Information Bottleneck for an Oblivious Relay with Channel State Information: the Vector Case

Hao Xu*, Tianyu Yang*, Giuseppe Caire*, and Shlomo Shamai (Shitz)†
*Faculty of Electrical Engineering and Computer Science, Technical University of Berlin, 10587 Berlin, Germany
†Department of Electrical Engineering, TechnionIsrael Institute of Technology, Haifa 3200003, Israel
E-mail: xuhao@mail.tu-berlin.de; tianyu.yang@tu-berlin.de; caire@tu-berlin.de; sshlomo@ee.technion.ac.il

2021 IEEE International Symposium on Information Theory (ISIT) | 978-1-5386-8209-8/21/$31.00 ©2021 IEEE | DOI: 10.1109/ISIT45174.2021.9518052

***Abstract*—This paper considers the information bottleneck (IB) problem of a Rayleigh fading multiple-input multiple-out (MIMO) channel. Due to the bottleneck constraint, it is impossible for the oblivious relay to inform the destination node of the perfect channel state information (CSI) in each channel realization. To evaluate the bottleneck rate, we provide an upper bound by assuming that the destination node can get the perfect CSI at no cost and two achievable schemes with simple symbol-by-symbol relay processing and compression. Numerical results show that the lower bounds obtained by the proposed achievable schemes can come close to the upper bound on a wide range of relevant system parameters.**

***Index Terms*—information bottleneck (IB), oblivious relay, Rayleigh fading, source coding, quantization.**

I. Introduction

For a Markov chain $X \rightarrow Y \rightarrow Z$ and an assigned joint probability distribution $p_{X,Y}$, consider the following information bottleneck (IB) problem

$$\max_{p_{Z|Y}} \quad I(X;Z) \tag{1a}$$

$$\text{s.t.} \quad I(Y;Z) \leq C, \tag{1b}$$

where C is the bottleneck constraint parameter and the optimization is with respect to the conditional probability distribution $p_{Z|Y}$ of Z given Y. Formulation (1) was introduced by Tishby in [1], and has been used to interpret the behavior of deep learning neural networks [2]. From a more fundamental information theoretic viewpoint, the IB arises from the classical remote source coding problem [3], [4] under logarithmic distortion [5].

An interesting application of the IB problem in communications consists of a source node, an oblivious relay, and a destination node, which is connected to the relay via an error-free link with capacity C. The source node sends codewords over a communication channel and an observation is made at the relay. X and Y are respectively the channel input from the source node and output at the relay. The relay is oblivious in the sense that it cannot decode the information message of the source node itself. This feature can be modeled rigorously by assuming that the source and destination nodes make use of a codebook selected at random over a library, while the relay is unaware of such random selection. For example, in a cloud radio access network (C-RAN), each remote radio head (RRH) acts as a relay and is usually constrained to implement only radio functionalities while the baseband functionalities are migrated to the cloud central processor, particularly as the network size gets large [6].

Due to the oblivious feature, the relaying strategies which require the codebooks to be known at the relay, e.g., decode-and-forward, compute-and-forward, etc. [7]–[9] cannot be applied. Instead, the relay has to perform oblivious processing, i.e., employ strategies in forms of compress-and-forward [10]–[13]. In particular, the relay must treat X as a random process, produce some useful representation Z, and convey it to the destination node subject to the link constraint C. Then, it makes sense to find Z such that $I(X;Z)$ is maximized.

The IB problem for this kind of communication scenario has been studied in [14]–[17]. In [14], the IB method was applied to reduce the fronthaul data rate of a C-RAN network. References [15] and [16] respectively considered Gaussian scalar and vector channels with IB constraint, and investigated the optimal trade-off between the compression rate and the relevant information. However, all references [14], [15], and [16] considered block fading channels, and assumed that the perfect channel state information (CSI) was known at both the relay and the destination node. In [17], the IB problem of a scalar Rayleigh fading channel was studied. Due to the bottleneck constraint, it is impossible to inform the destination node of the perfect CSI in each channel realization. An upper bound and two achievable schemes were provided in [17].

In this paper, we extend the work in [17] to the multiple-input multiple-out (MIMO) channel with independent and identically distributed (i.i.d.) Rayleigh fading. To evaluate the bottleneck rate, we first obtain an upper bound by assuming that the channel matrix is also known at the destination node with no cost. Then, we provide two achievable schemes where the first scheme transmits the compressed noisy signal as well as the quantized noise levels to the destination node, while the second scheme only transmits a compressed estimate. Numerical results show that with simple symbol-by-symbol relay processing and compression, the lower bounds obtained by the proposed achievable schemes can come close to the upper bound on a wide range of relevant system parameters.

II. Problem Formulation

We consider a system with a source node, an oblivious relay, and a destination node. For convenience, we call the source-relay channel, 'Channel 1', and the relay-destination channel,

'Channel 2'. For Channel 1, we consider the following Gaussian MIMO channel with i.i.d. Rayleigh fading

$$\boldsymbol{y} = \boldsymbol{H}\boldsymbol{x} + \boldsymbol{n}, \tag{2}$$

where $\boldsymbol{x} \in \mathbb{C}^{K\times 1}$ and $\boldsymbol{n} \in \mathbb{C}^{M\times 1}$ are respectively zero-mean circularly symmetric complex Gaussian input and noise with covariance matrices $\boldsymbol{I}_K$ and $\sigma^2\boldsymbol{I}_M$, i.e., $\boldsymbol{x} \sim \mathcal{CN}(\boldsymbol{0}, \boldsymbol{I}_K)$ and $\boldsymbol{n} \sim \mathcal{CN}(\boldsymbol{0}, \sigma^2\boldsymbol{I}_M)$. $\boldsymbol{H} \in \mathbb{C}^{M\times K}$ is a random matrix independent of both $\boldsymbol{x}$ and $\boldsymbol{n}$, and the elements of $\boldsymbol{H}$ are i.i.d. zero-mean unit-variance complex Gaussian random variables, i.e., $\boldsymbol{H} \sim \mathcal{CN}(\boldsymbol{0}, \boldsymbol{I}_K \otimes \boldsymbol{I}_M)$. Let $\rho = \frac{1}{\sigma^2}$ denote the signal-to-noise ratio (SNR). Let $\boldsymbol{z}$ denote a useful representation of $\boldsymbol{y}$ produced by the relay for the destination node. $\boldsymbol{x} \to (\boldsymbol{y}, \boldsymbol{H}) \to \boldsymbol{z}$ thus forms a Markov chain. We assume that the relay node has a direct observation of the channel matrix $\boldsymbol{H}$, while the destination node does not. Then, we consider the following IB problem

$$\max_{p(\boldsymbol{z}|\boldsymbol{y},\boldsymbol{H})} \quad I(\boldsymbol{x};\boldsymbol{z}) \tag{3a}$$

$$\text{s.t.} \quad I(\boldsymbol{y},\boldsymbol{H};\boldsymbol{z}) \leq C, \tag{3b}$$

where C is the bottleneck constraint, i.e., the link capacity of Channel 2. In this paper, we call $I(\boldsymbol{x};\boldsymbol{z})$ the bottleneck rate and $I(\boldsymbol{y},\boldsymbol{H};\boldsymbol{z})$ the compression rate. Obviously, for a joint probability distribution $p(\boldsymbol{x},\boldsymbol{y},\boldsymbol{H})$ determined by (2), problem (3) is a slightly augmented version of IB problem (1). In our problem, we aim to find a conditional distribution $p(\boldsymbol{z}|\boldsymbol{y},\boldsymbol{H})$ such that bottleneck constraint (3b) is satisfied and the bottleneck rate is maximized, i.e., as much as information of $\boldsymbol{x}$ can be extracted from representation $\boldsymbol{z}$.

III. Informed Receiver Upper Bound

As stated in [17], an obvious upper bound to problem (3) can be obtained by letting both the relay and the destination node know the channel matrix $\boldsymbol{H}$. We call the bound in this case the informed receiver upper bound. The IB problem in this case takes on the following form

$$\max_{p(\boldsymbol{z}|\boldsymbol{y},\boldsymbol{H})} \quad I(\boldsymbol{x};\boldsymbol{z}|\boldsymbol{H}) \tag{4a}$$

$$\text{s.t.} \quad I(\boldsymbol{y};\boldsymbol{z}|\boldsymbol{H}) \leq C. \tag{4b}$$

From the definition of $\boldsymbol{H}$ in (2), it is known that when $K \leq M$ (resp., when $K > M$), $\boldsymbol{H}^H\boldsymbol{H}$ (resp., $\boldsymbol{H}\boldsymbol{H}^H$) is a central complex Wishart matrix with M (resp., K) degrees of freedom and covariance matrix $\boldsymbol{I}_K$ (resp., $\boldsymbol{I}_M$), i.e., $\boldsymbol{H}^H\boldsymbol{H} \sim \mathcal{CW}_K(M, \boldsymbol{I}_K)$ (resp., $\boldsymbol{H}\boldsymbol{H}^H \sim \mathcal{CW}_M(K, \boldsymbol{I}_M)$) [18]. Let λ denote the unordered positive eigenvalue of $\boldsymbol{H}^H\boldsymbol{H}$ or $\boldsymbol{H}\boldsymbol{H}^H$. Its probability density function (pdf) is then given by [18, Theorem 2.17], [19]

$$f_\lambda(\lambda) = \frac{1}{T}\sum_{i=0}^{T-1} \frac{i!}{(i+S-T)!}\left[L_i^{S-T}(\lambda)\right]^2 \lambda^{S-T} e^{-\lambda}, \tag{5}$$

where $T = \min\{K, M\}$, $S = \max\{K, M\}$, and the Laguerre polynomials are

$$L_i^{S-T}(\lambda) = \frac{e^\lambda}{i!\lambda^{S-T}} \frac{d^i}{d\lambda^i}\left(e^{-\lambda}\lambda^{S-T+i}\right). \tag{6}$$

In [15], the IB problem for a scalar Gaussian channel with block fading has been studied. In the following theorem, we show that for the considered MIMO channel with Rayleigh fading, (4) can be decomposed into a set of parallel scalar IB problems, and the informed receiver upper bound can be obtained based on the result in [15].

Theorem 1. *For the considered MIMO channel with Rayleigh fading, the informed receiver upper bound is*

$$R^{ub} = T\int_{\frac{\nu}{\rho}}^{\infty} \left[\log(1+\rho\lambda) - \log(1+\nu)\right] f_\lambda(\lambda) d\lambda, \tag{7}$$

where ν is chosen such that the following bottleneck constraint is met

$$\int_{\frac{\nu}{\rho}}^{\infty} \left(\log\frac{\rho\lambda}{\nu}\right) f_\lambda(\lambda) d\lambda = \frac{C}{T}. \tag{8}$$

Lemma 1. *When $M \to +\infty$ or $\rho \to +\infty$, upper bound R^{ub} tends asymptotically to C. When $C \to +\infty$, R^{ub} approaches the capacity of Channel 1, i.e.,*

$$\begin{aligned} R^{ub} &\to I(\boldsymbol{x};\boldsymbol{y},\boldsymbol{H}) \\ &= T\int_0^{\infty} \log(1+\rho\lambda) f_\lambda(\lambda) d\lambda. \end{aligned} \tag{9}$$

IV. Achievable Schemes

In this section, we provide two achievable schemes where each scheme satisfies the bottleneck constraint and gives a lower bound to the bottleneck rate.

A. Quantized channel inversion (QCI) scheme when $K \leq M$

In our first achievable scheme, the relay first gets an estimate of the channel input using channel inversion and then transmits the quantized noise levels as well as the compressed noisy signal to the destination node.

In particular, we apply the pseudo inverse matrix of $\boldsymbol{H}$, i.e., $(\boldsymbol{H}^H\boldsymbol{H})^{-1}\boldsymbol{H}^H$, to $\boldsymbol{y}$, and get the zero-forcing estimate of $\boldsymbol{x}$ as follows

$$\begin{aligned} \tilde{\boldsymbol{x}} &= (\boldsymbol{H}^H\boldsymbol{H})^{-1}\boldsymbol{H}^H\boldsymbol{y} \\ &= \boldsymbol{x} + (\boldsymbol{H}^H\boldsymbol{H})^{-1}\boldsymbol{H}^H\boldsymbol{n} \\ &\triangleq \boldsymbol{x} + \tilde{\boldsymbol{n}}. \end{aligned} \tag{10}$$

For a given channel matrix $\boldsymbol{H}$, $\tilde{\boldsymbol{n}} \sim \mathcal{CN}(\boldsymbol{0}, \boldsymbol{A})$, where $\boldsymbol{A} = \sigma^2(\boldsymbol{H}^H\boldsymbol{H})^{-1}$. Let $\boldsymbol{A} = \boldsymbol{A}_1 + \boldsymbol{A}_2$, where $\boldsymbol{A}_1$ and $\boldsymbol{A}_2$ respectively consist of the diagonal and off-diagonal elements of $\boldsymbol{A}$, i.e., $\boldsymbol{A}_1 = \boldsymbol{A} \odot \boldsymbol{I}_K$ and $\boldsymbol{A}_2 = \boldsymbol{A} - \boldsymbol{A}_1$. If $\boldsymbol{H}$ can be perfectly transmitted to the destination node, the bottleneck rate could be obtained by following similar steps in Appendix A in [20]. However, since $\boldsymbol{H}$ follows a non-degenerate continuous distribution and the bottleneck constraint is finite, this is not possible. To reduce the number of bits per channel use required for informing the destination node of the channel information, we only convey a compressed version of $\boldsymbol{A}_1$ and consider a set of independent scalar Gaussian sub-channels.

Specifically, we force each diagonal entry of $\boldsymbol{A}_1$ to belong to a finite set of quantized levels by adding artificial noise, i.e., by introducing physical degradation. We fix a finite grid of J positive quantization points $\mathcal{B} = \{b_1, \cdots, b_J\}$, where $b_1 \leq b_2 \leq \cdots \leq b_{J-1} < b_J$, $b_J = +\infty$, and define the following ceiling operation

$$\lceil a \rceil_{\mathcal{B}} = \arg\min_{b \in \mathcal{B}} \{a \leq b\}. \tag{11}$$

Then, by adding a Gaussian noise vector $\tilde{\boldsymbol{n}}' \sim \mathcal{CN}(\boldsymbol{0}, \operatorname{diag}\{\lceil a_1 \rceil_{\mathcal{B}} - a_1, \cdots, \lceil a_K \rceil_{\mathcal{B}} - a_K\})$, which is independent of everything else, to (10), a degraded version of $\tilde{\boldsymbol{x}}$ can be obtained as follows

$$\begin{aligned} \hat{\boldsymbol{x}} &= \tilde{\boldsymbol{x}} + \tilde{\boldsymbol{n}}' \\ &= \boldsymbol{x} + \tilde{\boldsymbol{n}} + \tilde{\boldsymbol{n}}' \\ &\triangleq \boldsymbol{x} + \hat{\boldsymbol{n}}, \end{aligned} \tag{12}$$

where $\hat{\boldsymbol{n}} \sim \mathcal{CN}(\boldsymbol{0}, \boldsymbol{A}_1' + \boldsymbol{A}_2)$ for a given $\boldsymbol{H}$ and $\boldsymbol{A}_1' \triangleq \operatorname{diag}\{\lceil a_1 \rceil_{\mathcal{B}}, \cdots, \lceil a_K \rceil_{\mathcal{B}}\}$. Obviously, due to $\boldsymbol{A}_2$, the elements in noise vector $\hat{\boldsymbol{n}}$ are correlated.

To evaluate the bottleneck rate, we consider a new variable

$$\hat{\boldsymbol{x}}_g = \boldsymbol{x} + \hat{\boldsymbol{n}}_g, \tag{13}$$

where $\hat{\boldsymbol{n}}_g \sim \mathcal{CN}(\boldsymbol{0}, \boldsymbol{A}_1')$. Obviously, (13) can be seen as K parallel scalar Gaussian sub-channels with noise power $\lceil a_k \rceil_{\mathcal{B}}$ for each sub-channel. Since each quantized noise level $\lceil a_k \rceil_{\mathcal{B}}$ only has J possible values, it is possible for the relay to inform the destination node of the channel information via the constrained link. Note that from the definition of $\boldsymbol{A}$ in (10), it is known that a_k, $\forall\, k \in \mathcal{K} \triangleq \{1, \cdots, K\}$ are correlated. The quantized noise levels $\lceil a_k \rceil_{\mathcal{B}}$, $\forall\, k \in \mathcal{K}$ are thus also correlated. Hence, we can jointly source-encode $\lceil a_k \rceil_{\mathcal{B}}$, $\forall\, k \in \mathcal{K}$ to further reduce the number of bits used for CSI feedback. However, since the joint entropy of the quantization indices is difficult to obtain (even numerically, since it is a discrete joint distribution over J^K possible values), in this work we consider the (slightly) suboptimal, but far more practical, entropy coding of each sub-channel quantization index separately. The resulting optimization problem becomes

$$\max_{p(\hat{\boldsymbol{z}}_g|\hat{\boldsymbol{x}}_g)} \quad I(\boldsymbol{x}; \hat{\boldsymbol{z}}_g | \boldsymbol{A}_1') \tag{14a}$$

$$\text{s.t.} \quad I(\hat{\boldsymbol{x}}_g; \hat{\boldsymbol{z}}_g | \boldsymbol{A}_1') \leq C - \sum_{k=1}^{K} H_k, \tag{14b}$$

where H_k denotes the entropy of $\lceil a_k \rceil_{\mathcal{B}}$. Since we assume $K \leq M$ in this subsection, as stated in Section III, $\boldsymbol{H}^H\boldsymbol{H} \sim \mathcal{CW}_K(M, \boldsymbol{I}_K)$. The matrix $(\boldsymbol{H}^H\boldsymbol{H})^{-1}$ thus follows a complex inverse Wishart distribution and $a_k, \forall k \in \mathcal{K}$ are marginally identically inverse chi squared distributed with $M - K + 1$ degrees of freedom [21]. Let a denote a new variable with the same distribution as a_k. Its pdf is given by [1]

$$f_a(a) = \frac{(2/\sigma^2)^{-(M-K+1)/2}}{\Gamma\left(\frac{M-K+1}{2}\right)} a^{-(M-K+1)/2-1} e^{-\sigma^2/(2a)}. \tag{15}$$

[1]In Appendix C of the long version of this paper, i.e., [20], we provide more details about the derivation of this pdf.

Then, $H_k = H_0 \triangleq -\sum_{j=1}^{J} P_j \log P_j$, where the probability mass function (pmf) P_j can be calculated as follows

$$\begin{aligned} P_j &= \Pr\{\lceil a \rceil_{\mathcal{B}} = b_j\} \\ &= \Pr\{b_{j-1} < a \leq b_j\} \\ &= \int_{b_{j-1}}^{b_j} f_a(a) da. \end{aligned} \tag{16}$$

In the following theorem, we give a lower bound to the bottleneck rate by solving IB problem (14).

Theorem 2. *If $\boldsymbol{A}_1'$ is conveyed to the destination node for each channel realization, by solving IB problem (14), the following lower bound to the bottleneck rate can be obtained*

$$R^{lb1} = \sum_{j=1}^{J-1} K P_j \left[\log(1 + \rho_j) - \log(1 + \rho_j 2^{-c_j})\right]. \tag{17}$$

where $\rho_j = \frac{1}{b_j}$, $c_j = \left[\log \frac{\rho_j}{\nu}\right]^+$, and ν is chosen such that the following bottleneck constraint is met

$$\sum_{j=1}^{J-1} K P_j c_j = C - K H_0. \tag{18}$$

Since (13) can be seen as K parallel scalar Gaussian sub-channels, according to [15, (16)], the representation of $\hat{\boldsymbol{x}}_g$, i.e., $\hat{\boldsymbol{z}}_g$, can be constructed by adding independent fading and Gaussian noise to each element of $\hat{\boldsymbol{x}}_g$. Denote

$$\begin{aligned} \hat{\boldsymbol{z}}_g &= \boldsymbol{\Phi}\hat{\boldsymbol{x}}_g + \hat{\boldsymbol{n}}_g' \\ &= \boldsymbol{\Phi}\boldsymbol{x} + \boldsymbol{\Phi}\hat{\boldsymbol{n}}_g + \hat{\boldsymbol{n}}_g', \end{aligned} \tag{19}$$

where $\boldsymbol{\Phi}$ is a diagonal matrix with positive and real diagonal entries, and $\hat{\boldsymbol{n}}_g' \sim \mathcal{CN}(\boldsymbol{0}, \boldsymbol{I}_K)$. Note that $\hat{\boldsymbol{x}}_g$ in (13) and its representation $\hat{\boldsymbol{z}}_g$ in (19) are only auxiliary variables. What we are really interested in is the representation of $\hat{\boldsymbol{x}}$ and the corresponding bottleneck rate. Hence, we also add fading $\boldsymbol{\Phi}$ and Gaussian noise $\hat{\boldsymbol{n}}_g'$ to $\hat{\boldsymbol{x}}$ in (12) and get its representation as follows

$$\begin{aligned} \boldsymbol{z} &= \boldsymbol{\Phi}\hat{\boldsymbol{x}} + \hat{\boldsymbol{n}}_g' \\ &= \boldsymbol{\Phi}\boldsymbol{x} + \boldsymbol{\Phi}\hat{\boldsymbol{n}} + \hat{\boldsymbol{n}}_g'. \end{aligned} \tag{20}$$

In the following lemma we show that by transmitting quantized noise levels $\lceil a_k \rceil_{\mathcal{B}}$, $\forall k \in \mathcal{K}$ and representation $\boldsymbol{z}$ to the destination node, R^{lb1} is an achievable lower bound to the bottleneck rate and the bottleneck constraint is satisfied.

Lemma 2. *If $\boldsymbol{A}_1'$ is forwarded to the destination node for each channel realization, with signal vectors $\hat{\boldsymbol{x}}$ and $\hat{\boldsymbol{x}}_g$ in (12) and (13), and their representations $\boldsymbol{z}$ and $\hat{\boldsymbol{z}}_g$ in (20) and (19), we have*

$$I(\hat{\boldsymbol{x}}; \boldsymbol{z} | \boldsymbol{A}_1') \leq I(\hat{\boldsymbol{x}}_g; \hat{\boldsymbol{z}}_g | \boldsymbol{A}_1'), \tag{21}$$

$$I(\boldsymbol{x}; \boldsymbol{z} | \boldsymbol{A}_1') \geq I(\boldsymbol{x}; \hat{\boldsymbol{z}}_g | \boldsymbol{A}_1'), \tag{22}$$

where (21) indicates that $I(\hat{\boldsymbol{x}}; \boldsymbol{z} | \boldsymbol{A}_1') \leq C - KH_0$ and (22) gives $I(\boldsymbol{x}; \boldsymbol{z} | \boldsymbol{A}_1') \geq R^{lb1}$.

Lemma 3. *When $M \to +\infty$ or $\rho \to +\infty$, we can always find a sequence of quantization points $\mathcal{B} = \{b_1, \cdots, b_J\}$ such that $R^{lb1} \to C$. When $C \to +\infty$,*

$$R^{lb1} \to K\mathbb{E}\left[\log\left(1+\frac{1}{a}\right)\right] \leq I(\boldsymbol{x}; \boldsymbol{y}, \boldsymbol{H}), \tag{23}$$

where the expectation can be calculated by using the pdf of a in (15) and $I(\boldsymbol{x}; \boldsymbol{y}, \boldsymbol{H})$ is the capacity of Channel 1.

For the sake of simplicity, we may choose the quantization levels as quantiles such that we obtain the uniform pmf $P_j = \frac{1}{J}$. The lower bound (17) can thus be simplified as

$$R^{\text{lb1}} = \sum_{j=1}^{J-1} \frac{K}{J}\left[\log\left(1+\rho_j\right) - \log(1+\rho_j 2^{-c_j})\right], \tag{24}$$

and the bottleneck constraint (18) becomes

$$\sum_{j=1}^{J-1}\left[\log\frac{\rho_j}{\nu}\right]^+ = \frac{JC}{K} - JB, \tag{25}$$

where $B = \log J$ can be seen as the number of bits required for quantizing each diagonal entry of $\boldsymbol{A}_1$. Since $\rho_1 \geq \cdots \geq \rho_{J-1}$, from the strict convexity of the problem, we know that there must exist a unique integer $1 \leq l \leq J-1$ such that

$$\sum_{j=1}^{l}\log\frac{\rho_j}{\nu} = \frac{JC}{K} - JB,$$
$$\rho_j \leq \nu, \ \forall\, l+1 \leq j \leq J-1. \tag{26}$$

Hence, ν can be obtained from

$$\log\nu = \sum_{j=1}^{l}\frac{\log\rho_j}{l} - \frac{JC}{lK} + \frac{JB}{l}, \tag{27}$$

and R^{lb1} can be calculated as follows

$$R^{\text{lb1}} = \sum_{j=1}^{l}\frac{K}{J}\left[\log\left(1+\rho_j\right) - \log(1+\nu)\right]. \tag{28}$$

Then, we only need to test the above condition for $l = 1, 2, 3, \cdots$ till (26) is satisfied. Note that to ensure $R^{\text{lb1}} > 0$, $\frac{JC}{K} - JB$ in (25) has to be positive, i.e., $B < \frac{C}{K}$. Moreover, though choosing the quantization levels as quantiles makes it easier to calculate R^{lb1}, the results in Lemma 3 may not hold in this case since the choice of quantization points $\mathcal{B} = \{b_1, \cdots, b_J\}$ is restricted.

B. MMSE estimate at the relay

In the second achievable scheme, we assume that the relay first produces the MMSE estimate of $\boldsymbol{x}$ given $(\boldsymbol{y}, \boldsymbol{H})$, and then source-encode this estimate.

Denote

$$\boldsymbol{F} = \left(\boldsymbol{H}\boldsymbol{H}^H + \sigma^2\boldsymbol{I}_M\right)^{-1}\boldsymbol{H}. \tag{29}$$

The MMSE estimate of $\boldsymbol{x}$ is thus given by

$$\bar{\boldsymbol{x}} = \boldsymbol{F}^H\boldsymbol{y} = \boldsymbol{F}^H\boldsymbol{H}\boldsymbol{x} + \boldsymbol{F}^H\boldsymbol{n}. \tag{30}$$

Then, we consider the following modified IB problem

$$\max_{p(\boldsymbol{z}|\bar{\boldsymbol{x}})} \quad I(\boldsymbol{x}; \boldsymbol{z}) \tag{31a}$$
$$\text{s.t.} \quad I(\bar{\boldsymbol{x}}; \boldsymbol{z}) \leq C. \tag{31b}$$

Note that since matrix $\boldsymbol{H}\boldsymbol{H}^H + \sigma^2\boldsymbol{I}_K$ in (29) is always invertible, the results obtained in this subsection always hold no matter $K \leq M$ or $K > M$.

To evaluate the bottleneck rate $I(\boldsymbol{x}; \boldsymbol{z})$, we define an auxiliary Gaussian vector $\bar{\boldsymbol{x}}_g \sim \mathcal{CN}\left(\boldsymbol{0}, \mathbb{E}\left[\bar{\boldsymbol{x}}\bar{\boldsymbol{x}}^H\right]\right)$, let $\bar{\boldsymbol{z}}_g$ denote its representation, and choose $p(\boldsymbol{z}|\bar{\boldsymbol{x}})$ as well as $p(\bar{\boldsymbol{z}}_g|\bar{\boldsymbol{x}}_g)$ to be conditionally Gaussian distribution, i.e.,

$$\boldsymbol{z} = \bar{\boldsymbol{x}} + \boldsymbol{q},$$
$$\bar{\boldsymbol{z}}_g = \bar{\boldsymbol{x}}_g + \boldsymbol{q}, \tag{32}$$

where $\boldsymbol{q} \sim \mathcal{CN}(\boldsymbol{0}, D\boldsymbol{I}_K)$ is independent of everything else. Let

$$I(\bar{\boldsymbol{x}}_g; \bar{\boldsymbol{z}}_g) = \log\det\left(\boldsymbol{I}_K + \frac{\mathbb{E}\left[\bar{\boldsymbol{x}}\bar{\boldsymbol{x}}^H\right]}{D}\right) = C. \tag{33}$$

Then, rate $I(\bar{\boldsymbol{x}}_g; \bar{\boldsymbol{z}}_g)$ is achievable and D can be calculated from (33). Since $I(\bar{\boldsymbol{x}}; \boldsymbol{z}) \leq I(\bar{\boldsymbol{x}}_g; \bar{\boldsymbol{z}}_g)$, $I(\bar{\boldsymbol{x}}; \boldsymbol{z})$ is thus also achievable.

In the following, we obtain a lower bound to $I(\boldsymbol{x}; \boldsymbol{z})$ by evaluating $h(\boldsymbol{z}|\boldsymbol{H})$ and $h(\boldsymbol{z}|\boldsymbol{x})$ separately, and then using

$$I(\boldsymbol{x}; \boldsymbol{z}) = h(\boldsymbol{z}) - h(\boldsymbol{z}|\boldsymbol{x}) \geq h(\boldsymbol{z}|\boldsymbol{H}) - h(\boldsymbol{z}|\boldsymbol{x}). \tag{34}$$

First, since $\boldsymbol{z}$ is conditionally Gaussian given $\boldsymbol{H}$, we have

$$h(\boldsymbol{z}|\boldsymbol{H}) = \mathbb{E}\left[\log(\pi e)^K\det\left(\boldsymbol{F}^H\boldsymbol{H}\boldsymbol{H}^H\boldsymbol{F} + \sigma^2\boldsymbol{F}^H\boldsymbol{F} + D\boldsymbol{I}_K\right)\right]. \tag{35}$$

Next, using the fact that conditioning reduces entropy and Gaussian distribution maximizes the entropy over all distributions with the same variance [22, Theorem 8.6.5], we have

$$\begin{aligned} h(\boldsymbol{z}|\boldsymbol{x}) &= h\left(\boldsymbol{z} - \mathbb{E}(\boldsymbol{z}|\boldsymbol{x})|\boldsymbol{x}\right) \\ &= h\left(\left(\boldsymbol{F}^H\boldsymbol{H} - \mathbb{E}\left[\boldsymbol{F}^H\boldsymbol{H}\right]\right)\boldsymbol{x} + \boldsymbol{F}^H\boldsymbol{n} + \boldsymbol{q}|\boldsymbol{x}\right) \\ &\leq h\left(\left(\boldsymbol{F}^H\boldsymbol{H} - \mathbb{E}\left[\boldsymbol{F}^H\boldsymbol{H}\right]\right)\boldsymbol{x} + \boldsymbol{F}^H\boldsymbol{n} + \boldsymbol{q}\right) \\ &\leq \log(\pi e)^K\det(\boldsymbol{G}), \end{aligned} \tag{36}$$

where

$$\begin{aligned} \boldsymbol{G} &= \mathbb{E}\left[\left(\boldsymbol{F}^H\boldsymbol{H} - \mathbb{E}\left[\boldsymbol{F}^H\boldsymbol{H}\right]\right)\left(\boldsymbol{H}^H\boldsymbol{F} - \mathbb{E}\left[\boldsymbol{H}^H\boldsymbol{F}\right]\right) + \sigma^2\boldsymbol{F}^H\boldsymbol{F}\right] + D\boldsymbol{I}_K \\ &= \mathbb{E}\left[\boldsymbol{F}^H\boldsymbol{H}\boldsymbol{H}^H\boldsymbol{F}\right] - \mathbb{E}\left[\boldsymbol{F}^H\boldsymbol{H}\right]\mathbb{E}\left[\boldsymbol{H}^H\boldsymbol{F}\right] + \sigma^2\mathbb{E}\left[\boldsymbol{F}^H\boldsymbol{F}\right] + D\boldsymbol{I}_K. \end{aligned} \tag{37}$$

Combining (34), (35), and (36), we can get a lower bound to $I(\boldsymbol{x}; \boldsymbol{z})$ as shown in the following theorem.

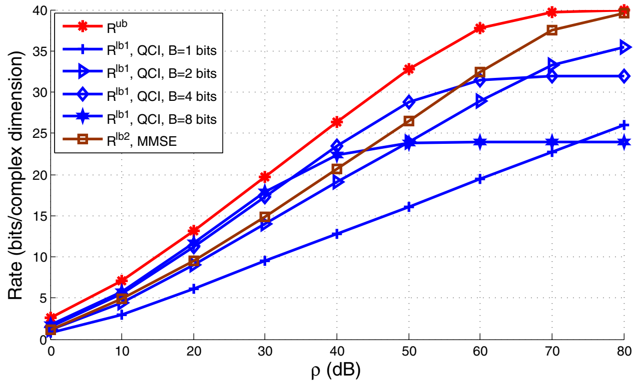

Fig. 1. Upper and lower bounds to the bottleneck rate versus ρ with $K = M = 2$ and $C = 40$ bits/complex dimension.

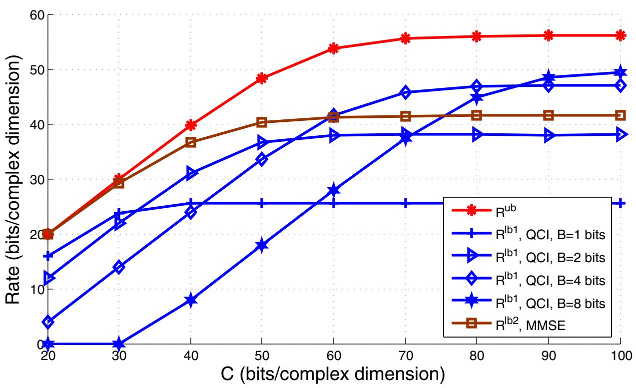

Fig. 2. Upper and lower bounds to the bottleneck rate versus C with $K = M = 4$ and $\rho = 40$dB.

Theorem 3. *With MMSE estimate at the relay, a lower bound to $I(\boldsymbol{x};\boldsymbol{z})$ can be obtained as follows*

$$R^{lb2} = T\mathbb{E}\left[\log\left(\vartheta + D\right)\right] + (K-T)\log D - K\log\left\{\frac{T}{K}\mathbb{E}\left[\vartheta\right] - \frac{T^2}{K^2}\left(\mathbb{E}\left[\vartheta\right]\right)^2 + D\right\}, \tag{38}$$

where $\vartheta = \frac{\lambda}{\lambda+\sigma^2}$, λ is defined in Section III,

$$D = \frac{\frac{T}{K}\mathbb{E}\left[\vartheta\right]}{2^{\frac{C}{K}} - 1}, \tag{39}$$

and the expectations can be calculated by using pdf (5).

Lemma 4. *When $M \to +\infty$ or when $K \le M$ and $\rho \to +\infty$, lower bound R^{lb2} tends asymptotically to C. When $K \le M$ and $C \to +\infty$,*

$$R^{lb2} \to K\mathbb{E}\left[\log\left(\vartheta\right)\right] - K\log\left\{\mathbb{E}\left[\vartheta\right] - \left(\mathbb{E}\left[\vartheta\right]\right)^2\right\}. \tag{40}$$

V. Numerical Results

In this section, we investigate the lower bounds obtained by the proposed achievable schemes and compare them with the upper bound. When performing the QCI scheme, we choose the quantization levels as quantiles for the sake of convenience.

In Fig. 1, the upper and lower bounds are depicted versus SNR ρ. It can be found that when ρ is small and 4 or 8 bits are applied to quantize the noise levels, the QCI scheme outperforms the MMSE scheme. As ρ grows large, R^{lb2} obtained by the MMSE scheme approaches C and is larger than R^{lb1}. This is because when ρ is small, the bottleneck rate is mainly limited by the capacity of channel 1, and the QCI scheme works well in this case since partial CSI, i.e., the noise level of each sub-channel, is conveyed to the destination node. When ρ is large, the MMSE scheme can get an accurate estimate and it does not require CSI feedback. The MMSE scheme thus performs better when ρ is large.

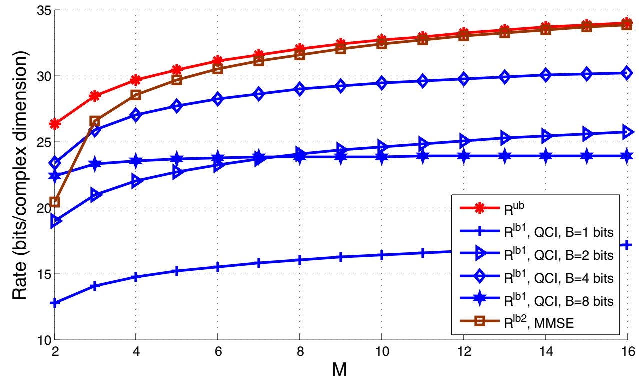

Fig. 3. Upper and lower bounds to the bottleneck rate versus M with $K = 2$, $\rho = 40$dB, and $C = 40$ bits/complex dimension.

The effect of the bottleneck constraint C is investigated in Fig. 2. It can be found that as C increases, all bounds grow and converge to different constants, which can be calculated based on Lemma 1, Lemma 3, and Lemma 4, respectively. Fig. 2 also shows R^{lb2} virtually achieves the upper bound when C is small, while when C is large, the QCI scheme outperforms the MMSE scheme thanks to CSI feedback.

Fig. 3 depicts the bounds versus the number of relay antennas M. As M increases, R^{lb2} quickly approaches R^{ub}. It is also shown that the result for the limit case in Lemma 3, i.e., when $M \to +\infty$, we can always find suitable quantization points $\mathcal{B} = \{b_1, \cdots, b_J\}$ such that $R^{\text{lb1}} \to C$, does not hold here. This is because when performing the QCI scheme, we choose the quantization levels as quantiles. The choice of quantization points $\mathcal{B} = \{b_1, \cdots, b_J\}$ is thus restricted.

VI. Conclusions

This work extends the IB problem of the scalar case in [17] to the case of MIMO Rayleigh fading channels. Due to the information bottleneck constraint, the destination node cannot get the perfect CSI from the relay. Our results show that with simple symbol-by-symbol oblivious relay processing and compression, we can get bottleneck rate close to the upper bound on a wide range of relevant system parameters.

Acknowledgments

This work was supported by the Alexander von Humboldt Foundation and the work of S. Shamai has been supported by the European Union's Horizon 2020 Research and Innovation Programme with grant agreement No. 694630.

References

[1] N. Tishby, F. C. Pereira, and W. Bialek, "The information bottleneck method," *arXiv preprint physics/0004057*, 2000.

[2] R. Shwartz-Ziv and N. Tishby, "Opening the black box of deep neural networks via information," *arXiv preprint arXiv:1703.00810*, 2017.

[3] R. Dobrushin and B. Tsybakov, "Information transmission with additional noise," *IRE Trans. Inf. Theory*, vol. 8, no. 5, pp. 293–304, Sep. 1962.

[4] H. Witsenhausen, "Indirect rate distortion problems," *IEEE Trans. Inf. Theory*, vol. 26, no. 5, pp. 518–521, Sep. 1980.

[5] T. A. Courtade and T. Weissman, "Multiterminal source coding under logarithmic loss," *IEEE Trans. Inf. Theory*, vol. 60, no. 1, pp. 740–761, Jan. 2014.

[6] I. E. Aguerri, A. Zaidi, G. Caire, and S. S. Shitz, "On the capacity of cloud radio access networks with oblivious relaying," *IEEE Trans. Inf. Theory*, vol. 65, no. 7, pp. 4575–4596, July 2019.

[7] B. Nazer and M. Gastpar, "Compute-and-forward: Harnessing interference through structured codes," *IEEE Trans. Inf. Theory*, vol. 57, no. 10, pp. 6463–6486, Oct. 2011.

[8] S.-N. Hong and G. Caire, "Compute-and-forward strategies for cooperative distributed antenna systems," *IEEE Trans. Inf. Theory*, vol. 59, no. 9, pp. 5227–5243, Sep. 2013.

[9] B. Nazer, A. Sanderovich, M. Gastpar, and S. Shamai, "Structured superposition for backhaul constrained cellular uplink," in *Proc. IEEE Int. Symp. Inf. Theory (ISIT)*, Seoul, South Korea, June 2009, pp. 1530–1534.

[10] O. Simeone, E. Erkip, and S. Shamai, "On codebook information for interference relay channels with out-of-band relaying," *IEEE Trans. Inf. Theory*, vol. 57, no. 5, pp. 2880–2888, May 2011.

[11] S.-H. Park, O. Simeone, O. Sahin, and S. Shamai, "Robust and efficient distributed compression for cloud radio access networks," *IEEE Trans. Veh. Technol.*, vol. 62, no. 2, pp. 692–703, Feb. 2013.

[12] Y. Zhou, Y. Xu, W. Yu, and J. Chen, "On the optimal fronthaul compression and decoding strategies for uplink cloud radio access networks," *IEEE Trans. Inf. Theory*, vol. 62, no. 12, pp. 7402–7418, Dec. 2016.

[13] I. E. Aguerri and A. Zaidi, "Lossy compression for compute-and-forward in limited backhaul uplink multicell processing," *IEEE Trans. Commun.*, vol. 64, no. 12, pp. 5227–5238, Dec. 2016.

[14] J. Demel, T. Monsees, C. Bockelmann, D. Wuebben, and A. Dekorsy, "Cloud-ran fronthaul rate reduction via ibm-based quantization for multicarrier systems," in *Proc. 24th International ITG Workshop on Smart Antennas*, Hamburg, Germany, Feb. 2020, pp. 1–6.

[15] A. Winkelbauer and G. Matz, "Rate-information-optimal Gaussian channel output compression," in *Proc. 48th Annu. Conf. Inf. Sci. Syst. (CISS)*, Princeton, NJ, USA, Mar. 2014, pp. 1–5.

[16] A. Winkelbauer, S. Farthofer, and G. Matz, "The rate-information tradeoff for Gaussian vector channels," in *Proc. IEEE Int. Symp. Inf. Theory*, Honolulu, USA, June 2014, pp. 2849–2853.

[17] G. Caire, S. Shamai, A. Tulino, S. Verdu, and C. Yapar, "Information bottleneck for an oblivious relay with channel state information: the scalar case," in *Proc. IEEE Int. Conf. Science of Electrical Engineering in Israel (ICSEE)*, Eilat, Israel, Dec. 2018, pp. 1–5.

[18] A. M. Tulino, S. Verdú *et al.*, *Random matrix theory and wireless communications*. Now Publishers, 2004.

[19] E. Telatar, "Capacity of multi-antenna gaussian channels," *Europ. Trans. Telecommun.*, vol. 10, no. 6, pp. 585–595, Nov.-Dec. 1999.

[20] H. Xu, T. Yang, G. Caire, and S. S. Shitz, "Information bottleneck for an oblivious relay with channel state information: the vector case," https://arxiv.org/abs/2101.09790, Jan. 2021.

[21] L. E. Brennan and I. S. Reed, "An adaptive array signal processing algorithm for communications," *IEEE Trans. Aerosp. Electron. Syst.*, no. 1, pp. 124–130, Jan. 1982.

[22] T. M. Cover and J. A. Thomas, *Elements of information theory*. John Wiley & Sons, 2012.