau\mid Y_d}) + \sum_{n=1}^{\infty} (-1)^{n+1} \sigma^{2n} \lambda_1(S_n), \tag{65}
\]
and
\[
\lambda_i(S_{Y,\tau\mid Y_d, X}) + \sum_{n=1}^{\infty} (-1)^{n+1} \sigma^{2n} \lambda_r(S'_n) \leq \lambda_i(S_{Y,\tau\mid Y_d, X}) + \sum_{n=1}^{\infty} (-1)^{n+1} \sigma^{2n} \lambda_1(S'_n), \tag{66}
\]
By the squeeze theorem, \( a_{i,0} = \lambda_i(S_{Y,\tau\mid Y_d}) \) and \( a'_{i,0} = \lambda_i(S_{Y,\tau\mid Y_d, X}) \). Moreover, \( a_{i,1} \) and \( a'_{i,1} \) are positive, because both \( S_1 \) and \( S'_1 \) are positive definite matrices. \(\square\)

**APPENDIX B\**

**PROOF OF THEOREM**

Before counting the number of asymptotic deterministic components, we characterize the rank of the two conditional covariance matrices in the following lemma.

**Lemma 3** Consider \( \bar{Y}_\tau \) and \( Y_d \) as defined in \((42)-(43)\). We have \( \text{rank}(S_{\bar{Y},\tau\mid Y_d}) = \min (r, (s^t + t - d)^+) \) and \( \text{rank}(S_{Y,\tau\mid Y_d, X}) = \min (r, (t - d)^+) \), almost surely.

**Proof:** Since \( \text{rank}(S_X) = s^t \) and \( \text{rank}(S_{X,t}) = t \), we have
\[
\text{rank}(S_{\bar{Y},\tau\mid Y_d}) = \min (r + d, s^t + t), \tag{67}
\]
\[
\text{rank}(S_{Y,\tau\mid Y_d}) = \min (d, s^t + t), \tag{68}
\]
\[
\text{rank}(S_{\bar{Y},\tau\mid Y_d, X}) = s^t + \min (r + d, t), \tag{69}
\]
\[
\text{rank}(S_{Y,\tau\mid Y_d, X}) = s^t + \min (d, t). \tag{70}
\]
When $S$ and $(34)$ are characterized in Lemma 3. Here, we argue that the largest we need to use bounds $S$ with equality if the null space of deterministic, but not asymptotically reversely degraded. By taking limit from both sides of (76) and using Lemma 2, we have due to [24, Corollary 2.5].

Now, we proceed to prove Theorem 9. The rank of the conditional covariance matrices $S_{Y_t|Y_d}$ and $S_{Y_t|Y_d,X}$ in (33) and (34) are characterized in Lemma 3. Here, we argue that the largest $i$ for which we have

$$\lim_{\sigma^2 \to 0} \lambda_i(S_{Y_t|Y_d}S_{Y_t|Y_d,X}) = \infty$$

is the difference of the two ranks, $r' - r''$. To relate the limit of the generalized eigenvalues to ranks of conditional covariances, we need to use bounds

$$\frac{\lambda_{r''+i}(S_{Y_t|Y_d})}{\lambda_{r'+i}(S_{Y_t|Y_d,X})} \leq \frac{\lambda_i(S_{Y_t|Y_d}S_{Y_t|Y_d,X})}{\lambda_i(S_{Y_t|Y_d})} = \frac{\lambda_i(S_{Y_t|Y_d})}{\lambda_i(S_{Y_t|Y_d,X})}$$

and

$$\frac{\lambda_i(S_{Y_t|Y_d}S_{Y_t|Y_d,X})}{\lambda_i(S_{Y_t|Y_d})} \leq \frac{\lambda_i(S_{Y_t|Y_d})}{\lambda_i(S_{Y_t|Y_d,X})}$$

due to [24, Corollary 2.5].

For $i \leq r' - r''$, we have $\lambda_{r'+i}(S_{Y_t|Y_d,X}) = 0$ and $\lambda_{r''+i}(S_{Y_t|Y_d}) > 0$. In this case, the $i^{th}$ element is asymptotically deterministic, but not asymptotically reversely degraded. By taking limit from both sides of (76) and using Lemma 2 we have

$$\lim_{\sigma^2 \to 0} \lambda_i(S_{Y_t|Y_d}S_{Y_t|Y_d,X}) \geq \lim_{\sigma^2 \to 0} \frac{\lambda_{r''+i}(S_{Y_t|Y_d})}{\lambda_{r'+i}(S_{Y_t|Y_d,X})} = \lim_{\sigma^2 \to 0} \frac{\lambda_{r''+i}(S_{Y_t|Y_d}) + a_{r''+1}\sigma^2}{a_{r'+1}\sigma^2} = \infty.$$

For $i > r' - r'' = \min \left( r, s', (r + d - t)^+, (s' + t - d)^+ \right)$, the generalized eigenvalues remain finite. To see this, we should consider three possible cases.

1) When $i > s'$, by Theorem 8, the $i^{th}$ element is reversely degraded and we have $\lambda_i = 1$ at all values of $\sigma^2$.

2) When $i > r + d - t$, using (40)-(41) it is easy to see that $r'' > r - i$ and $r'' > 0$. Therefore, both $\lambda_{r''+1}(S_{Y_t|Y_d,X}) > 0$ and $\lambda_i(S_{Y_t|Y_d,X}) > 0$. By Lemma 2 and (77)

$$\lim_{\sigma^2 \to 0} \lambda_i \leq \lim_{\sigma^2 \to 0} \frac{\lambda_i(S_{Y_t|Y_d,X}) + a_1\sigma^2}{\lambda_i(S_{Y_t|Y_d,X}) + a_{r''+1}\sigma^2} < \infty.$$ (80)

3) When $i > s' + t - d$, using (40) it is easy to see that $r'' < i$, therefore, $\lambda_i(S_{Y_t|Y_d,X}) = 0$. In this case, the $i^{th}$ element is asymptotically reversely degraded. By Lemma 2 and (78)

$$\lim_{\sigma^2 \to 0} \lambda_i \leq \lim_{\sigma^2 \to 0} \frac{a_1\sigma^2}{\lambda_i(S_{Y_t|Y_d,X}) + a_{r'}\sigma^2} < \infty.$$ (81)
Appendix C
Proof of Theorem 11

Fix the transmit covariance $S_X$ to a full-rank matrix. Consider i.i.d. quantization of elements of the relay’s observed vector, i.e., set $S_Q = qL_r$. To calculate DoF gain due to compress-and-forward relaying (15), we select $q$ such that the relaying capacity constraint in (9) is satisfied with equality

$$f_c(S_X, qL_r) = C_0(\rho).$$

(82)

Hence, $q$ depends on $\rho$. To characterize the asymptotic scaling of such a $q$ with $\rho$ (or equivalently with $\sigma^2$), we let

$$x = \lim_{\sigma^2 \to 0} \frac{\log(q)}{\log(\sigma^2)},$$

(83)

and obtain

$$x = \begin{cases} \frac{\sigma}{\sigma + r'}, & \alpha \leq r' \\ \frac{\sigma}{\sigma + r''}, & \alpha \geq r' \end{cases}$$

(84)

To see this, note that the DoF of $C_0(\rho)$ is $\alpha$ by (15). For the DoF of $f_c(S_X, qL_r)$, using Lemma 2 in Appendix A, as $\sigma^2 \to 0$ we have

$$f_c(S_X, qL_r) = \log \frac{|S_{Y_r}|Y_d + qL_r|}{|qL_r|}$$

$$= \sum_{i=1}^{r'} \log \frac{\lambda_i' + a_i'^2 + O(\sigma^4) + \sigma^2x}{\sigma^2x}$$

$$+ \sum_{i=r'+1}^{r} \log \frac{\lambda_i' + a_i'^2 + O(\sigma^4) + \sigma^2x}{\sigma^2x}.$$  

(85)

Therefore, the DoF of $f_c(S_X, qL_r)$ is

$$\lim_{\sigma^2 \to 0} \frac{f_c(S_X, qL_r)}{-\log(\sigma^2)} = r'x + (r - r')(x - 1)^+.$$  

(86)

Solving $r'x + (r - r')(x - 1)^+ = \alpha$, yields $x$.

Now, the desired DoF gain is

$$\Delta DoF_{i.i.d.} = \lim_{\sigma^2 \to 0} \frac{I(X; \hat{Y}_r | Y_d)}{-\log(\sigma^2)}.$$  

(87)

Using Lemma 2 as $\sigma^2 \to 0$,

$$I(X; \hat{Y}_r | Y_d) = \log \frac{|S_{Y_r}|Y_d + qL_r|}{|S_{Y_r}|Y_d, X + qL_r|}$$

$$= \sum_{i=1}^{r'} \log \frac{\lambda_i' + a_i'^2 + O(\sigma^4) + \sigma^2x}{\lambda_i'' + a_i''^2 + O(\sigma^4) + \sigma^2x}$$

$$+ \sum_{i=r'+1}^{r''} \log \frac{\lambda_i' + a_i'^2 + O(\sigma^4) + \sigma^2x}{\lambda_i'' + a_i''^2 + O(\sigma^4) + \sigma^2x}$$

$$+ \sum_{i=r'+1}^{r''} \log \frac{\lambda_i' + a_i'^2 + O(\sigma^4) + \sigma^2x}{\lambda_i'' + a_i''^2 + \sigma^2x + O(1)}.$$  

(88)

Hence, the DoF of $I(X; \hat{Y}_r | Y_d)$ is

$$\Delta DoF_{i.i.d.} = (r' - r'') \min(1, x) = (r' - r'') \min \left(1, \frac{\alpha}{r'} \right).$$  

(89)

Finally, by comparing (39) and (46) we have $r' - r'' = DoF_R - DoF_D.$
APPENDIX D
PROOF OF THEOREM 12

Using the cut-set upper bound (5), we have

$$\Delta D o F \leq \min (D o F_R - D o F_D, \alpha).$$

(90)

To almost surely achieve this upper bound by compress-and-forward, we transform $Y_r$ using a matrix $\tilde{C}_r \in C^{r \times r}$, then describe $\tilde{C}_r Y_r$ by

$$\tilde{Y}_r = \tilde{C}_r Y_r + Q,$$

(91)

with $Q \sim \mathcal{CN}(0_{r \times 1}, q_{I_r})$ independent of other variables and

$$\tilde{r} = \min (r, s, (r + d - t)^+).$$

(92)

Similar to the proof of Theorem 11 in Appendix C, we achieve

$$\Delta D o F = \left( \text{rank}(\tilde{C}_r S_{Y_r | Y_d, X} \tilde{C}_r^\dagger) - \text{rank}(\tilde{C}_r S_{Y_r | Y_d, X} \tilde{C}_r^\dagger) \right) \min \left( \frac{\alpha}{\text{rank}(\tilde{C}_r S_{Y_r | Y_d, X} \tilde{C}_r^\dagger)}, \frac{1}{\text{rank}(\tilde{C}_r S_{Y_r | Y_d, X} \tilde{C}_r^\dagger)} \right).$$

(93)

In Theorem 11 without combining the relay's observed vector, we had $\text{rank}(S_{Y_r | Y_d, X}) = r''$ and $\text{rank}(S_{Y_r | Y_d}) = r'$. The key step here is to design the combining matrix $\tilde{C}_r$ at the relay such that $\text{rank}(\tilde{C}_r S_{Y_r | Y_d, X} \tilde{C}_r^\dagger) = 0$, while $\text{rank}(\tilde{C}_r S_{Y_r | Y_d} \tilde{C}_r^\dagger) = r' - r''$. This can be obtained by distributed zero-forcing of interference, i.e., $\tilde{C}_r$ is such that for some $A \in C^{r \times d}$,

$$\left[ \begin{array}{c} \tilde{C}_r \\ A \end{array} \right] \left[ \begin{array}{c} H_{tr} \\ H_{td} \end{array} \right] S_X^\dagger \tilde{X}_r = 0,$$

(94)

while

$$\text{rank} \left( \left[ \begin{array}{c} \tilde{C}_r \\ A \end{array} \right] \left[ \begin{array}{c} H_{sr} \\ H_{sd} \end{array} \right] S_X^\dagger \tilde{X}_r \right) = \tilde{r},$$

(95)

almost surely. In other words, $\tilde{C}_r$ is chosen such that the observed interference at the relay is aligned with the row space of the observed interference at the destination, i.e.,

$$\text{rowspan} \left( \tilde{C}_r H_{tr} S_{X_r}^\dagger \right) \subseteq \text{rowspan} \left( H_{td} S_{X_r}^\dagger \right).$$

(96)

Note that since $\left[ H_{tr} \dagger, H_{td} \right] S_X^\dagger \tilde{X}_r$ has an $(r + d - t)^+$ dimensional left null space, such a zero-forcing matrix always exists.

Now, we argue that $\text{rank}(\tilde{C}_r S_{Y_r | Y_d, X} \tilde{C}_r^\dagger) = 0$ and $\text{rank}(\tilde{C}_r S_{Y_r | Y_d} \tilde{C}_r^\dagger) = r' - r''$.

We have

$$S_{\tilde{C}_r \tilde{Y}_r, \tilde{Y}_d} = \left[ \begin{array}{c} \tilde{C}_r H_{sr} \\ H_{sd} \end{array} \right] S_X \left[ \begin{array}{c} H_{sr} \dagger \tilde{C}_r^\dagger \\ H_{sd} \dagger \end{array} \right] S_X \left[ \begin{array}{c} \tilde{C}_r H_{tr} \\ H_{td} \end{array} \right] S_X \left[ \begin{array}{c} H_{tr} \dagger \tilde{C}_r^\dagger \\ H_{td} \dagger \end{array} \right],$$

(97)

and

$$\text{rank}(S_{\tilde{C}_r \tilde{Y}_r, \tilde{Y}_d}) = \min (\tilde{r} + d, \min (s, \tilde{r} + d) + \min (d, t))$$

(98)

almost surely. Similar to the proof of Lemma 3, we use rank additivity of generalized Schur complement

$$\text{rank} \left( \begin{array}{c} \tilde{C}_r S_{Y_r | Y_d} \tilde{C}_r^\dagger \\ S_{Y_r, Y_d} \end{array} \right) = \text{rank} \left( \begin{array}{c} \tilde{C}_r S_{Y_r | Y_d} \tilde{C}_r^\dagger \\ S_{Y_r, Y_d} \end{array} \right) - \text{rank} \left( S_{Y_r, Y_d} \right)$$

$$= \min (\tilde{r} + d, d + s, s + t) - \min (d, s + t)$$

$$= r' - r''.$$  

(99)

Also, we have

$$\text{rank} \left( S_{\tilde{C}_r \tilde{Y}_r, \tilde{Y}_d} \right)$$

$$= \text{rank} \left( \begin{array}{cccc} \tilde{C}_r H_{tr} S_X H_{tr} \dagger \tilde{C}_r^\dagger & \tilde{C}_r H_{tr} S_X H_{td} \dagger & \tilde{C}_r H_{sr} S_X \\ H_{td} S_X \tilde{C}_r \dagger & H_{td} S_X H_{td} \dagger & H_{sd} S_X \\ 0 & 0 & S_X \end{array} \right)$$

$$= s' + \min (d, t)$$

$$= \text{rank} \left( S_{Y_d, X} \right),$$

(100)
almost surely. Similar to the proof of Lemma 3 by rank additivity of generalized Schur complement
\[ \text{rank} \left( \tilde{C}_rS_{Y_i|Y_d,X}C^\dagger_r \right) = \text{rank} \left( S_{\tilde{C}_rY_r,Y_d,X} \right) - \text{rank} \left( S_{Y_d,X} \right) = 0. \] (101)

Therefore, the DoF gain (93) can be written as
\[ \Delta \text{DoF} = (r' - r'') \min \left( \frac{\alpha}{r'} - \frac{\alpha}{r''} \right) = \min \left( \text{DoF}_R - \text{DoF}_D, \alpha \right). \] (102)

The last equality follows by noting \( r' - r'' = \text{DoF}_R - \text{DoF}_D \) from (39) and (46).

**APPENDIX E**

**PROOF OF THEOREM 14**

We first argue that a sub-optimal evaluation of compress-and-forward rate in (8) is to within a constant gap of the cut-set upper bound (5). Then, we show that this is true for the achievable rate expression (6) when optimized by Algorithm 1 as well.

Let \( X \sim \mathcal{CN}(0_{r \times 1}, S^{\text{CSB}}_X) \), where \( S^{\text{CSB}}_X \) is the global maximizer of the cut-set bound (5). Then, consider matrix \( C_r \) that simultaneously diagonalizes \( S_{Y_i|Y_d,X} \) and \( S_{Y_r|Y_d,X} \), and the generalized eigenvalues \( \lambda_i \)'s in decreasing order as introduced in Lemma 1. Let \( Q \sim \mathcal{CN}(0_{r \times 1}, S^{\text{CG}}_Q) \), where \( S^{\text{CG}}_Q \triangleq C_r^{-1} \Sigma^{\text{CG}}_Q C_r^{-1} \) with diagonal \( \Sigma^{\text{CG}}_Q \) such that the \( i \text{th} \) diagonal element is
\[ \Sigma^{ii}_{Q,Q} = \begin{cases} \frac{\lambda_i}{\lambda_i+1} & \lambda_i > 1 \\ +\infty & \lambda_i = 1 \end{cases} \] (103)

Under the above input and quantization distributions, the gap between the cut-set upper bound (5) and the achievable rate (8) is bounded as below
\[
\begin{align*}
&\min \left\{ I(X; Y_r, Y_d), I(X; Y_d) + C_0 \right\} - \\
&\min \left\{ I(X; \breve{Y}_r, Y_d), I(X; Y_d) + C_0 - I(Y_r; \breve{Y}_r|Y_d, X) \right\} \\
&\leq \max \left\{ I(X; Y_r | Y_d) - I(X; \breve{Y}_r | Y_d), I(Y_r; \breve{Y}_r | Y_d, X) \right\} \\
&= \max \left\{ \log \frac{|S_{Y_i|Y_d,X}|}{|S_{Y_i,Y_d,X}|}, \log \frac{|S_Q + S_{Y_i,Y_d,X}|}{|S_Q|} \right\} \\
&\overset{(a)}{=} \max \left\{ \sum_{i=1}^{r} \lambda_i (\Sigma^{ii}_{Q} + 1), \sum_{i=1}^{r} \log \frac{\Sigma^{ii}_{Q} + 1}{\Sigma^{ii}_{Q}} \right\} \\
&\leq \sum_{i=1}^{r} \max \left\{ \lambda_i (\Sigma^{ii}_{Q} + 1), \log \frac{\Sigma^{ii}_{Q} + 1}{\Sigma^{ii}_{Q}} \right\} \\
&\overset{(b)}{=} \sum_{i=1}^{r} \log \left( 2 - \frac{1}{\lambda_i} \right) \overset{(c)}{=} r - (r-s)^+ = \min(r,s). \\
\end{align*}
\]

Here, Lemma 1 is used in equality (a), equality (b) follows by the choice of \( \Sigma_Q \) in (103); note that for \( i \in \{1, \ldots, r\} \), if \( \lambda_i > 1 \), the first and the second arguments of the maximization are, respectively, increasing and decreasing in \( \Sigma^{ii}_{Q} \). Therefore, the corresponding term is minimized by equating the two and solving for \( \Sigma^{ii}_{Q} \). If \( \lambda_i = 1 \), the corresponding term is zero by letting \( \Sigma^{ii}_{Q} \to \infty \). By Theorem 3 the number of such components is at least \( (r-s)^+ \), hence (c).

Now, we argue that the achievable rate by Algorithm 1 is also to within a constant gap of the cut-set bound. Initialize Algorithm 1 with \( S^{\text{CSB}}_X \) and update the optimal quantization noise covariance, denoted by \( S^{\text{CG}}_Q \). By Theorem 3 \( S^{\text{CG}}_Q \) is the global optimum of (8) at \( S_X = S^{\text{CSB}}_X \) as well. Hence, the achievable rate by Algorithm 1, i.e., \( R_{CF} = I(X; \breve{Y}_r, Y_d) \) evaluated at \( (S^{\text{CSB}}_X, S^{\text{CG}}_Q) \), is no smaller than (8) evaluated at \( (S^{\text{CSB}}_X, S^{\text{CG}}_Q) \), which is already shown to be within \( \min(r,s) \) bits of the cut-set upper bound.

**ACKNOWLEDGMENT**

We thank Professor Jun Chen for informing us of [8].
REFERENCES

[1] T. Cover and A. El Gamal, “Capacity theorems for the relay channel,” *IEEE Trans. Inf. Theory*, vol. 25, no. 5, pp. 572–584, Sep. 1979.
[2] G. Kramer, M. Gastpar, and P. Gupta, “Cooperative strategies and capacity theorems for relay networks,” *IEEE Trans. Inf. Theory*, vol. 51, no. 9, pp. 3037 – 3063, Sept. 2005.
[3] L. Zhang, J. Jiang, A. Goldsmith, and S. Cui, “Study of Gaussian relay channels with correlated noises,” *IEEE Trans. Commun.*, vol. 59, no. 3, pp. 863–876, Mar. 2011.
[4] S. Simoens, O. Muez-Medina, J. Vidal, and A. del Coso, “Compress-and-forward cooperative MIMO relaying with full channel state information,” *IEEE Trans. Signal Process.*, vol. 58, no. 2, pp. 781–791, Feb. 2010.
[5] M. Gastpar, P. Dragotti, and M. Vetterli, “The distributed Karhunen-Loève transform,” *IEEE Trans. Inf. Theory*, vol. 52, no. 12, pp. 5177 – 5196, Dec. 2006.
[6] R. A. Horn and C. R. Johnson, *Matrix Analysis*. Cambridge University Press, 1990.
[7] A. Ayoughi and W. Yu, “Optimized MIMO transmission and compression for interference mitigation with user cooperation,” in *Proc. IEEE Int. Conf. Commun. (ICC)*, 2015.
[8] C. Tian and J. Chen, “Remote vector Gaussian source coding with decoder side information under mutual information and distortion constraints,” *IEEE Trans. Inf. Theory*, vol. 55, no. 10, pp. 4676–4680, Oct. 2009.
[9] L. Zhou and W. Yu, “Capacity of the Gaussian relay channel with correlated noises to within a constant gap,” *IEEE Commun. Lett.*, vol. 16, no. 1, pp. 2–5, Nov. 2012.
[10] X. Jin and Y. Kim, “The approximate capacity of the MIMO relay channel,” *IEEE Trans. Inf. Theory*, vol. 63, no. 2, pp. 1167–1176, 2017.
[11] S. Mohajer, S. N. Diggavi, C. Fragouli, and D. N. C. Tse, “Approximate capacity of a class of Gaussian interference-relay networks,” *IEEE Trans. Inf. Theory*, vol. 57, no. 5, pp. 2837–2864, 2011.
[12] T. Gou, S. A. Jafar, C. Wang, S.-W. Jeon, and S.-Y. Chung, “Aligned interference neutralization and the degrees of freedom of the $2 \times 2 \times 2$ interference channel,” *IEEE Trans. Inf. Theory*, vol. 58, no. 7, pp. 4381–4395, 2012.
[13] B. Rankov and A. Wittneben, “Spectral efficient protocols for half-duplex fading relay channels,” *IEEE J. Sel. Areas Commun.*, vol. 25, no. 2, pp. 379–389, 2007.
[14] K. S. Gomadam and S. A. Jafar, “The effect of noise correlation in amplify-and-forward relay networks,” *IEEE Trans. Inf. Theory*, vol. 55, no. 2, pp. 731–745, 2009.
[15] T. M. Cover and Y. Kim, “Capacity of a class of deterministic relay channel,” in *IEEE Int. Symp. Inf. Theory (ISIT)*, Jun. 2007, pp. 591–595.
[16] A. Ayoughi and W. Yu, “A DoF analysis of compress-and-forward in MIMO Gaussian relay channel with correlated noises,” in *Proc. Can. Workshop Inf. Theory (CWIT)*, 2015.
[17] A. Sanderovich, S. Shamai, Y. Steinberg, and G. Kramer, “Communication via decentralized processing,” *IEEE Trans. Inf. Theory*, vol. 54, no. 7, pp. 3008–3023, 2008.
[18] A. El Gamal and Y. Kim, *Network Information Theory*. Cambridge University Press, 2011.
[19] C. R. Rao, *Linear Statistical Inference and Its Applications*, 2nd ed. Wiley, 1973.
[20] A. del Coso and S. Simoens, “Distributed compression for MIMO coordinated networks with a backhaul constraint,” *IEEE Trans. Wireless Commun.*, vol. 8, no. 9, pp. 4698–4709, Sep. 2009.
[21] S. Boyd and L. Vandenberghe, *Convex Optimization*. Cambridge University Press, 2004.
[22] F. Rellich, *Perturbation theory of eigenvalue problems*. CRC Press, 1969.
[23] D. Carlson, E. Haynsworth, and T. Markham, “A generalization of the Schur complement by means of the Moore-Penrose inverse,” *SIAM J. Appl. Math.*, vol. 26, no. 1, pp. 169–175, Jan. 1974.
[24] L.-Z. Lu and C. E. M. Pearce, “Some new bounds for singular values and eigenvalues of matrix products,” *Ann. Operations Res.*, vol. 98, no. 1-4, pp. 141–148, 2000.