CCIT Report #668 November 2007

Distributed MIMO receiver - Achievable rates and upper bounds

Amichai Sanderovich,Shlomo Shamai (Shitz),Yossef Steinberg

Technion, Haifa, Israel

Abstract

In this paper we investigate the achievable rate of a system that includes a nomadic transmitter with several antennas, which is received by multiple agents, exhibiting independent channel gains and additive circular-symmetric complex Gaussian noise. In the nomadic regime, we assume that the agents do not have any decoding ability. These agents process their channel observations and forward them to the final destination through lossless links with a fixed capacity. We propose new achievable rates based on elementary compression and also on a Wyner-Ziv (CEO-like) processing, for both fast fading and block fading channels, as well as for general discrete channels. The simpler two agents scheme is solved, up to an implicit equation with a single variable. Limiting the nomadic transmitter to a circular-symmetric complex Gaussian signalling, new upper bounds are derived for both fast and block fading, based on the vector version of the entropy power inequality. These bounds are then compared to the achievable rates in several extreme scenarios. The asymptotic setting with numbers of agents and transmitter's antennas taken to infinity is analyzed. In addition, the upper bounds are analytically shown to be tight in several examples, while numerical calculations reveal a rather small gap in a finite 2×2 setting. The advantage of the Wyner-Ziv approach over elementary compression is shown where only the former can achieve the full diversity-multiplexing tradeoff. We also consider the non-nomadic setting, with agents that can decode. Here we give an achievable rate, over fast fading channel, which combines broadcast with dirty paper coding and the decentralized reception, which was introduced for the nomadic setting.

Index Terms

MIMO, Decentralized detection, wireless networks, Wyner-Ziv, CEO, compress-and-forward

I. Introduction

In this paper we deal with a network in which a nomadic transmitter has several antennas and is communicating to a remote destination, where no direct link exists between the transmitter and the final destination, as is depicted in figure 1. The final destination receives all of its inputs from several separated agents, which are connected to it through fixed lossless links with a given capacity. This setting is identical to the setting of [1], only that here we focus on fading channels. Namely, the channel between the transmitting antennas and the agents is a Rayleigh fading channel with independent channel gains, where the extension to other fading statistics is straight forward. In this contribution we consider both fast fading and block fading channels. The channel fading coefficients, or channel state information (CSI) are known in full to the agents and the final destination, but not to the transmitter. This setting is closely related to the setting of the multiple input multiple output (MIMO) channel, which is thoroughly treated in the literature, see [2]. The multiplexing gain is a common asymptotic measure of performance of MIMO systems. It assesses the capacity increase, for high signal to noise ratios, due to the use of multiple antennas [3] in the scheme. In this paper, we analyze the multiplexing gain for the suggested network, where recent examples for the multiplexing gains of multi terminal networks are [4] and [5]. The results reported here have implications on other MIMO-related channels, such as the MIMO broadcast channel [6], the MIMO relay channel [7], and ad-hoc networks [8]. All these works deal with situations where multiple antennas are transmitting and the signals are received in a distributed fashion, either by relays, destinations or any combination of the above. In addition, results regarding ad-hoc networks [9], relay channels [10] and joint cell-site processing [11] are closely related, providing yet another aspect of the achievable rates in wireless networks, where relays form, in a distributed manner, the required spacial dimensions.

Our model assumes that the transmitter is nomadic, which means that the agents do not possess the codebook in use, and thus do not have any decoding ability [1]. A good way to model a nomadic setting is by letting the transmitter use random encoding. Such model excludes any decoding from the agents. Given that the codebook is random, we further assume that it is Gaussian. In this case, as the model becomes close to source coding and the Gaussian CEO (Chief Executive Officer) [12], we were able to obtain analytic expressions for an achievable rate and for upper bounds. Relevant works here are distributed source coding by Wyner and Ziv (WZ) [13], [14] who deals with the multiple terminals WZ problem and the Gaussian CEO by [15], among others.

The achievable rates derived in this paper extend the achievable rates from [1] to the case of fading channels, and multiple antennas at the transmitter and at the receiver. The techniques that are used for the derivation are based on the well known CEO or WZ distributed source coding. These techniques, although intended for source coding problems, enable better utilization of system resources also for channel coding problems, as done for example by [16],[17] and [18].

The upper bounds in this paper were derived using the vector version of the entropy power inequality, which was used for several known problems which are based on Gaussian statistics. These include the MIMO broadcast channel [6] and the Gaussian CEO with quadratic distortion [15]. Several generalizations to the original entropy power inequality exist, among them are [19],[20] and [21].

The Gaussian signaling used by the transmitter results with the channel outputs being Gaussian and for the nomadic setting, also memoryless. Notice that unlike traditional source coding problems that use the CEO or WZ techniques, and examine the resulting distortions, we focus on the allowed communication rates. Thus any upper bounds or even optimality shown for a source coding problem, although strongly connected, is not identical to our problem. Therefore the technique used to show optimality of the distributed WZ with two terminals problem ([14]) does not carry over to our setting.

This paper is organized as follows, in section II the setting is described and the basic definitions and notations are given. Section III describes the elementary compression approach and gives several results about the achievable rates when using this approach. Section IV improves upon the approach taken in section III by including CEO compression (as in the CEO problem) at the agents and the final destination. An upper bound to the achievable rate, when using nomadic transmitter and non-decoding agents, is given in section V, and then demonstrated by a numerical example, to be rather close to the achievable rate when using the CEO compression. In the last section, an achievable rate for when the agents are informed of the codes used by the transmitter, and the transmitter is informed of both agents' processing and channel coefficients is given in section VII. Concluding remarks are then made in section VIII.

II. Setting and Model definition

Throughout this paper, boldfaced letters are used to denote vectors $\boldsymbol{X}$ of length n, calligraphic letters $\mathcal{T}$ to denote sets, capital letters X are usually used for random variables, and lower case letters for realizations of random variables x, indices i, j, k, and counters n, r, t. Subscript denotes an element within a vector and superscript X^r denotes the set $X_1, \ldots, X_r$.

The statistical mean is denoted by E, * denotes the transpose conjugate and $\mathcal{CN}(\Xi, \Sigma)$ stands for complex Gaussian random variable with mean Ξ and covariance Σ.

An example for the model appears in Figure 1. The model consists of a transmitter S which has t transmitting antennas and which transmits during n channel uses. In each channel use, the transmitter sends a vector $X \in \mathbb{C}^{[t\times 1]}$ to the channel, where $\frac{1}{n}\sum_{k=1}^{n} \mathrm{E}[X(k)^* X(k)] \leq P$. The transmitter uses circular-symmetric complex Gaussian signalling, which is known to be optimal for various problems involving the Gaussian channel. The communication rate is denoted by R. The message to be sent M is encoded by a random encoding function $\boldsymbol{X} = \phi_{S,F}(M)$ such that for all messages M, the outputs of the encoding function are randomly and independently chosen according to probability $P_{\boldsymbol{X}}(\boldsymbol{x})$.We indicate the random encoding function by a random variable F. That is,

$$\phi_{S,F} : [1, \ldots, 2^{nR}] \to \mathcal{X}^n. \tag{1}$$

The agents are not informed about the selected encoding F, but are fully aware of P_X.

We have r agents $A_1, \ldots, A_r$, each receiving the scalar channel outputs

$$Y_i(k) = H_i(k)X(k) + N_i(k), \quad i = 1, \ldots, r,\ k = 1, \ldots, n \tag{2}$$

where $H_i(k) \in \mathbb{C}^{[1\times t]}$ is the vector of the channel transfer coefficients, which are either ergodic (fast fading) or static, non-ergodic (block fading). In both cases, the coefficients are distributed independently from each other, and from any other variable, according to circular-symmetric complex Gaussian distribution $\mathcal{CN}(0, 1)$. Similarly, the noises are distributed as $N_i(k) \sim \mathcal{CN}(0, 1)$, and are independent of each other and along time. For the sake of brevity, we drop the time index k from now on.

Most of the results which are reported here can be easily extended by including other fading distributions, such as Ricean, invoking the results of [22].

The r agents are connected to a remote destination D with lossless links, each with capacity C_i bits per channel use. The final destination D decodes the message M from the r messages, which are sent from the r agents, where decoding function is $\phi_{D,F} : [1, \ldots, 2^{\sum^r nC_i}] \to [1, \ldots, 2^{nR}]$. This setting is depicted in figure 1.

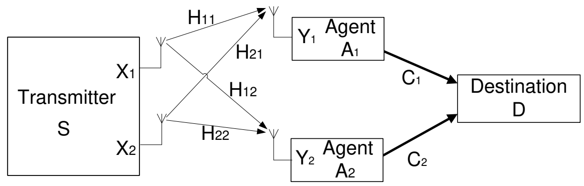

Fig. 1. A system that includes a transmitter with $t = 2$ and two agents A_1 and A_2 ($r = 2$), connected to the final destination with capacities of C_1 and C_2, respectively. The channel fading coefficients H are designated by $\{H_{i,j}\}$.

For fast fading channels, the rate R is said to be achievable, if for every $\epsilon > 0$, there exists n sufficiently large such that

$$\frac{1}{2^{nR}} \sum_{m=1}^{2^{nR}} \Pr(\hat{M} \neq m | M = m) \leq \epsilon, \tag{3}$$

where $\Pr(\hat{M} \neq m | M = m)$ includes averaging over the channel and the random coding. In parallel, the rate-vs-outage probability of ϵ, for block fading is said to be achievable if there exists n sufficiently large such that

$$\frac{1}{2^{nR}} \sum_{m=1}^{2^{nR}} \Pr(\hat{M} \neq m | M = m) \leq \epsilon. \tag{4}$$

The transmitter is nomadic [1], that is the codebook that is used f, is unknown to the agents, but is fully known to the final destination. This way the agents treat input signals not accounting for the coded transmission, in a CEO or multiple WZ approach. All the reported results in this paper assume that the transmitter is limited to using only a circular-symmetric complex Gaussian codebook. Notice that the Gaussian codebook is not necessarily optimal, (a counter example exists for the non fading case, where using binary signaling at the transmitter with a simple two level demapper at the agents can outperform the Gaussian signaling scheme, see [1]). However, the Gaussian codebook does provide a good candidate, as for $C_i \to \infty$ and $C \to 0$ the Gaussian codebook is indeed optimal.

In addition to the nomadism, the transmitter has no information regarding $\boldsymbol{H} = \{H(k)\}_{k=1}^{n}$, where

$$H(k) = \begin{bmatrix} H_1(k) \\ \vdots \\ H_r(k) \end{bmatrix},$$

while the final destination is fully informed about $\boldsymbol{H}$. By default, each agent has the full CSI $\boldsymbol{H}$. However, many of the presented schemes require each agent to know only its own channel coefficients $\boldsymbol{H}_i$, as is stated in the text.

Although the transmitter is unaware of the channel realizations $\boldsymbol{H}$, it does have the full knowledge of the channel statistics, as well as $\{C_i\}$, which is used to calculate the rate in which the transmitter will encode its messages. Alternatively, higher layer control layers can indicate the code-rate which is to be used, based on an ACK/NACK mechanism.

As said, the multiplexing gain of any scheme describes the scaling laws of its capacity, as P is increased [3].

Definition: The multiplexing gain of a scheme is defined as

$$m = \lim_{P\to\infty} \frac{R}{\log(P)}, \tag{5}$$

whereas the diversity is defined by

$$d = \lim_{P\to\infty} -\frac{\log\left(\Pr\{\text{outage}\}\right)}{\log(P)}. \tag{6}$$

We will use $\doteq$ and $\dot{\leq}$ to denote equality and respectively inequality, under the operation $\lim_{P\to\infty} \frac{\log_2(\cdot)}{\log_2(P)}$. The norm $|V|^2$ for a vector V is defined as $|V|^2 = \sum_i |V_i|^2$.

III. ELEMENTARY COMPRESSION SCHEME

In this section, a scheme that incorporates elementary compression at the agents is analyzed. By elementary compression, we mean compression process that does not use the correlations between $\{\boldsymbol{Y}_i\}$, and thus does not require the agents to have full CSI, rather, they just need $\boldsymbol{H}_i$. In addition, the implementation of such compression and especially the decompression are rather simple and realized with low complexity algorithms at the agents and the final destination.

A. Ergodic Channel

We first propose an achievable rate for general ergodic channels.

Proposition 1: An achievable rate for an ergodic channel, with elementary compression is

$$R_{EC} = I(U^r; X|H), \tag{7}$$

(EC stands for elementary-compression) with the constraints:

$$I(U_i; Y_i|H) \leq C_i,\ \ i = 1, \dots, r, \tag{8}$$

where

$$P_{X,Y^r,U^r,H}(x, y^r, u^r, h) = P_X(x)P_H(h)\prod_{i=1}^{r} P_{Y_i|X,H}(y_i|x,h)P_{U_i|Y_i,H}(u_i|y_i,h). \tag{9}$$

The proof involves the random generation of codewords, $\boldsymbol{U}_i$ according to $\prod_{k=1}^{n} P_{U_i|H}(U_i(k)|H(k))$, as done in standard rate-distortion problems with non-casual side information ($\boldsymbol{H}$). These codebooks are made available to both all encoders and decoder. The proof appears in Appendix II.

Applying Proposition 1 to the Gaussian MIMO channel, one gets

Proposition 2: An achievable rate for ergodic setting using elementary compression is equal to:

$$R_{EC} = \max_{Q\in\mathcal{P}, \{q_i:\mathbb{C}^{[r\times t]}\to\mathbb{R}_+\}_{i=1}^{r}} \mathrm{E}_H\left[\log_2\det\left(I_r + \mathrm{diag}\left(1-2^{-q_i(H)}\right)_{i=1}^{r} HQH^*\right)\right], \tag{10}$$

when the maximization in (10) is such that

$$\mathrm{E}_H\left[\log_2\left(\left(2^{q_i(H)}-1\right)\left(H_iQH_i^*+1\right)+1\right)\right] \leq C_i,\ \ i = 1, \dots, r, \tag{11}$$

where

$$\mathcal{P} = \{Q : Q_{i,j} = 0 \text{ for } i \neq j,\ Q_{i,i} \geq 0,\ \mathrm{trace}(Q) \leq P\}. \tag{12}$$

Here each agent employs the elementary compression scheme which is based on an underlying additive circular-symmetric complex Gaussian noise channel $U_i = Y_i + D_i$, where D_i is the compression noise. As in the Gaussian CEO problem, there is a difference between the used formulation and the backward channel $Y_i = U_i + D_i$, used for standard rate-distortion compression.

Another issue here is that the known fading affects the variance of the compression noise. The quantization noise is circularly-symmetric complex Gaussian with variance $P_{D_i}(k)$ that depends on $\boldsymbol{H}(k)$. Let us further define $q_i(h) \triangleq I(Y_i; U_i|X, H = h)$, which, due to the Gaussian model, is equivalent to:

$$\frac{1}{1 + P_{D_i}} = 1 - 2^{-q_i}. \tag{13}$$

Notice that for all $i = 1, \ldots, r$, q_i is a function of $H = \{H_i\}_{i=1}^r$ and thus is a random variable.

It is easy to verify that the optimization problem in Proposition 2 includes a concave objective function (10) but a non-convex domain (11). Notice that since HQH^* is distributed the same as $HVQV^*H^*$ for any unitary V, we can still limit the search for optimal Q to non-ordered elements of a diagonal Q (which is the set $\mathcal{P}$).

Remark 1: Despite the name elementary compression, it requires an infinite number of codebooks at the agents and the final destination, since they should correspond to infinitely many fading coefficients.

1) An Achievable Rate when $r, t \to \infty$: Let us consider the case where $r = \tau t$, symmetric agents with constant total capacity from the agents to the final destination ($C_i = C_t/r$, $i = 1, \ldots, r$).Such scheme can account for bottleneck effects in the channel between the agents and the final destination. Let us take $r \to \infty$, and find the limiting rate which is reliably supported by the scheme ($\tilde{\tau} \triangleq \frac{\min\{r,t\}}{r}$). First consider that when $t \to \infty$, we have $\frac{|H_{i_*}|^2}{t} \to 1$, almost surely. Applying this to (13) and (11), and also setting P_D to be identical for all the agents (the maximal P_D, unlike what was done in (25), which set P_D to be the minimal), we get

$$\lim_{r\to\infty} R_{EC} \leq \tilde{\tau} \lim_{r\to\infty} r\mathrm{E}_H \left[\log_2\left(1 + \frac{P\lambda/t}{1 + P_{D_*}}\right)\right] = \tilde{\tau} \lim_{r\to\infty} r\mathrm{E}_\nu \left[\log_2\left(1 + \frac{\tau\tilde{\tau}P\nu}{1 + \frac{P+1}{2^{C_t/r}-1}}\right)\right], \tag{14}$$

where λ is one of the unordered eigenvalues $\{\lambda_i\}$ and $\nu \triangleq \frac{\lambda}{\min\{t,r\}}$ is a random variable with some finite mean. We can exchange the order of the expectation and the limit due to dominant convergence

$$\lim_{r\to\infty} R_{EC} \leq \tilde{\tau}\mathrm{E}_\nu \left[\lim_{r\to\infty} r\log_2\left(1 + \frac{\tau\tilde{\tau}P\nu}{1 + \frac{P+1}{2^{C_t/r}-1}}\right)\right] = \tilde{\tau}\mathrm{E}_\nu \left[C_t\frac{\tau\tilde{\tau}P\nu(1+P)}{(1+P)^2}\right] = C_t\frac{P}{1+P}, \tag{15}$$

where $\mathrm{E}\nu = \max\{\tau, \frac{1}{\tau}\}$. Since also $\mathrm{argmax}_{1\leq i\leq r}\frac{|H_i|^2}{t} \to 1$, the inequality in equation (15) is in fact an equality. Thus we have the following corollary:

Corollary 1: In the limit of $r, t \to \infty$, an achievable rate using elementary compression is $C_t\frac{P}{1+P}$.

Discussion: This result can be explained by noticing that the MIMO channel capacity is approximately linear with r when P is fixed, which leaves the fixed C_t to limit the performance, where we can not get to C_t because of the nomadic setting. In addition, this rate reaches C_t in the limit of large P, as expected. Notice that the rate (15) does not depend on the ratio between the number of receive and transmit antennas, τ. This is because the signal to noise ratio (SNR) at the final destination, from every antenna, is very small ($C_t/r = \log(1 + \frac{1+P}{D}) \to C_t/r = \frac{1+P}{D}$). So that the total SNR at the final destination is $\frac{P}{1+D} = \frac{PC_t}{C_t+(1+P)r}$, and the achievable rate can be calculated as (small P'):

$$\log_2 |I + P'/tHH^*| \to rP'. \tag{16}$$

Notice that (16) indeed does not depend on τ. Taking $P' = \frac{PC_t}{C_t+(1+P)r}$ in (16) results with (15).

B. Block fading channel

For block fading channel, the Shannon capacity is zero, and the concept of rate-vs-outage is the leading figure of merit.

1) Rate vs Outage:

Proposition 3: The rate-vs-outage region for the block fading channel, is calculated using the same equations (10) and (11), used for the fast fading channel only without the expectation over H. This results with an achievable outage probability for rate R, calculated as

$$\Pr(outage) = \min_{Q\in\mathcal{P}} \Pr\left(R > \log_2 \det\left(I_r + \text{diag}\left(\frac{2^{C_i}-1}{2^{C_i}+H_iQH_i^*}\right)_{i=1}^{r} HQH^*\right)\right). \tag{17}$$

The underlying MIMO channel enables us to analyze the proposed schemes using the diversity multiplexing tradeoff.

2) Diversity Multiplexing Tradeoff (DMT): An analysis for the diversity-multiplexing tradeoff is given next. The diversity and multiplexing are defined in the end of section II. Since our links are lossless, any outage event in the system is due to the underlying block fading channel. Thus, we fix all these links to carry information in the rate

$$C_i = \frac{m}{r}\log_2(P) + \epsilon, \; i = 1, \ldots, r \tag{18}$$

where $m \leq \min\{r,t\}$ is the multiplexing gain which is used by the system, and $\epsilon > 0$ is some fixed positive constant.

Proposition 4: The DMT $d(m)$ of any scheme with C_i as in (18) and non-ergodic block fading underlying channel, is upper bounded by the minimum between the piecewise linear function of $(k, (r-k)(t-k))$, for $k = 0, \ldots, \min\{t,r\}$ and

$$t\left(1 - \frac{m}{r}\right), \tag{19}$$

where m stands for the multiplexing gain.

For example, the maximum diversity achieved here is with $m = 0$, which results with $d(0) = t$, which is smaller than rt. This result can be understood by considering that when $m = 0$, the capacity of the links between the agents and the final destination are very small. So that getting good channel between the transmitter and only one agent will not suffice to forward the information. So we need a good channel at every agent, which results with diversity order of t and not rt.

An implication of the result is with respect to the MIMO broadcast channel. In order to achieve the full multiplexing gain in a MIMO broadcast channel, the transmitter is required to have full CSI [6]. Here, an elementary compression scheme, with limited cooperation between the destinations achieves the full multiplexing gain without channel state knowledge at the transmitter (which usually requires some feedback). Further, such cooperation is usually easier to obtain when the destinations are co-located.

Proof: The proof is based on the cut-set bound [23]. For any covariance constraint $\mathrm{E}[XX^*] = Q$, and channel H, any achievable rate is upper bounded by the cut-set bound, for any cut $\mathcal{S} \subseteq \{1, \ldots, r\}$

$$R_c = I(X; Y_{\mathcal{S}}|H) + \sum_{j \in \mathcal{S}^C} C_j = \log_2 \det(I_{|\mathcal{S}|} + H_{\mathcal{S}} Q H_{\mathcal{S}}^*) + (r - |\mathcal{S}|) \left[\frac{m}{r} \log_2(P) + \epsilon\right]. \quad (20)$$

So that for any scheme that achieves the rate $R(H)$ for channel H, with input covariance Q, the probability of outage is limited by

$$\forall\ R^* > 0 : \ \Pr\{R(H) < R^*\} \geq \Pr\{R_c(H) < R^*\}. \quad (21)$$

Now we can calculate the upper bound on the DMT:

$$d(m) \leq -\lim_{P\to\infty} \frac{\log(\Pr(\text{outage}))}{\log(P)} =$$
$$\min_{\mathcal{S}\subseteq\{1,\ldots,r\}} -\lim_{P\to\infty} \frac{\min_{Q\in\mathcal{P}} \log \Pr\left(\log_2 \det\left(I_{|\mathcal{S}|} + H_{\mathcal{S}} Q H_{\mathcal{S}}^*\right) < \frac{m}{r}|\mathcal{S}|\log(P) - |\mathcal{S}^C|\epsilon\right)}{\log(P)}. \quad (22)$$

Using [3], for each $\mathcal{S}$ we get that the diversity $d_{\mathcal{S}}(m)$ is the piecewise linear function connecting points $(k, (|\mathcal{S}| - k)(t-k))$, with $\frac{|\mathcal{S}|m}{r}$ as the argument. Next, we need to minimize this $d_{\mathcal{S}}(m)$ over all subsets $\mathcal{S}$. Since $\Pr\{0 < -r\epsilon\} = 0$, we can limit the search space to subsets that include at least one element. Define $s = |\mathcal{S}|$, so that we can use

$$\Pr\{\text{outage with } s\} \,\dot{\geq}\, s P^{-d_{\mathcal{S}}(m)} \doteq P^{-d_{\mathcal{S}}(m)}. \quad (23)$$

Let us use the underlying functions of $d_{\mathcal{S}}(m)$, before applying the piecewise linear operation

$$\min_{1\leq s\leq r} \left(s - \frac{sm}{r}\right)\left(t - \frac{sm}{r}\right). \quad (24)$$

The minimum of (24) is obtained by either taking $s = 1$ or $s = r$, regardless of m. Since the piecewise linear function exhibits the same behavior, we get Proposition 4. ■

Corollary 2: The elementary compression achieves the full multiplexing gain, but fails to achieve the DMT.

Proof for Corollary 2

1) Next we show that elementary compression suffices to achieve the full multiplexing gain $\bar{m} = \min\{r, t\}$. The first step is to lower bound (10) by a specific choice of Q and P_{D_i}. We can lower bound (10) by taking $Q = \frac{P}{t} I_t$ and the following suboptimal quantization noise power $P_{D_i} = \frac{P/t|H_i|^2+1}{2^{C_i}-1}$ and by further taking $\breve{P}_D \triangleq P_{D_{\breve{i}}}$, $\breve{i} = \mathrm{argmax}\{P_{D_i}\}$

$$R_{EC} \geq \mathrm{E}_H \log_2 \det\left(I_r + \frac{P}{t}\frac{1}{1+\breve{P}_D}\mathrm{diag}(\lambda_1, \ldots, \lambda_r)\right) =$$
$$\mathrm{E}_H\left[\sum_{i=1}^r \log_2\left(1 + \frac{P\lambda_i/t}{1+\breve{P}_D}\right)\right] = \mathrm{E}_H\left[\sum_{i=1}^r \log_2\left(1 + \left(1 + \frac{|H_{\breve{i}}|^2\frac{P}{t}+1}{P^{\frac{\bar{m}}{r}}2^{\epsilon}-1}\right)^{-1}\frac{P\lambda_i}{t}\right)\right], \quad (25)$$

where $\{\lambda_i\}$ are the eigenvalues of HH^*. Now since for $i = 1, \ldots, \bar{m}$ we have that $\lambda_i > 0$,

$$\lim_{P\to\infty} \frac{\log_2\left(1 + \left(1 + \frac{|H_{i^*}|^2 \frac{P}{t}+1}{P^{\frac{\bar{m}}{r}}2^{\epsilon}-1}\right)^{-1} \frac{P\lambda_i}{t}\right)}{\log_2(P)} = \frac{\bar{m}}{r}, \tag{26}$$

we get

$$\lim_{P\to\infty} \frac{R_{EC}}{\log_2(P)} = \bar{m}. \tag{27}$$

■

2) As for the DMT achieved by elementary compression (d_{EC}), upper bound the outage probability from equation (17), and calculate the resulting diversity

$$\begin{aligned} d_{EC}(m) &\le - \lim_{P\to\infty} \frac{\log \Pr\left(\log_2 \det\left(I_r + P\mathrm{diag}\left\{\frac{2^{C_i}-1}{P/t|H_i|^2}\right\}_{i=1}^{r} HH^*\right) < m\log(P)\right)}{\log(P)} \\ &= - \lim_{P\to\infty} \frac{\log \Pr\left(\log_2 \det\left(I_r + t\left(P^{\frac{m}{r}}2^{\epsilon}-1\right)H_{\measuredangle}H_{\measuredangle}^*\right) < m\log(P)\right)}{\log(P)}, \end{aligned} \tag{28}$$

where the inequality in (28) is since $\frac{2^{C_i}-1}{P/t|H_i|^2+2^{C_i}} \le \frac{2^{C_i}-1}{P/t|H_i|^2}$ and since $\log_2\det(I+HQH^*) \le \log_2\det(I+PHH^*)$ for any diagonal Q with $\mathrm{trace}(Q) \le P$ [3]. The matrix $H_{\measuredangle}$ is defined as

$$H_{\measuredangle} = \begin{bmatrix} \frac{H_1}{|H_1|} \\ \vdots \\ \frac{H_r}{|H_r|} \end{bmatrix}.$$

Next, since $\det(P^{\frac{m}{r}}I_r) = P^m$ we have the equality

$$\begin{aligned} &\Pr\left(\log_2 \det\left(I_r + t\left(P^{\frac{m}{r}}2^{\epsilon}-1\right)H_{\measuredangle}H_{\measuredangle}^*\right) < m\log(P)\right) = \\ &\qquad \Pr\left(\log_2 \det\left(P^{-\frac{m}{r}}I_r + t\left(2^{\epsilon}-P^{-\frac{m}{r}}\right)H_{\measuredangle}H_{\measuredangle}^*\right) < 0\right). \end{aligned} \tag{29}$$

Taking the limit with respect to P, one gets

$$\lim_{P\to\infty} \Pr\left(\log_2 \det\left(P^{-\frac{m}{r}}I_r + t\left(2^{\epsilon}-P^{-\frac{m}{r}}\right)H_{\measuredangle}H_{\measuredangle}^*\right) < 0\right) = \Pr\left(\log_2\det\left(H_{\measuredangle}H_{\measuredangle}^*\right) < -(t2^{\epsilon})^r\right). \tag{30}$$

Using Hadamard's inequality, as long as $r > 1$, the limit of the probability in (30) is strictly larger than zero, so that when taking the logarithm and dividing by $\log(P)$, one gets that

$$d_{EC}(m) = 0, \tag{31}$$

for all $m > 0$. So the optimal DMT is not achievable using elementary compression, for more than a single agent $r > 1$, and multiplexing gain of more than zero. ■

3) This sub-optimal DMT is since there exist correlations between the received signals at the different agents $\frac{|H_iH_j^*|}{|H_i||H_j|} > 0$, we get that $|H_{\measuredangle}H_{\measuredangle}^*| < 1$. As these correlations decrease, for example, by taking t to be very large compared with r, the outage probability becomes smaller. In the next section, we will exploit these correlations by a CEO approach, to reach the optimal DMT.

IV. CEO BASED SCHEME

In this section we consider the same setting as in the previous section, but use the technique from [24], that is compression followed by bining, for better utilization of the capacity of the links, exploiting the correlations between the received signals at the agents.

A. Ergodic Channel

Proposition 5: An achievable rate, for the ergodic channel, when using CEO compression is:

$$R_{CEO} = \max_{P_{U_i|Y_i,H}(u_i|y_i,h)} \left\{ \min_{\mathcal{S}\subseteq\{1,\ldots,r\}} \left\{ \sum_{i\in\mathcal{S}^C} [C_i - I(Y_i;U_i|X,H)] + I(U_{\mathcal{S}};X|H) \right\} \right\}, \tag{32}$$

where

$$P_{X,Y^r,H,U^r}(x,y^r,h,u^r) = P_X(x)P_H(h)\prod_{i=1}^{r} P_{Y_i|H,X}(y_i|h,x)P_{U_i|H,Y_i}(u_i|h,y_i). \tag{33}$$

Proof guidelines: The proof involves the random generation of $\boldsymbol{U}_i$ according to $\prod_{k=1}^{n} P_{U_i|H}(u_i(k)|h(k))$, and then randomly partitioning the resulting code book into 2^{nC_i} bins, as done in a WZ or a CEO based quantization. Then, each agent selects $\boldsymbol{U}_i$ which is jointly typical with the received $(\boldsymbol{Y}_i, \boldsymbol{H})$. It proceeds by sending the corresponding bin index to the final destination through the lossless link. The final destination knows $\boldsymbol{H}$ and the bins in which $\boldsymbol{U}^r$ fall in. Finally, the destination looks for $(\boldsymbol{X}, \boldsymbol{U}^r)$ which is jointly typical, and from deciding on the transmitted $\boldsymbol{X}$, declares the decoded message.

The formal proof is by degenerating Proposition 15, such that W^r are constants, and the random encoding, which is represented by f is known to all parties. ■

Focusing on the Gaussian channel, for the fast fading channel using (32) the following proposition is derived.

Proposition 6: An achievable rate when using CEO compression over Gaussian channel with fast fading is:

$$R_{CEO} = \max_{\{q_i:\mathbb{C}^{[r\times t]}\to\mathbb{R}_+\}_{i=1}^r} \left\{ \min_{\mathcal{S}\subseteq\{1,\ldots,r\}} \left\{ \mathrm{E}_H \left[\sum_{i\in\mathcal{S}^C} [C_i - q_i(H)] + \log_2\det\left(I_{|\mathcal{S}|} + \frac{P}{t}\mathrm{diag}\left(1-2^{-q_i(H)}\right)_{i\in\mathcal{S}} H_{\mathcal{S}}H_{\mathcal{S}}^* \right) \right] \right\} \right\}, \tag{34}$$

where $H_{\mathcal{S}} = \{H_i\}_{i\in\mathcal{S}}$.

The Proposition is proved by using the underlying channel $P_{U|Y,H}$ for the compression, such that the quantization noise is independent of the signal, as done for the elementary compression scheme in section III. Similarly, define P_{D_i} as the power of the circular-symmetric complex Gaussian quantization noise and $q_i(H)$ is the corresponding parameter, calculated as (13).

The rate in (34) is calculated assuming signalling with $Q = \frac{P}{t}I_t$. The proof that such signaling indeed maximizes the achievable rate is relegated to Appendix III. This means that the achievable rate from Proposition 6 applies also to the sum-rate of multi access channel, see [22]. Notice that although introducing correlation in Q improves the compression, since it uses the correlation to save bandwidth, it comes on the expense of the achievable rate, due to the reduced degrees of freedom. Thus, the total rate is still maximized by taking $Q = \frac{P}{t}I_t$.

Remark 2: The optimization over q_i in the above problem is a concave problem, and thus can be efficiently solved. The optimization results with an achievable rate, while assuming full knowledge of CSI ($\boldsymbol{H}$) in the final destination and in all the agents. However, this requirement does not impose severe limitations. This becomes evident in the sequel where Correlations 5, 7 describe special cases, where $q_i(H)$ is fixed, so only $\boldsymbol{H}_i$ is required at the agent.

Notice that (34) includes joint optimization over all possible channel realizations. A simpler non-optimal approach is to optimize separably for every channel

$$R_{CEO,2} = \max_{Q\in\mathcal{P}} \mathrm{E}_H\left[\max_{\{0\le q_i\}_{i=1}^r}\min_{\mathcal{S}\subseteq\{1,\ldots,r\}}\left\{\sum_{i\in\mathcal{S}^C}[C_i-q_i]+\log_2\det\left(I_{|\mathcal{S}|}+\mathrm{diag}\left(1-2^{-q_i}\right)_{i\in\mathcal{S}}H_{\mathcal{S}}QH_{\mathcal{S}}^*\right)\right\}\right]. \tag{35}$$

Unlike many channel coding problems over fast fading channels [25],[26], where there is no loss in optimality when using different codebook for every channel realization, here there is a strict gain to using a single codebook, such that the decoding is done jointly over the different realizations of H. So that $R_{CEO,2} < R_{CEO}$, with high probability.

Remark 3: As in [1], $\mathrm{E}q_i$ in both (34) and (35), can be interpreted as the rate wasted on the compression of the additive noise by the i^{th} agent's processing. So that, for example taking $\mathcal{S} = \{\phi\}$ in (34), results with the achievable rate of $\sum_{i=1}^r[C_i - \mathrm{E}_H q_i]$. Of course, this represents only one of $2^{|\mathcal{S}|}$ elements within the minimum.

Remark 4: When the agents do not have H, but rather only H_i, that is, each agent has only its own channel to the transmitter, and not the channels of the other agents, the optimization in Proposition 6 is done in this case over $q_i : \mathbb{C}^{[1\times t]} \to \mathbb{R}_+$.

Next, we give a solution to the optimization problem issued by Proposition 6, for the symmetric case with $r = 2$. Such setting results with

Proposition 7: An achievable rate for the symmetric setting with $r = 2$ and ergodic setting is equal to

$$R_{CEO} = 2(C - \mathrm{E}_H[q_1(H,\theta)]), \tag{36}$$

where ($\lceil a\rceil^+ = \max\{a, 0\}$)

$$q_i(H,\theta) = \begin{cases} \left\lceil -\log_2\left(\frac{\theta}{1+\theta}\frac{1+\frac{P}{t}|H_i|^2}{\frac{P}{t}|H_i|^2}\right)\right\rceil^+ & \theta > F_H(\frac{\Delta}{\Delta+|H_i|^2}) \\ \left\lceil -\log_2(\frac{\Delta+|H_{3-i}|^2}{\Delta}F_H(\theta))\right\rceil^+ & \theta \le F_H(\frac{\Delta}{\Delta+|H_i|^2}), \end{cases} \tag{37}$$

with

$$F_H(\theta) \triangleq \frac{1}{2(1+\theta)}\left(1+2\theta-\sqrt{(1+2\theta)^2-4\theta(1+\theta)\frac{(\Delta+\frac{t}{P}+|H_1|^2+|H_2|^2)\Delta}{(\Delta+|H_1|^2)(\Delta+|H_2|^2)}}\right) \tag{38}$$

$$\Delta \triangleq \frac{P}{t}\det(HH^*) \tag{39}$$

$$\tag{40}$$

and $\theta > 0$ which is set such that

$$\mathrm{E}_H\log_2\det\left(I_2+\frac{P}{t}\mathrm{diag}\left(1-2^{-q_i(H,\theta)}\right)_{i=1}^2 HH^*\right) = 2(C-\mathrm{E}_H[q_1(H,\theta)]). \tag{41}$$

The proof is relegated to Appendix IV.

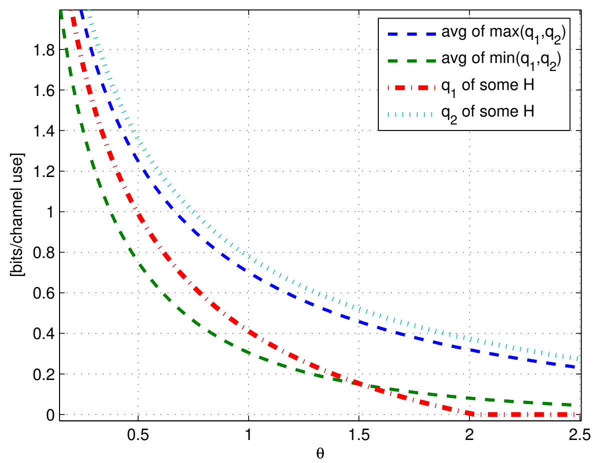

Fig. 2. The resulting compression parameters q_1 and q_2, as function of θ, the Lagrangian for some specific H, and also average results when averaging over 1000 channels H, $t = 2$ and $P = 7$ dB.

An intuition into the solution offered by Proposition 7 is by considering both q_1 and q_2. For that, assume $|H_1|^2 > |H_2|^2$, then $F_H^{-1}\left(\frac{\Delta}{\Delta+|H_1|^2}\right) < F_H^{-1}\left(\frac{\Delta}{\Delta+|H_2|^2}\right)$ and

$$\theta \leq F_H^{-1}\left(\frac{\Delta}{\Delta+|H_1|^2}\right) : \quad \begin{cases} q_1 &= -\log_2\left(\frac{\Delta+|H_2|^2}{\Delta} F_H(\theta)\right) \\ q_2 &= -\log_2\left(\frac{\Delta+|H_1|^2}{\Delta} F_H(\theta)\right) \end{cases} \tag{42}$$

$$F_H^{-1}\left(\frac{\Delta}{\Delta+|H_1|^2}\right) < \theta < \frac{P}{t}|H_1|^2 : \quad \begin{cases} q_1 &= -\log_2\left(\frac{\theta}{1+\theta}\frac{1+\frac{P}{t}|H_1|^2}{\frac{P}{t}|H_1|^2}\right) \\ q_2 &= 0 \end{cases} \tag{43}$$

$$\frac{P}{t}|H_1|^2 \leq \theta : \quad \begin{cases} q_1 &= 0 \\ q_2 &= 0. \end{cases} \tag{44}$$

This reveals the structure of the optimal solution, which can be described as a variant of the famous "water-filling". This is since as in classic water-filling, depending on the available bandwidth, the parameter θ determines how the compression depends on the channel realizations. When C is very large, θ is very small, and fewer channel realizations result with $q_1 = q_2 = 0$ (44). When $q_1 = q_2 = 0$ the scheme does not relay any information regarding the channel outputs, thus saving bandwidth for better channel realizations.

This is demonstrated in Figure 2, for $P = 7$ dB, and 2×2 system, where the averaged maximum and minimum of q_1 and q_2, over 1000 channels is depicted, as function of the Lagrangian θ. It is observed that the average difference between the two compression parameters q_1 and q_2 is about 0.4 bits/channel use. Figure 2 also draws q_1 and q_2 for some specific channel H. It is seen that q_2 is always larger than q_1, since $|H_1|^2 < |H_2|^2$, for the specific channel. Form $\theta = 2$ on, $q_1 = 0$, which means that no information is sent from agent A_1 to the final destination for this channel realization, when $\theta \geq 2$.

1) An Achievable Rate when $r, t \to \infty$: For the case where $r/t = \tau$, $C_i = C_t/r$ and $r \to \infty$, we repeat the suboptimal assignment and again fix $q_i = q^* = \varepsilon C_t/r$.

Next, we define $q_t = rq^* = \varepsilon C_t$. Now we can write for any $\mathcal{S}$:

$$\log_2 \left| I_{|\mathcal{S}|} + \frac{P}{t}(1-2^{q_t/r}) H_{\mathcal{S}} H_{\mathcal{S}}^* \right| = \sum_{i=1}^{|\mathcal{S}| \vee t} \log_2(1+P/t(1-2^{q_t/r})\lambda_i) \to_{r\to\infty} \tau_{\mathcal{S}} r \mathrm{E} \log_2 \left(1 + P(1-2^{q_t/r})\tau\tau_{\mathcal{S}}\nu_{\mathcal{S}}\right) \tag{45}$$

Where $\tau_{\mathcal{S}} \triangleq \frac{|\mathcal{S}| \vee t}{r}$, $\nu_{\mathcal{S}} \triangleq \frac{\lambda}{|\mathcal{S}| \vee t}$ and $\vee$ denotes min.

Now we can exchange the order of the expectation and the limit due to dominant convergence:

$$\tau_{\mathcal{S}} \mathrm{E} \left[\lim_{r\to\infty} r \log_2 \left(1 + P(1-2^{q_t/r})\tau\tau_{\mathcal{S}}\nu_{\mathcal{S}}\right)\right] = \tau\tau_{\mathcal{S}}^2 q_t P \mathrm{E}[\nu_{\mathcal{S}}] = P q_t \frac{|\mathcal{S}|}{r}, \tag{46}$$

since $\mathrm{E}[\nu_{\mathcal{S}}] = \max\left\{\frac{|\mathcal{S}|}{t}, \frac{t}{|\mathcal{S}|}\right\}$. On the other hand, for that same $\mathcal{S}$:

$$\sum_{i\in\mathcal{S}} [C_i - q_i] = \frac{|\mathcal{S}|}{r}(C_t - q_t). \tag{47}$$

Next, we set q_t, such that the right hand sides of (46) and (47) are equal. This results with the achievable rate of $R_{CEO} = C_t - q_t = C_t \frac{P}{P+1}$. Notice that this rate is identical to the elementary compression (15). One would expect that the Wyner-Ziv approach will improve as τ is increased, because then the correlations between the received signals is increased, improving the compression rates. However, from (15), it is observed that for small powers, the mutual information is independent of t, so that also the correlations between the received signals Y_i are independent of τ. In addition, from the discussion below Correlation 1, it is evident that the equivalent signal to noise ratio when received at the final destination is very low, so that the inter-agent correlations are also low, diminishing the effect of the CEO compression.

B. Block Fading Channel

As in elementary compression, here we again use the rate-vs-outage figure of merit, and then also give the DMT for the CEO based scheme.

1) Rate vs Outage: For the non-ergodic block fading channel, equation (34), stands for the averaged mutual information. Since the rate-vs-outage is not concave with respect to Q, as in the fast fading channel, $Q = \frac{P}{t} I$ is no longer optimal [2], and we need to optimize also over Q.

Proposition 8: An achievable rate R is correctly received over a block fading channel, with an outage probability of at most ϵ, as long as the following holds (obtained from (34)):

$$\Pr\left(\max_{Q\in\mathcal{P}, \{0\le q_i \le C_i\}_{i=1}^r} \left\{ \min_{\mathcal{S}\subseteq\{1,\dots,r\}} \left\{ \log_2 \det\left(I_{|\mathcal{S}|} + \mathrm{diag}\left(1-2^{-q_i}\right)_{i\in\mathcal{S}} H_{\mathcal{S}} Q H_{\mathcal{S}}^* \right) + \sum_{i\in\mathcal{S}^C} [C_i - q_i] \right\}\right\} < R \right) \le \epsilon \tag{48}$$

where the probability is with respect to H.

2) Diversity Multiplexing Tradeoff (DMT): The CEO approach can get to the upper bound of the DMT, and thus gives the optimal DMT.

Proposition 9: The full Diversity Multiplexing Tradeoff $d(m)$ *is the minimum between the piecewise linear function of* $(k,(r-k)(t-k)$*, for* $k=0,\ldots,\min\{t,r\}$ *and*

$$t\left(1-\frac{m}{r}\right), \tag{49}$$

where $0\leq m\leq\min\{r,t\}$*. This tradeoff can not be achieved using the elementary compression, only using the CEO approach.*

This Proposition is proved by showing that the upper bound on the DMT from Proposition 4 is achievable.

Proof: Consider again $C_i=\frac{m}{r}\log(P)+\epsilon$ and then fix $q_i=0.5\epsilon$ in equation (48). Let us write the diversity here as d_{CEO}, where CEO stands for chief executive officer

$$\begin{aligned}d_{CEO}(m)&=-\lim_{P\to\infty}\frac{\log(\Pr(\text{outage}))}{\log(P)}=\\&\min_{\mathcal{S}\subseteq\{1,\ldots,r\}}-\lim_{P\to\infty}\frac{\min_{Q\in\mathcal{P}}\Pr\left(\log_2\det\left(I_{|\mathcal{S}|}+(1-2^{-0.5\epsilon})H_{\mathcal{S}}QH_{\mathcal{S}}^*\right)<\frac{m}{r}|\mathcal{S}|\log(P)-0.5|\mathcal{S}^C|\epsilon\right)}{\log(P)}.\end{aligned}\tag{50}$$

The difference between the upper bound in equation (22) and (50) is with the attenuation of $(1-2^{0.5\epsilon})$. Since this attenuation diminishes as P gets large, it is evident that we get the same diversity as the upper bound.

Next, we show the achievability of the full multiplexing gain, thus proving the DMT. We get the following achievable rate:

$$R_{CEO}=\bar{m}\log_2(P)+o(\log_2(P)), \tag{51}$$

where $\bar{m}=\min\{r,t\}$ and $\lim_{P\to\infty}\frac{o(\log_2(P))}{\log_2(P)}=0$. This is since

$$\min_{\mathcal{S}\subseteq\{1,\ldots,r\}}\left\{|\mathcal{S}|\frac{\bar{m}}{r}\log_2(P)+\min\{r-|\mathcal{S}|,\bar{m}\}\log_2(P)+o(\log_2(P))\right\}=\bar{m}\log_2(P)+o(\log_2(P)) \tag{52}$$

is fulfilled with $\mathcal{S}=\phi$ and $\mathcal{S}=\{1,\ldots,r\}$. ■

C. An Achievable Rate For the Case of Multiple Antennas Also At the Agents

The case of multiple antennas at the agents is different than the above case, where only a single antenna was used by the agents, in that now the agents can use more elaborated processing in order to improve the overall performance. We consider here only ergodic channel, where the block fading case follows the same line.

The channel can still be described by (2), only that now, $Y_i(k)$ is a vector, taking values from $\mathbb{C}^{[r_i\times 1]}$, $N_i(k)\sim\mathcal{CN}(0,I_{r_i})$, and $H_i(k)\in\mathbb{C}^{[r_i\times t]}$, again with elements that are independently and identically distributed, according to the circular-symmetric complex Gaussian distribution with variance of 1.

The difference between this scheme and the previous one, is that now each agent can add non-white quantization noise (but still input independent) to the received vector, where such dependency can improve the resulting achievable rate, by improving the estimation at the final destination, through better utilization of the lossless links.

Proposition 10: An achievable rate, over an ergodic channel, with several receiving antennas at each agent, is

$$R_{CEO} = \max_{\{\Lambda_i(H):\mathbb{C}^{[r\times t]}\to\mathbb{B}_i\}_{i=1}^r} \min_{\mathcal{S}\subseteq\{1,\dots,r\}}$$
$$\left[\mathrm{E}_H\left\{\sum_{i\in\mathcal{S}^C}[C_i - \log_2|I_{m_i}+\Lambda_i^{-1}|] + \log_2\left|I_{\sum_{i\in\mathcal{S}} m_i} + \frac{P}{t}\mathrm{diag}\left((I_{m_i}+\Lambda_i)^{-1}_{i\in\mathcal{S}}\right)H_{\mathcal{S}}H_{\mathcal{S}}^*\right|\right\}\right], \quad (53)$$

where

$$H_{\mathcal{S}} = \begin{pmatrix} \vdots \\ \Gamma_i u_i \\ \vdots \end{pmatrix}_{i\in\mathcal{S}} \quad (54)$$

and

$$\mathbb{B}_i = \{M:\ M\in\mathbb{C}^{m'\times m'},\ m'\le\min\{r_i,t\}, M\succeq 0\}. \quad (55)$$

To achieve this rate, each agent performs singular value decomposition of $H_i = v_i\Gamma_i u_i$, so that $v_i\in\mathbb{C}^{[r_i\times r_i]}$ and $u_i\in\mathbb{C}^{[t\times t]}$ are unitary matrices, for calculating $v_i^*Y_i$. Then each agent looks for $\boldsymbol{U}_i^n$ which is jointly typical with $(\boldsymbol{v}_i^*)\boldsymbol{Y}_i$, when U_i and v_iY_i are distributed as

$$U_i = v_i^*Y_i + D_i. \quad (56)$$

Here D_i is random vector, independent with Y_i, distributed as $\mathcal{NC}(0,\Lambda_i)$. Define $m_i = \mathrm{rank}(\Gamma_i)$ and redefine the matrix Γ_i to include only the non-zero elements in Γ_i. The matrix $\Lambda_i\in\mathbb{C}^{[m_i\times m_i]}$ represents m_i random variables, like in the previous section, only here it is a vector instead of a scalar.

Note that $Q=\frac{P}{t}I_t$ is optimal in (53) as in (34). By assigning $r_i=1$, $\Lambda_i=P_{D_i}$ and noticing that $\Gamma_i u_i = H_i$, we see that indeed (53) coincides with (34), as expected.

V. Upper Bounds

In this section several upper bounds are derived, for both fast fading and block fading cases.

A. Cut-Set Upper Bound

The simple cut-set upper bound [23], although quite intuitive often provides good results. This bound is very general, and is not limited to the nomadic setting.

Corollary 3: Cut-set: Any achievable rate in the system is upper bounded by the cut-set bound,

$$R \le \min_{\mathcal{S}\subseteq\{1,\dots,r\}}\left[I(X;Y_{\mathcal{S}}|H) + \sum_{i\in\mathcal{S}^C} C_i\right]. \quad (57)$$

For the ergodic fast fading channel, this upper bound equals

$$R \le \min_{\mathcal{S}\subseteq\{1,\dots,r\}}\left[\mathrm{E}_H\log_2\det\left(I_{|\mathcal{S}|}+\frac{P}{t}H_{\mathcal{S}}H_{\mathcal{S}}^*\right) + \sum_{i\in\mathcal{S}^C} C_i\right]. \quad (58)$$

Where for the block fading channel, the rate vs outage is limited by

$$\Pr(outage) = \min_{Q\in\mathcal{P}}\Pr\left(R > \min_{\mathcal{S}\subseteq\{1,\dots,r\}}\left[\log_2\det\left(I_{|\mathcal{S}|}+H_{\mathcal{S}}QH_{\mathcal{S}}^*\right) + \sum_{i\in\mathcal{S}^C} C_i\right]\right). \quad (59)$$

The proof is based on [23], considering also the proof of Proposition 4, and is omitted due to its simplicity.

B. Upper Bounds for Nomadic Transmitter

The upper bounds here are calculated assuming nomadic transmitter, who uses circular-symmetric complex Gaussian codebook. Thus they show what cannot be achieved, no matter what processing is used at the agents, as long as they are ignorant of the codebook used. In the following, we first upper bound general channels, and then apply the bound for ergodic channel and the block fading channel.

Proposition 11: The achievable rate for reliable communication is upper bounded by:

$$R \leq \min_{\mathcal{S}\subseteq\{1,\dots,r\}} \left\{ \sum_{i\in\mathcal{S}} [C_i - q_i] + \frac{1}{n} I(\boldsymbol{X}; V_{\mathcal{S}^C}|\boldsymbol{H}) + \frac{1}{n} \right\}. \tag{60}$$

Proof: We first give an information theoretic upper bound for the achievable rate, based on [1]. We define V_i to be the message sent from agent A_i after receiving n channel outputs. Notice that $\boldsymbol{H}$ is fully known to all agents and to the final destination, so they can use it to calculate the $\{V_i\}$.

For any subset $\mathcal{S} \subseteq \{1,\dots,r\}$, the following chain of inequalities holds:

$$\begin{aligned}
\sum_{i\in\mathcal{S}} C_i &\geq \frac{1}{n} I(\boldsymbol{Y}^r; V_{\mathcal{S}}|V_{\mathcal{S}^C}, \boldsymbol{H}) && (61)\\
&= \frac{1}{n} I(\boldsymbol{Y}^r; V^r|\boldsymbol{H}) - \frac{1}{n} I(\boldsymbol{Y}^r; V_{\mathcal{S}^C}|\boldsymbol{H}) && (62)\\
&= \frac{1}{n} I(\boldsymbol{Y}^r, \boldsymbol{X}; V^r|\boldsymbol{H}) - \frac{1}{n} I(\boldsymbol{Y}^r, \boldsymbol{X}; V_{\mathcal{S}^C}|\boldsymbol{H}) && (63)\\
&= \frac{1}{n} I(\boldsymbol{X}; V^r|\boldsymbol{H}) - \frac{1}{n} I(\boldsymbol{X}; V_{\mathcal{S}^C}|\boldsymbol{H}) + \frac{1}{n} I(\boldsymbol{Y}^r; V^r|\boldsymbol{X}, \boldsymbol{H}) - \frac{1}{n} I(\boldsymbol{Y}_{\mathcal{S}^C}; V_{\mathcal{S}^C}|\boldsymbol{X}, \boldsymbol{H}) && (64)\\
&= \frac{1}{n} I(\boldsymbol{X}; V^r|\boldsymbol{H}) - \frac{1}{n} I(\boldsymbol{X}; V_{\mathcal{S}^C}|\boldsymbol{H}) + \sum_{i=1}^{r} q_i - \sum_{i\in\mathcal{S}^C} q_i && (65)\\
&= \frac{1}{n} I(\boldsymbol{X}; V^r|\boldsymbol{H}) - \frac{1}{n} I(\boldsymbol{X}; V_{\mathcal{S}^C}|\boldsymbol{H}) + \sum_{i\in\mathcal{S}} q_i. && (66)
\end{aligned}$$

where (63) is because V_i is a function of $\boldsymbol{Y}_i$ and $\boldsymbol{H}$, so we have the Markov chain $V_i - \{Y_i, H\} - X$ and q_i is defined by $q_i \triangleq \frac{1}{n} I(\boldsymbol{Y}_i; V_i|\boldsymbol{X}, \boldsymbol{H})$. By changing order we get

$$\frac{1}{n} I(\boldsymbol{X}; V^r|\boldsymbol{H}) \leq \sum_{i\in\mathcal{S}} [C_i - q_i] + \frac{1}{n} I(\boldsymbol{X}; V_{\mathcal{S}^C}|\boldsymbol{H}). \tag{67}$$

Next we utilize Fano's inequality

$$\begin{aligned}
R &\leq \frac{1}{n} H(M) = \frac{1}{n} I(M; V^r, F|\boldsymbol{H}) + \frac{1}{n} H(M|F, V^r, \boldsymbol{H}) && (68)\\
&\leq \frac{1}{n} I(M; V^r, F|\boldsymbol{H}) + P_e && (69)\\
&\leq \frac{1}{n} I(M, F; V^r|\boldsymbol{H}) + P_e && (70)\\
&\leq \frac{1}{n} I(\boldsymbol{X}(M, F); V^r|\boldsymbol{H}) + P_e && (71)\\
&\leq \sum_{i\in\mathcal{S}} [C_i - q_i] + \frac{1}{n} I(\boldsymbol{X}; V_{\mathcal{S}^C}|\boldsymbol{H}) + P_e. && (72)
\end{aligned}$$

■

The following lemma, which is proved in the Appendix, is required for obtaining computable upper bounds (single letter upper bound).

Lemma 1: If the transmitter is nomadic, so the agents have no decoding ability, and the transmitter uses Gaussian codebooks, the following inequality holds for any $\mathcal{S} \subseteq \{1, \ldots, r\}$:

$$\frac{1}{n} I(\boldsymbol{X}; V_{\mathcal{S}} | \boldsymbol{H} = \boldsymbol{h}) \leq m \log_2 \left(\prod_{k=1}^{n} \left| I_{|\mathcal{S}|} + \Lambda_{\mathcal{S}}(k) \right|^{\frac{1}{nm}} - \prod_{k=1}^{n} |W_{\mathcal{S}}(k)|^{\frac{1}{nm}} \right) \tag{73}$$

where $\Lambda_{\mathcal{S}}(k) \triangleq H_{\mathcal{S}}(k) Q H^*_{\mathcal{S}}(k)$,

$$W_{\mathcal{S}}(k) \triangleq \begin{cases} Q H_{\mathcal{S}}(k)^* \mathrm{diag}\left(2^{-q_i(\boldsymbol{h})}\right)_{i \in \mathcal{S}} H_{\mathcal{S}}(k) & |\mathcal{S}| > t \\ \mathrm{diag}\left(2^{-q_i(\boldsymbol{h})}\right)_{i \in \mathcal{S}} H_{\mathcal{S}}(k) Q H_{\mathcal{S}}(k)^* & |\mathcal{S}| \leq t \end{cases} \tag{74}$$

$q_i(\boldsymbol{h}) \triangleq \frac{1}{n} I(Y_i^n; V_i | \boldsymbol{X}, \boldsymbol{H} = \boldsymbol{h})$ and $m \triangleq \min\{t, |\mathcal{S}|\}$.

Since HQH^* is distributed the same as $HU^* \Sigma U H^*$, when U is a unitary matrix and Σ is diagonal, Q can be restricted to be diagonal in (73). However, unlike the achievable rate, which is a concave function of Q, so that $Q \propto I$ is optimal, the right hand side of (73) is not concave in Q, thus in the sequel, we let Q be such that $Q \in \mathcal{P}$. Notice that the inequality in (73) is tight when the channel is $H = (1, \ldots, 1)^T$, which corresponds to the Gaussian CEO problem with quadratic distortion [15].

1) Upper Bound for Fast Fading Channel: We begin the derivation of an upper bound for the fast fading channel by evaluating the bound of Lemma 1 for the fast fading:

Corollary 4: In the limit of $n \to \infty$, due to the ergodic fading process:

$$\lim_{n \to \infty} \frac{1}{n} I(\boldsymbol{X}; V_{\mathcal{S}} | \boldsymbol{H} = \boldsymbol{h}) \leq F(\mathcal{S}, q_{\mathcal{S}}) \tag{75}$$

where

$$F(\mathcal{S}, q_{\mathcal{S}}) \triangleq m \log_2 \left(2^{\frac{1}{m} \mathrm{E}_{H(1)} \log_2 |I + \Lambda_{\mathcal{S}}|} - 2^{\frac{1}{m} \mathrm{E}_{H(1)} \log_2 |W_{\mathcal{S}}|} \right), \tag{76}$$

and we use the notation $q_i \triangleq q_i(\boldsymbol{h})$ and $q_{\mathcal{S}} \triangleq \{q_i\}_{i \in \mathcal{S}}$, and $\Lambda_{\mathcal{S}} = \Lambda_{\mathcal{S}}(1)$, $W_{\mathcal{S}} = W_{\mathcal{S}}(1)$. Consequently, (75) can be averaged over the channels:

$$\lim_{n \to \infty} \frac{1}{n} I(\boldsymbol{X}; V_{\mathcal{S}} | \boldsymbol{H}) \leq F(\mathcal{S}, q_{\mathcal{S}}). \tag{77}$$

The dependence of F from (76) on q_i, stems from the definition of q_i, as the bandwidth used for the noise compression, and is essential for the bound, as it is used for connecting the bandwidth for the signal compression to the achievable rate. Combining proposition 11 with Corollary 4 above, we get the following proposition:

Proposition 12: The achievable rate of a nomadic transmitter, over fast fading channel, is upper bounded by:

$$R \leq \max_{Q \in \mathcal{P}, \{0 \leq q_i \leq C_i\}} \left\{ \min_{\mathcal{S} \subseteq \{1, \ldots, r\}} \left\{ F(\mathcal{S}^C, q_{\mathcal{S}}) + \sum_{i \in \mathcal{S}} [C_i - q_i] \right\} \right\}. \tag{78}$$

Remark 5: When $C_i = C$ for $i = 1, \ldots, r$, then the argument which is maximized over $\{q_i\}_{i=1}^r$ in (78), is symmetric in $\{q_i\}_{i=1}^r$. Since the argument is also concave in $\{q_i\}_{i=1}^r$, for $C_i = C$, equation (78) is maximized by $q_i = q^*$ for $i = 1, \ldots, r$. So that for the symmetric case:

$$R \leq \max_{Q \in \mathcal{P}, 0 \leq q^* \leq C} \min_{\mathcal{S} \subseteq \{1, \ldots, r\}} \left\{ F(\mathcal{S}^C, q^*) + |\mathcal{S}| [C - q^*] \right\}. \tag{79}$$

Following remark 5, we give a special case where the upper bound in proposition 12 is tight. Notice that in this case, the optimal compression strategy used by the agents, is with fixed $q^* = q_i$. This means that the each agent is required to know only its own H_i, and not the other agents $\{H_j\}_{j\neq i}$. Furthermore, notice that this conclusion is due to the tight upper bound, and is not trivially obtained from the achievable rate (34) alone.

Corollary 5: The CEO approach is optimal for infinite transmission power, $Q = \frac{P}{t}I$, and $C_i = C$, $i = 1, \ldots, r$ Here we take $P \to \infty$, and fixed t and r.

Proof: We show it for $r \leq t$, where the proof for $r > t$ follows the same lines.

The achievable rate: Taking $P \to \infty$ and optimizing over q_{CEO} (where $q_i = q_{CEO},\ i = 1, \ldots, r$ in equation (34)) instead of over $\{q_i\}$, results with:

$$\frac{1}{n}I(\boldsymbol{X}; V_{\mathcal{S}}|\boldsymbol{H}) = m\log_2(P) + \mathrm{E}_H \log_2 |\frac{1}{t}H_{\mathcal{S}}H^*_{\mathcal{S}}| + m\mathrm{E}_H \log_2\left(1 - 2^{-q_{CEO}}\right) + o(P), \tag{80}$$

where $o(P) \to 0$ when $P \to \infty$.

The upper bound: On the other hand, taking $P \to \infty$ equation (75) becomes

$$F(\mathcal{S}, q_{\mathcal{S}}) = m\log_2\left(2^{\log_2(P) + \frac{1}{m}\mathrm{E}_H \log_2 |\frac{1}{t}H_{\mathcal{S}}H^*_{\mathcal{S}}|}\left(2^{o(P)} - \prod_{i\in\mathcal{S}} 2^{-\frac{q_i}{|\mathcal{S}|}}\right)\right) =$$

$$m\log_2(P) + \mathrm{E}_H \log_2 |\frac{1}{t}H_{\mathcal{S}}H^*_{\mathcal{S}}| + m\log_2\left(2^{o(P)} - \prod_{i\in\mathcal{S}} 2^{-\frac{q_i}{|\mathcal{S}|}}\right), \tag{81}$$

Since $C_i = C$, equation (78) is a concave symmetric function of $\{q_i\}$, the solution is when all $\{q_i\}$ are identical, denoted as $q_i = q_{UB}$. So (81) becomes

$$F(\mathcal{S}, q_{UB}) = m\log_2(P) + \mathrm{E}_H \log_2 |\frac{1}{t}H_{\mathcal{S}}H^*_{\mathcal{S}}| + m\log_2\left(2^{o(P)} - 2^{-q_{UB}}\right). \tag{82}$$

which is identical, in the limit, to (80). Substituting (80) in (34) and (82) in (79) gives the desired equality. ■

For $P \to \infty$ and $C_i = C$, there is no need to perform expectation over H of the rightmost element of (80), since taking $q_{CEO} = q_{UB}$ results with the optimal rate. This means that for large P and symmetric links, the compression parameters are independent of H, which in turn means that the i-th agent needs to know only its own H_i. Notice that the channel state information (CSI, H_i) is still required at i^{th} agent, for the determination of the codebook of U (see [1]). This is unlike the classical Gaussian Wyner Ziv problem, which does not benefit from side information at the encoder.

The upper bound of proposition 12 is not tight because the upper bound in Lemma 1 was obtained using the vector version of the entropy power inequality. This inequality is known to be tight only for proportional correlation matrices, which is not our case. Thus the entropy power inequality introduces a gap that prevents the bound to be tight. This gap can be mitigated by taking into account smaller matrices. The following proposition improves upon proposition 12 by optimizing also over sub-matrices of $\mathcal{S}$:

Proposition 13: An achievable rate of a nomadic transmitter, which uses circular-symmetric complex Gaussian *signalling with total power P, through agents with bandwidths $\{C_i\}$ is upper bounded by:*

$$R_u \triangleq \max_{Q\in\mathcal{P}\{0\leq q_i\leq C_i\}_{i=1}^r} \left\{ \min_{\substack{\cup_{j=1}^r \mathcal{Z}_j \subseteq \{1,\dots,r\},\\ i\neq j:\ \mathcal{Z}_j\cap\mathcal{Z}_i=\phi}} \left\{ \sum_{j=1}^r F(\mathcal{Z}_j, q_{\mathcal{Z}_j}) + \sum_{i\in\cap_{j=1}^r \mathcal{Z}_j^c} [C_i - q_i] \right\} \right\} \tag{83}$$

where $F(\mathcal{Z}_j, q_{\mathcal{Z}_j})$ is defined as before, in equation (75).

The proof is very simple, considering for every group of disjoint subsets ($\{\mathcal{Z}_j\}_{j=1}^r:\ \mathcal{Z}_j\cap\mathcal{Z}_i=\phi$ when $i\neq j$) that cover $\cup_{j=1}^r\mathcal{Z}_j=\mathcal{S}$ we can write:

$$I(\boldsymbol{X}; V_{\mathcal{S}}|\boldsymbol{H}) \leq \sum_{j=1}^r I(\boldsymbol{X}; V_{\mathcal{Z}_j}|\boldsymbol{H}), \tag{84}$$

which is due to the Markov chain $V_j - X - V_i$ when $i \neq j$, and then using the upper bound of proposition 12 again, for every element. Since the entropy power inequality, which is used in proposition 12 is not tight (in general) for the Gaussian vector case, but is tight for the Gaussian scalar case, this upper bound can improve upon the latter. For the symmetric case, where $C_i = C$ for $i = [1,\dots,r]$, due to the concavity of (83), the maximum in (83) is achieved with $q_i = q^*,\ i = [1,\dots,r]$, so that (83) is written as:

$$R_u = \max_{Q\in\mathcal{P},0\leq q^*\leq C} \left\{ \min_{\substack{\sum_{j=1}^r jk_j \leq r,\\ k_j\geq 0}} \left\{ \sum_{j=1}^r k_j F(j,q^*) + (r - \sum_{j=1}^r jK_j)(C-q^*) \right\} \right\} \tag{85}$$

By solving the above optimization problem for $\{k_j\}_{j=1}^r$ and then solving for q^* by explicitly writing $F(j,q^*)$ we can simplify (85) to

Corollary 6: The achievable rate of nomadic transmitter in the symmetric case, $C_i = C,\ i = 1,\dots,r$, is upper bounded by

$$R_{us} \triangleq rC + r\max_{Q\in\mathcal{P}} \left\{ \min_{1\leq j\leq r} \left\{ \frac{1}{j}\mathrm{E}_{H_j}\log_2|I_j + H_jQH_j^*| - \log_2\left(2^C + 2^{\frac{1}{j}\mathrm{E}_{H_j}\log_2|H_jQH_j^*|}\right) \right\} \right\} \tag{86}$$

where H_j is the fading coefficients seen by any subset of j agents (since the channel is ergodic, it does not matter which subset).

The improvement of the bound from proposition 13 over the bound from proposition 12, is seen in the next corollary, where the inequality (84) is in fact an equality, and a conclusive result is obtained.

Corollary 7: The CEO approach is optimal for $Q = \frac{P}{t}I$ and $t\to\infty$ while r is fix.

The bound (83) is tight, when $t\to\infty$ and Q is a multiplicity of the identity matrix. This is since HQH^* is proportional to the identity matrix, each agent receives independent reception. This means r parallel links that can be optimized separately. Namely, when $t\to\infty$ while r is fixed we get

$$\lim_{t\to\infty}\frac{1}{t}HH^* = I_r. \tag{87}$$

Proof:

The achievable rate: assigning the limit (87) in (34), we get:

$$\lim_{t\to\infty} R(H) = \max_{\{0\le q_i\le C_i\}}\left\{\min_{\mathcal{S}}\left\{\sum_{i\in\mathcal{S}^C}[C_i-q_i]+\sum_{i\in\mathcal{S}}\log_2(1+P(1-2^{-q_i}))\right\}\right\}. \tag{88}$$

Notice that (88) is independent of the channel realization H.

The upper bound: On the other hand, taking $Q=\frac{P}{t}I_t$ and $\frac{1}{t}H_iQH_i^*=1$ for the calculation of $F(\{i\},q_i)$ in (75) gives $\log_2(1+P(1-2^{-q_i}))$. Assigning back to equation (83), with $\mathcal{Z}_i=\{i\}$ results with:

$$\lim_{t\to\infty} R_u = \max_{\{0\le q_i\le C_i\}}\left\{\min_{\mathcal{S}}\left\{\sum_{i\in\mathcal{S}^C}[C_i-q_i]+\sum_{i\in\mathcal{S}}\log_2(1+P(1-2^{-q_i}))\right\}\right\}, \tag{89}$$

which equals (88) and proves the optimality. ■

2) Upper Bound for Block Fading Channels: In this subsection, we will consider the case of H distributed independently, but once per block, such that $\boldsymbol{H}=H$. The resulting rate in equation (34) is actually the average rate, supported by the scheme. In the sequel of this subsection, we will upper bound the rate-vs.-outage of the scheme.

For the upper bound, we again use:

$$R(\boldsymbol{H}=\boldsymbol{h}) \le \max_{\{q_i\}_1^r}\min_{\mathcal{S}}\left\{\frac{1}{n}I(V_{\mathcal{S}};\boldsymbol{X}|\boldsymbol{H}=\boldsymbol{h})+\sum_{i\in\mathcal{S}^C}[C_i-q_i]\right\}. \tag{90}$$

For $I(V_{\mathcal{S}};\boldsymbol{X}|\boldsymbol{H}=\boldsymbol{h})$, we use the upper bound of equation (73). Since $\boldsymbol{H}=H$, we get:

$$G(\mathcal{S},q_{\mathcal{S}}) \triangleq m\log_2\left(\left|I_{|\mathcal{S}|}+\Lambda_{\mathcal{S}}\right|^{\frac{1}{m}}-|W_{\mathcal{S}}|^{\frac{1}{m}}\right) \tag{91}$$

$$\frac{1}{n}I(\boldsymbol{X};V_{\mathcal{S}}|H=\boldsymbol{h}) \le G(\mathcal{S},q_{\mathcal{S}}) \tag{92}$$

where $\Lambda_{\mathcal{S}}=H_{\mathcal{S}}QH_{\mathcal{S}}^*$, as before and $W_{\mathcal{S}}$ is defined by $W_{\mathcal{S}}(1)$ from equation (74). Combining (90) and (92) and noticing that H is a random variable, we get the following upper bound on the outage ϵ vs. rate R:

Proposition 14: An upper bound on the achievable rate R, for given outage probability ϵ is the minimal R which fulfills:

$$P\left(\max_{Q\in\mathcal{P},\{0\le q_i\le C_i\}}\left\{\min_{\mathcal{S}\subseteq\{1,\ldots,r\}}\left\{G(\mathcal{S},q_{\mathcal{S}})+\sum_{i\in\mathcal{S}^C}[C_i-q_i]\right\}\right\}<R\right)\le\epsilon. \tag{93}$$

Actually, we can improve upon (93), the same it was done in proposition 13:

$$P\left(\max_{Q\in\mathcal{P},\{0\le q_i\le C_i\}}\left\{\min_{\substack{\cup_{j=1}^r\mathcal{Z}_j\subseteq\{1,\ldots,r\},\\ i\neq j:\ \mathcal{Z}_j\cap\mathcal{Z}_i=\phi}}\left\{\sum_{j=1}^r G(\mathcal{Z}_j,q_{\mathcal{Z}_j})+\sum_{i\in\cap_{j=1}^r\mathcal{Z}_j^c}[C_i-q_i]\right\}\right\}<R\right)\le\epsilon, \tag{94}$$

but since the problem is not symmetric (due to the non-ergodic H), we can not further simplify it, as in Corollary 6. However, the limiting behavior of (87) is true also for the block fading case. Thus the optimality of the CEO approach when $t\to\infty$ from correlation 7 is assured for the block fading case as well.

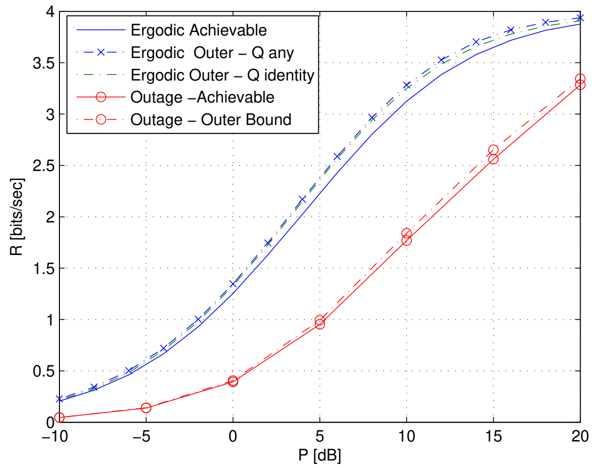

Fig. 3. The achievable rates compared to the upper bounds over a 2×2 system with $C=2$: for fast fading Rayleigh channel with upper bound according to an arbitrary Q (Q singular) and to a fix $Q=\frac{P}{t}I_t$ (Q identity), and for block fading Rayleigh channel, with outage probability of 10^{-2}, where the upper bound was calculated from (94). All as a function of P in dB, where the outage probability and the average over H were done by Monte Carlo simulations over H.

C. Discussion

When considering the upper bound, several clarifications are in order. It is known [15],[1] that when no fading is present, and the transmitter has only a single antenna, the upper bound is in fact tight. It means that when the sum $\sum_{i=j}^{r} Y_j$ is sufficient statistics, the capacity is established. This situation changes when considering fading channels. It is evident from [27], that when $Y_1 - Y_2$ is sufficient statistics, using our technique, which is based on the Berger-Tung CEO, is strictly sub-optimal and lattice approach can outperform the random binning. Therefore, it is not expected that ultimate performance is achieved, although the upper bound proximity to the achievable rate.

VI. Numerical Example

The achievable rates and the upper bounds for both fast fading and block fading channels, were calculated for a 2×2 system, with $C_1 = C_2 = 2$, for several signal to noise ratios (P in dB), and the results are presented in figure 3. For the fast fading, both achievable rate and upper bound are obtained by averaging over 30 blocks, each containing 50 channel realizations (the expectation expressed by E_H in (86) and (34)). It is seen there that the upper bound is convex, and that it is close to the achievable rate, when using CEO compression. For the lower and higher P the bound is tighter.

For block fading channel, the upper bound from (94) is depicted along with the achievable rate (48), for outage probability of $\epsilon = 10^{-2}$. The probability was calculated using Monte Carlo simulations over 10000 different realizations of H. It is seen there that the bound is again very tight for the low SNR region, and the gap becomes higher, with larger SNR, although it remains rather small, no more than 1 dB throughout the figure.

VII. Agents with Code Knowledge, and Fully Informed Transmitter

In this section we consider the same model, as in the previous sections, with two differences. One difference is that we drop the nomadity assumption, and let the agents be able to decode messages. The second difference is that we assume full CSI ($\boldsymbol{H}$) at the transmitter, in a non casual sense, so that the transmitter and the agents have the same channel state information.

We get to the following proposition, which is proved in the Appendix.

Proposition 15: In the ergodic regime, when the transmitter has full CSI, and the agents are cognizant of the codebook used, the rate (95) is achievable

$$R_{cog} = \max_{\pi} \min_{\mathcal{S}} \left\{ \sum_{i\in\mathcal{S}} [C_i - I(U_i;Y_i|X,W^r,H)] + I(U_{\mathcal{S}^C};X|W^r,H) + \sum_{i\in\mathcal{S}^C} [I(W_i;Y_i|H) - I(W_i;W_{\tilde{\mathcal{T}}(\pi,i)}|H)] \right\}, \tag{95}$$

where π *is a permutation of* $[1,\dots,r]$,

$$\tilde{\mathcal{T}}(\pi,i) \triangleq \{\pi_1,\dots,i\}, \tag{96}$$

and

$$P_{W^r,X,Y^r,U^r|H}(w^r,x,y^r,u^r|h) = P_{W^r|H}(w^r|h)P_{X|W^r,H}(x|w^r,h)\prod_{i=1}^{r}[P_{Y_i|X,H}(y_i|x,h)P_{U_i|Y_i,W_i,H}(u_i|y_i,w_i,h)]. \tag{97}$$

The transmitter sends messages to the agents via the broadcast channel [6], by using the dirty paper coding (DPC) technique [28]. On top, the transmitter also sends information to be decoded only at the final destination, invoking the nomadic techniques of the previous scheme. We actually extended the results of [1], to include also DPC and a random ergodic channel. In [1] Corollary 4, the superposition coding combined with the CEO technique, was used for that setting, when no fading was present, and when the channel was degraded. The main difference between superposition coding and DPC is in that superposition coding lets the destined terminal cancel the interfering transmissions (which are destined to terminals with weaker channels) and the DPC performs precoding, so that interference transmissions are canceled at the transmitter (thus the name dirty paper coding).

Next, for the fading Gaussian channel, the combined final destination decoding and DPC results with the rate

$$R_{DPC,1} = \max_{Q,\pi,\{B_i,q_i\}_{i=1}^r} \min_{\mathcal{S}} \mathrm{E}_H \left\{ \sum_{i\in\mathcal{S}} [C_i - q_i] + \log_2 \left| I_{|\mathcal{S}^C|} + \mathrm{diag}(1-2^{-q_i}) H_{\mathcal{S}^C} \left(Q - \sum_{i=1}^r B_i \right) H^*_{\mathcal{S}^C} \right| + \sum_{i\in\mathcal{S}^C} \log_2 \left(\frac{1 + H_i \left(Q - \sum_{j\in\tilde{\mathcal{T}}(\pi,i)} B_j \right) H_i^*}{1 + H_i \left(Q - \sum_{j\in\tilde{\mathcal{T}}(\pi,i)} B_j - B_i \right) H_i^*} \right) \right\}, \tag{98}$$

where the maximization is over

$$q_i \quad : \quad \mathbb{C}^{[r\times t]} \to \mathbb{R}_+, \tag{99}$$

$$Q, B_i \quad : \quad \mathbb{C}^{[r\times t]} \to \mathbb{C}^{[t\times t]}, \tag{100}$$

such that $Q, B_i \succeq 0$, $Q - \sum_1^r B_i \succeq 0$ and $\mathrm{E}_H[\mathrm{trace}(Q)] \le P$. The rate in (98) can be increased by convex hull [28], since in general, this problem is non concave.

This rate is achieved by using W_i, as in [28], and then $P_{U_i|Y_i,W_i,H} = P_{U_i|Y_i,H}$ remains the same as in Proposition 1. The situation in the compression stage, is similar to when using Wyner-Ziv source compression with decoder side information over Gaussian sources, where supplying the side information (W_i) to the encoder does not improve the rate distortion.

Although calculating (98) is hard, due to the non-convexity of the problem, note that a sub-optimal rate can be calculated for the symmetric case ($C_i = C$), by using the DPC such that the maximal sum-rate is obtained, and so that $Q = I_t \frac{P'}{t} - \sum_{j=1}^{r} B_j \succeq 0$, and letting $\mathrm{E}_H[P'] \leq P$. Since the problem is symmetric and the channel ergodic, each agent decodes the same rate. The DPC sum-rate can be obtained by the dual multi-access (concave) MIMO channel [29].

VIII. Conclusion

In this paper we showed the effectiveness of several compression techniques for decentralized reception in fast fading and block fading MIMO channels. We proved that in many cases, the elementary compression is sufficient to get the full-multiplexing gain. In addition, we showed the advantages of the CEO approach, which were evident in an asymptotic analysis and in a finite example. We presented upper-bounds for both fast fading channel and block fading channel, which are based on the nomadic characteristic of the scheme, along with the EPI, and which turned out to be quite tight even for relatively small 2×2 scheme. Achievable rate for a non-nomadic scheme was finally derived, combining the decentralized processing with the DPC.

Acknowledgment

This research was supported by the EU 6th framework program via the NEWCOM network of excellence.

Appendix I
Useful definitions and Lemmas

Let $P_{A_1,A_2,\ldots,A_L}(a_1,a_2,\ldots,a_L)$ be the probability function of the random variables $A_1,\ldots,A_L$ which take values in $\mathcal{A}_1,\ldots,\mathcal{A}_L$, respectively.

Definitions:

1) The marginal probabilities are then defined as

$$P_{A_l}(a_l) = \sum_{a_{\mathcal{L}\setminus l} \in \mathcal{A}_{\mathcal{L}\setminus l}} P_{A_1,A_2,\ldots,A_L}(a_1,a_2,\ldots,a_L) \tag{101}$$

($\mathcal{L}$ is the set $\{1,\ldots,L\}$).

2) The conditional probabilities are defined as:

$$P_{A_l|A_\mathcal{S}}(a_l|a_\mathcal{S}) = \frac{P_{A_l,A_\mathcal{S}}(a_l,a_\mathcal{S})}{P_{A_\mathcal{S}}(a_\mathcal{S})}, \tag{102}$$

for some $\mathcal{S} \subseteq \mathcal{L}$ and $l \notin \mathcal{S}$ and $P_{A_\mathcal{S}}(a_\mathcal{S}) \neq 0$.

3) As commonly done (see [23], section 13, problem 10), define the ϵ-typical (strongly conditional typical) set $\mathbf{T}_\epsilon$ of $\boldsymbol{a}_\mathcal{L}$ as the set for which $N(a_\mathcal{S}, h|\boldsymbol{a}_\mathcal{S}, \boldsymbol{h}) = 0$ for any $a_\mathcal{S} \in \mathcal{A}_\mathcal{S}, \boldsymbol{h} \in \mathcal{H}$ such that $P_{A_\mathcal{S}|H}(a_\mathcal{S}|h) = 0$, and also

$$\mathbf{T}_\epsilon(\boldsymbol{h}) \triangleq \left\{\boldsymbol{a}_\mathcal{L} : \ \forall \mathcal{S} \subseteq \mathcal{L}, \ \forall a_\mathcal{S} \in \mathcal{A}_\mathcal{S}, h \in \mathcal{H} \ \frac{1}{n} \left|N(a_\mathcal{S}, h|\boldsymbol{a}_\mathcal{S}, \boldsymbol{h}) - P_{A_\mathcal{S}|H}(a_\mathcal{S}|h) N(h|\boldsymbol{h})\right| < \frac{\epsilon}{|\mathcal{A}_\mathcal{S}|}\right\}, \tag{103}$$

where $N(a_\mathcal{S}|\boldsymbol{a}_\mathcal{S})$ denotes the counting operator of the number of occurrences of the symbol $a_\mathcal{S}$ in the vector $\boldsymbol{a}_\mathcal{S}$.

Lemmas:

Lemma 2: For any $\epsilon > 0$, there exist n^* such that for all $n > n^*$ and randomly generated $\boldsymbol{a}_\mathcal{L}$ according to $\prod P_{A_\mathcal{L}|H}(a_\mathcal{L}(k)|h(k))$

$$\Pr\{\boldsymbol{a}_\mathcal{L} \in \mathbf{T}_\epsilon(\boldsymbol{h})\} \geq 1 - \epsilon. \tag{104}$$

Lemma 3: Fix some $\mathcal{S} \subseteq \mathcal{L}$ and probability

$$P_{A_\mathcal{L}, W_\mathcal{L}|H}(a_\mathcal{L}, w_\mathcal{L}|h). \tag{105}$$

Define the jointly ϵ-typical set $\mathbf{T}_\epsilon(\boldsymbol{h})$, as before, by the joint probability (105).

Let $\boldsymbol{a}_\mathcal{L}^n$ be generated according to

$$\boldsymbol{a}_\mathcal{L} \sim \prod_{k=1}^{n} \left\{P_{A_{\mathcal{S}^C}|W_\mathcal{L}, H}(a_{\mathcal{S}^C}(k)|w_\mathcal{L}(k), h(k)) \prod_{l \in \mathcal{S}} P_{A_l|W_l, H}(a_l(k)|w_l(k), h(k))\right\}, \tag{106}$$

where the conditional and marginal probabilities are calculated from (105) and $\boldsymbol{w}_\mathcal{L}$ is a given vector which was randomly generated and that belongs to the set $\mathbf{T}_\epsilon(\boldsymbol{h})$, as defined by (103) (that is, there exist $\boldsymbol{a}_\mathcal{L}$ that are jointly typical with $\boldsymbol{w}_\mathcal{L}$).

Then the probability of the vector $\boldsymbol{a}_\mathcal{L}$ which is distributed according to (106) to be in $\mathbf{T}_\epsilon(\boldsymbol{h})$, which is defined according to (105) is bounded by:

$$\Pr\{(\boldsymbol{a}_{1,\ldots,L}, \boldsymbol{w}_\mathcal{L}) \in \mathbf{T}_\epsilon(\boldsymbol{h})\} \geq 2^{-n[H(A_{\mathcal{S}^C}|W_\mathcal{L}, H) - H(A_\mathcal{L}|W_\mathcal{L}, H) + \sum_{l \in \mathcal{S}} H(A_l|W_l, H) + \epsilon_1]} \tag{107}$$

$$\Pr\{(\boldsymbol{a}_{1,\ldots,L}, \boldsymbol{w}_\mathcal{L}) \in \mathbf{T}_\epsilon(\boldsymbol{h})\} \leq 2^{-n[H(A_{\mathcal{S}^C}|W_\mathcal{L}, H) - H(A_\mathcal{L}|W_\mathcal{L}, H) + \sum_{l \in \mathcal{S}} H(A_l|W_l, H) - \epsilon_1]} \tag{108}$$

where $\epsilon_1 \to 0$ as $\epsilon \to 0$.

Lemma 4: Generalized Markov Lemma

Let

$$P_{A_\mathcal{S}, W_\mathcal{S}, Y_\mathcal{S}|H}(a_\mathcal{S}, w_\mathcal{S}, y_\mathcal{S}|h) = P_{W_\mathcal{S}, Y_\mathcal{S}|H}(w_\mathcal{S}, y_\mathcal{S}|h) \prod_{l \in \mathcal{S}} P_{A_t|W_t, Y_t, H}(a_t|w_t, y_t, h). \tag{109}$$

Given randomly generated $\boldsymbol{w}_\mathcal{S}\boldsymbol{y}_\mathcal{S}$ according to $P_{W_\mathcal{S}, Y_\mathcal{S}|H}$, for every $i \in \mathcal{S}$, randomly and independently generate $N_i \geq 2^{nI(A_i;Y_i|W_i,H)}$ vectors $\tilde{\boldsymbol{a}}_i$ according to $\prod_{k=1}^{n} P_{A_i|W_i,H}(\tilde{a}_i(k)|w_i(k), h(k))$, and index them by $\tilde{\boldsymbol{a}}_i^{(t)}$ $(1 \leq$

$t \leq N_i$). Then there exist $|\mathcal{S}|$ functions $t_i^* = \phi_i(\boldsymbol{y}_i, \boldsymbol{w}_i, \tilde{\boldsymbol{a}}_i^{(1)}, \ldots, \tilde{\boldsymbol{a}}_i^{(N_t)})$ taking values in $[1 \ldots N_t]$, such that for sufficiently large n,

$$\Pr((\{\boldsymbol{a}_i^{(t_i^*)}\}_{i\in\mathcal{S}}, \boldsymbol{w}_\mathcal{S}, \boldsymbol{y}_\mathcal{S}) \in \mathbf{T}_\epsilon(\boldsymbol{h})) \geq 1 - \epsilon. \quad (110)$$

Proof: See [23] and [30] for the proofs of Lemmas 2-3, while Lemma 4 is a simple extension of Lemma 3.4 (Generalized Markov Lemma) in [31]. ■

In the following, we use only ϵ and remove the distinction between ϵ and ϵ_1, for the sake of brevity.

APPENDIX II
PROOF OF PROPOSITION 1

A. Code construction:

Fix $\delta > 0$.

1) For the transmitter, for any codebook used, f
   - Randomly choose $2^{nR_{CEO}}$ vectors $\boldsymbol{x}$, with probability $P_\mathbf{X}(\boldsymbol{x}) = \prod_k P_X(x(k))$.
   - Index these vectors by M_{CEO} where $M_{CEO} \in [1, 2^{nR_{EC}}]$.
2) For the compressor at the agents

   For every channel realization $\boldsymbol{h}$
   - Randomly generate 2^{nC_i} vectors $\boldsymbol{u}_i$ of length n according to $\prod_k P_{U_i|H}(u_i(k)|h(k))$.
   - Index all the generated $\boldsymbol{u}_i$ with $z_i \in [1, 2^{nC_i}]$.

B. Encoding:

Let M be the message to be sent, and f is the codebook used. The transmitter then sends $\boldsymbol{x}(M, f)$ to the channel.

C. Processing at the agents:

The i^{th} agent chooses any of the z_i such that

$$\big(\boldsymbol{u}_i(z_i, \boldsymbol{h}), \boldsymbol{y}_i\big) \in \mathbf{T}_\epsilon^{EC,i}(\boldsymbol{h}), \quad (111)$$

where $\mathbf{T}_\epsilon^{EC,i}(\boldsymbol{h})$ is defined in the standard way, as (103). The event where no such z_i is found is defined as the error event E_1.

After deciding on z_i the agent forwards it to the final destination through the lossless link.

D. Decoding (at the destination):

The destination retrieves z^r from the lossless links, and uses $\boldsymbol{h}$ and the random encoding f.

The destination then finds $\hat{M}$ such that

$$\big(\boldsymbol{x}(\hat{M}, f), \boldsymbol{u}^r(\hat{z}^r)\big) \in \mathbf{T}_\epsilon^{EC,3}(\boldsymbol{h}). \quad (112)$$

Where $\mathbf{T}_{\epsilon}^{EC,3}$ is defined in the standard way, as (103). If there is no such $\hat{M}$, or if if there is more than one, the destination chooses one arbitrarily. Define error E_2 as the event where $\hat{M} \neq M_{CEO}$.

Correct decoding means that the destination decides $\hat{M} = M$. An achievable rate R was defined as when the final destination receives the transmitted message with an error probability which is made arbitrarily small for sufficiently large block length n.

E. Error analysis

The error probability is upper bounded by:

$$\Pr\{\text{error}\} = \Pr\left(\cup_{i=1}^{2} E_i\right) \leq \sum_{i=1}^{2} \Pr(E_i). \tag{113}$$

Where:

1) E_1: No $\boldsymbol{u}_i(z_i, \boldsymbol{h})$ is jointly typical with $\boldsymbol{y}_i$.
2) E_2: Decoding error $\boldsymbol{x}(\hat{M}, f) \neq \boldsymbol{x}(M, f)$, so that $\hat{M} \neq M$.

Next, we will upper bound the probabilities of the individual error events by arbitrarily small ϵ.

1) E_1: According to Lemma 4, the probability $\Pr\{E_1\}$ can be made as small as desired, for n sufficiently large, as long as

$$C_i > I(U_i; Y_i | H). \tag{114}$$

2) E_2: Consider the case where $\hat{M} \neq M$. There are $2^{nR_{CEO}}$ such vectors, and the probability of $(\boldsymbol{x}(\hat{M}, f), \boldsymbol{u}^r(z^r))$ to be jointly typical is upper bounded by (Lemma 3) $2^{-n[I(X;U^r|H)-\epsilon]}$. Thus the rate R_{CEO} is achievable if:

$$R_{CEO} < I(X; U^r | H) - \epsilon, \tag{115}$$

which proves Proposition 1. ■

Appendix III
Proof of Optimality of $Q = \frac{P}{t} I_t$ for the Ergodic Channel.

First consider that since the channel is unknown to the transmitter, and VH is distributed as H when V is unitary (eigenvectors of a non diagonal Q) all through this work, Q can be limited to be diagonal.

Next, for any given $q_i(H)$ and $\mathcal{S}$, we have that

$$\mathrm{E}_H \left[\log_2 \det \left(I_{|\mathcal{S}|} + \operatorname{diag}\left(1 - 2^{-q_i(H)}\right)_{i \in \mathcal{S}} H_{\mathcal{S}} Q H_{\mathcal{S}}^* \right) \right] \tag{116}$$

is a concave function of Q, which is thus maximized by $Q = \frac{P}{t} I_t$ [2]. Thus it also maximizes the maximum over all $q_i(H)$ and $\mathcal{S}$ concluding the proof. ■

Notice that this proof does not extend to (10) and to (35), so that there, the optimal Q may not be proportional to identity, but is still diagonal, though.

APPENDIX IV
PROOF FOR PROPOSITION 7

In this Appendix, we give a closed solution to the $r = 2$, symmetric case. We extend what was done in [1] to the ergodic channel case, with $t > 1$. Equation (34) for the symmetric case can be written as:

$$R_{CEO} = \max_{0 \le q^* \le C} \left\{ \min_{\mathcal{S} \subseteq \{1,\ldots,r\}} \{ |\mathcal{S}^C| [C - q^*] + F_{\mathcal{S}}(q^*) \} \right\}, \tag{117}$$

where

$$F_{\mathcal{S}}(q^*) = \max_{\{q_i : \mathbb{C}^{[r \times t]} \to \mathbb{R}_+\}_{i=1}^r} \mathrm{E}_H \log_2 \det \left(I_{|\mathcal{S}|} + \frac{P}{t} \mathrm{diag} \left(1 - 2^{-q_i(H)} \right)_{i \in \mathcal{S}} H_{\mathcal{S}} H_{\mathcal{S}}^* \right) \tag{118}$$

such that

$$\mathrm{E}_H[q_i(H)] = q^*, \; i = 1, \ldots, r. \tag{119}$$

Since the channel is ergodic, and the scheme symmetric, the users will be equivalent, and due to the concavity of the problem, the optimal solution is characterized by $q^* = \mathrm{E}_H[r_i(H)]$. That is, equal bandwidth that is wasted by all users on the noise quantization. By writing the equation this way, the ergodic nature of the channel is used, such that the channel randomness is limited to within $F_{\mathcal{S}}$. Since $F_{\mathcal{S}}$ is an increasing function of q^*, when solving it, the solution of (34) is readily available numerically. So we are left with the concave problem of finding $F_{\mathcal{S}}$.

Since $F_{\{1,\ldots,r\}}(q^*)$ is an increasing function of q^*, and $r(C - q^*)$ is a decreasing function of q^*, the point $F_{\{1,\ldots,r\}}(q^*) = r(C - q^*)$ exists, and further, it is an upper bound to the achievable rate. Next, using Hadamard inequality we have that for any $\mathcal{S}$

$$\begin{aligned} \log_2 \det \left(I_2 + \frac{P}{t} \mathrm{diag}(1 - 2^{-q_i})_{i=1}^r H H^* \right) \le \log_2 \det \left(I_{\mathcal{S}} + \frac{P}{t} \mathrm{diag}(1 - 2^{-q_i})_{i \in \mathcal{S}} H_{\mathcal{S}} H_{\mathcal{S}}^* \right) \\ + \log_2 \det \left(I_{\mathcal{S}^C} + \frac{P}{t} \mathrm{diag}(1 - 2^{-q_i})_{i \in \mathcal{S}^C} H_{\mathcal{S}^C} H_{\mathcal{S}^C}^* \right). \end{aligned} \tag{120}$$

Since the channel is ergodic, the minimum in (117) is over functionals of the channel probability, rather then channel realizations. In addition, the channel probability is symmetric with regards to the agents, leading to $F_{\{1,\ldots,r\}}(q^*)$, which is the minimum among all the subsets $\mathcal{S}$. So that the achievable rate can be calculated by solving the following problem

$$\max_{\{q_i : \mathbb{C}^{[r \times t]} \to \mathbb{R}_+\}_{i=1}^r} \mathrm{E}_H \log_2 \det \left(I_2 + \frac{P}{t} \mathrm{diag} \left(1 - 2^{-q_i(H)} \right)_{i=1}^r H H^* \right) \tag{121}$$

such that $q_i(H) \ge 0$ and

$$\mathrm{E}_H[q_i(H)] = q^*, \; i = 1, \ldots, r. \tag{122}$$

Let us limit the discussion to the case of $r = 2$. The solution can be obtained through Lagrange multipliers, as follows ($\theta \ge 0$)

$$\nabla \log_2 \det \left(I_2 + \frac{P}{t} \mathrm{diag} \left(1 - 2^{-q_i(H)} \right)_{i=1,2} H H^* \right) - \theta I_2 = \mu(H). \tag{123}$$

So for any $\mu_i(H) = 0$, such that $q_i(H) > 0$, we get that ($\bar{i} = 3 - i$)

$$i = 1, 2: \;\; \frac{2^{-q_i}(\Delta_{2+i} - 2^{-q_{\bar{i}}} \Delta_2)}{\Delta} = \theta, \tag{124}$$

and $\mathrm{E}_H(q_i) = q^*$, where

$$\begin{aligned}
\Delta &\triangleq \det\left(I_2 + \frac{P}{t}\mathrm{diag}\left(1 - 2^{-q_i(H)}\right)_{i=1,2} HH^*\right) && (125)\\
\Delta_1 &\triangleq \det\left(I_2 + \frac{P}{t}HH^*\right) && (126)\\
\Delta_2 &\triangleq \det\left(\frac{P}{t}HH^*\right) && (127)\\
\Delta_3 &\triangleq \det\left(\mathrm{diag}([0,1]) + \frac{P}{t}HH^*\right) && (128)\\
\Delta_4 &\triangleq \det\left(\mathrm{diag}([1,0]) + \frac{P}{t}HH^*\right). && (129)
\end{aligned}$$

We note that (124) determines a one-to-one connection between θ and q^*. In addition, note that

$$\Delta = \Delta_1 + 2^{-q_1-q_2}\Delta_2 - 2^{-q_1}\Delta_3 - 2^{-q_2}\Delta_4,$$

and that

$$\Delta_3 = \Delta_2 + \frac{P}{t}|H_1|^2 \quad (130)$$

$$\Delta_4 = \Delta_2 + \frac{P}{t}|H_2|^2. \quad (131)$$

The solution of (124) is

$$q_i = -\log_2\left(\frac{\Delta_{\bar{i}+2}}{2\Delta_2(1+\theta)}\left((1+2\theta) - \sqrt{(1+2\theta)^2 - 4\theta(1+\theta)\frac{\Delta_1\Delta_2}{\Delta_3\Delta_4}}\right)\right). \quad (132)$$

We note that $\frac{\Delta_1\Delta_2}{\Delta_3\Delta_4} \leq 1$ with equality if and only if HH^* is a diagonal matrix. So the square root in equation (132) is guaranteed to be positive real. By a simple derivative, it is easily verified that $F_H(\theta)$, defined by (38), is monotonically increasing with θ.

Then, in case any of $q_i,\ i = 1,2$ from (132) turns out negative (say $F_H(\theta) > \frac{\Delta_2}{\Delta_{2+i}}$ which leads to $q_{\bar{i}} < 0$), then the solution is $q_{\bar{i}} = 0$ and q_i is equal to

$$q_i = -\log_2\left(\frac{\theta}{1+\theta}\frac{1+\frac{P}{t}|H_i|^2}{\frac{P}{t}|H_i|^2}\right). \quad (133)$$

If (133) is negative as well, the solution is $q_i = 0$. As θ gets smaller, more channels will result with (132) solved with $q_i > 0$, which means better compression.

Overall, we can write

$$q_1(H,\theta) = \begin{cases} \left\lceil -\log_2\left(\frac{\theta}{1+\theta}\frac{1+\frac{P}{t}|H_1|^2}{\frac{P}{t}|H_1|^2}\right)\right\rceil^+ & F_H(\theta) > \frac{\Delta_2}{\Delta_3} \\ \left\lceil -\log_2\left(\frac{\Delta_4}{\Delta_2}F_H(\theta)\right)\right\rceil^+ & F_H(\theta) \leq \frac{\Delta_2}{\Delta_3}. \end{cases} \quad (134)$$

Now θ is determined by the equation

$$\mathrm{E}_H \log_2 \det\left(I_2 + \frac{P}{t}\mathrm{diag}\left(1 - 2^{-q_i(H,\theta)}\right)_{i=1}^2 HH^*\right) = 2(C - \mathrm{E}_H[q_i(H,\theta)]) \quad (135)$$

and the achievable rate is

$$R_{CEO} = 2(C - \mathrm{E}_H[q_i(H,\theta)]). \quad (136)$$

This concludes the proof. ■

APPENDIX V
PROOF OF LEMMA 1

The proof is divided into two sections, we start by proving for the case where $|\mathcal{S}| \leq t$. This division is since the first case is easier to show, and thus gives better understanding of the guidelines and techniques, which are identical, albeit more involved, for the case of $|\mathcal{S}| \geq t$.

For the sake of the proof, define:

- $Z \triangleq H_{\mathcal{S}} X$, where $I(\boldsymbol{Y}_{\mathcal{S}}; \boldsymbol{X}|\boldsymbol{H}) = I(\boldsymbol{Y}_{\mathcal{S}}; \boldsymbol{Z}|\boldsymbol{H})$.
- $\Lambda_z \triangleq \mathrm{E}[ZZ^*] = H_{\mathcal{S}} Q H_{\mathcal{S}}^* = \frac{P}{t} H_{\mathcal{S}} H_{\mathcal{S}}^*$ (equal to $\Lambda_{\mathcal{S}}$).
- $\hat{Z} \triangleq AY$, where A is the best estimator of Z from Y, calculated as $A = \Lambda_z (I + \Lambda_z)^{-1}$.

Since $|\mathcal{S}| \leq t$ we have that $|\Lambda_z| > 0$. Note that since $\hat{Z}$ is the best estimator

$$Z = \hat{Z} + \hat{N} \tag{137}$$

where $\hat{Z}$ and $\hat{N}$ are independent, and since $\mathrm{E}[\hat{Z}\hat{Z}^*] = \Lambda_z (I + \Lambda_z)^{-1} \Lambda_z$, we get $\mathrm{E}[\hat{N}\hat{N}^*] = \Lambda_z (I + \Lambda_z)^{-1}$. Now we can rely on the independence in (137) and the vector entropy power inequality:

$$2^{\frac{1}{n|\mathcal{S}|} h(\boldsymbol{Z}|V_{\mathcal{S}}, \boldsymbol{H}=\boldsymbol{h})} \geq 2^{\frac{1}{n|\mathcal{S}|} h(\hat{\boldsymbol{Z}}|V_{\mathcal{S}}, \boldsymbol{H}=\boldsymbol{h})} + (\pi e) \prod_{k=1}^{n} \left(\frac{|\Lambda_z(k)|}{|I + \Lambda_z(k)|} \right)^{\frac{1}{n|\mathcal{S}|}}. \tag{138}$$

Next we express the required quantity $\lambda \triangleq \frac{1}{n} I(\boldsymbol{Z}; V_{\mathcal{S}}|\boldsymbol{H} = \boldsymbol{h})$ in both sides of (138). For the left hand side,

$$\frac{1}{n|\mathcal{S}|} h(\boldsymbol{Z}|V_{\mathcal{S}}, \boldsymbol{H} = \boldsymbol{h}) = \frac{1}{n|\mathcal{S}|} h(\boldsymbol{Z}|\boldsymbol{H} = \boldsymbol{h}) - \frac{\lambda}{|\mathcal{S}|} = \frac{1}{n|\mathcal{S}|} \log_2 \left(\prod^{n} |\Lambda_z(k)| \right) + \log_2(\pi e) - \frac{\lambda}{|\mathcal{S}|}. \tag{139}$$

The right hand side is more elaborated, and will be done in two stages. First note that:

$$h(\hat{\boldsymbol{Z}}|V_{\mathcal{S}}, \boldsymbol{H} = \boldsymbol{h}) = h(\hat{\boldsymbol{Z}}|\boldsymbol{Z}, V_{\mathcal{S}}, \boldsymbol{H} = \boldsymbol{h}) + I(\boldsymbol{Z}; \hat{\boldsymbol{Z}}|V_{\mathcal{S}}, \boldsymbol{H} = \boldsymbol{h}). \tag{140}$$

We know that $h(\boldsymbol{Z}|\hat{\boldsymbol{Z}}, \boldsymbol{H} = \boldsymbol{h}) = h(\boldsymbol{Z}|\hat{\boldsymbol{Z}}, V_{\mathcal{S}}, \boldsymbol{H} = \boldsymbol{h})$, from the definition of $\hat{Z}$ and $V_{\mathcal{S}}$. This means that:

$$\frac{1}{n} I(\boldsymbol{Z}; \hat{\boldsymbol{Z}}|V_{\mathcal{S}}, \boldsymbol{H} = \boldsymbol{h}) = \frac{1}{n} I(\boldsymbol{Z}; \hat{\boldsymbol{Z}}|\boldsymbol{H} = \boldsymbol{h}) - \lambda = \frac{1}{n} \log_2 \left(\prod^{n} |I + \Lambda_z(k)| \right) - \lambda. \tag{141}$$

Second, we have that $\hat{Z} = AY_{\mathcal{S}}$, so

$$h(\hat{\boldsymbol{Z}}|\boldsymbol{Z}, V_{\mathcal{S}}, \boldsymbol{H} = \boldsymbol{h}) = h(\boldsymbol{Y}_{\mathcal{S}}|\boldsymbol{Z}, V_{\mathcal{S}}, \boldsymbol{H} = \boldsymbol{h}) + 2\log_2 \left(\prod^{n} |A(k)| \right) =$$
$$\sum_{i \in \mathcal{S}} h(\boldsymbol{Y}_i|\boldsymbol{Z}, V_i, \boldsymbol{H} = \boldsymbol{h}) + 2\log_2 \left(\prod^{n} \frac{|\Lambda_z(k)|}{|I + \Lambda_z(k)|} \right). \tag{142}$$

define $q_i(\boldsymbol{h}) \triangleq \frac{1}{n} I(\boldsymbol{Y}_i; V_i|\boldsymbol{X}, \boldsymbol{H} = \boldsymbol{h})$ and since we used additive noise with unit variance,

$$\frac{1}{n} h(\boldsymbol{Y}_i|\boldsymbol{Z}, V_i, \boldsymbol{H} = \boldsymbol{h}) = \log_2(\pi e) - q_i(\boldsymbol{h}). \tag{143}$$

rewrite (142) as

$$\frac{1}{n} h(\hat{\boldsymbol{Z}}|\boldsymbol{Z}, V_{\mathcal{S}}, \boldsymbol{H} = \boldsymbol{h}) = |\mathcal{S}| \log_2(\pi e) - \sum_{i \in \mathcal{S}} q_i(\boldsymbol{h}) + \frac{2}{n} \log_2 \left(\prod^{n} \frac{|\Lambda_z(k)|}{|I + \Lambda_z(k)|} \right). \tag{144}$$

Now using (141) and (144) in the right hand side, written in (140), we get to:

$$2^{\frac{1}{n|\mathcal{S}|}h(\hat{\boldsymbol{Z}}|V_{\mathcal{S}},\boldsymbol{H}=\boldsymbol{h})} = \pi e \prod_{i\in\mathcal{S}} 2^{-\frac{q_i(\boldsymbol{h})}{|\mathcal{S}|}} \left(\prod^{n} \frac{|\Lambda_z(k)|}{|I+\Lambda_z(k)|}\right)^{\frac{2}{n|\mathcal{S}|}} \left(\prod^{n} |I+\Lambda_z(k)|\right)^{\frac{1}{n|\mathcal{S}|}} 2^{-\frac{\lambda}{|\mathcal{S}|}}. \tag{145}$$

Finally we combine left hand side (139) and right hand side (145), and get

$$\pi e 2^{-\frac{1}{|\mathcal{S}|}\lambda}\left(\prod^{n}|\Lambda_z(k)|\right)^{\frac{1}{n|\mathcal{S}|}} \geq \pi e 2^{-\frac{\lambda}{|\mathcal{S}|}} \prod_{i\in\mathcal{S}} 2^{-\frac{q_i(\boldsymbol{h})}{|\mathcal{S}|}} \left(\prod^{n} \frac{|\Lambda_z(k)|}{|I+\Lambda_z(k)|}\right)^{\frac{2}{n|\mathcal{S}|}} \left(\prod^{n} |I+\Lambda_z(k)|\right)^{\frac{1}{n|\mathcal{S}|}} + \pi e \prod_{k=1}^{n}\left(\frac{|\Lambda_z(k)|}{|I+\Lambda_z(k)|}\right)^{\frac{1}{n|\mathcal{S}|}}. \tag{146}$$

Reordering the equation we get to (73), which proves Lemma 1 for when $|\mathcal{S}| \leq t$.

We continue to the case where $|\mathcal{S}| > t$, where we have more agents than transmitters, so that $|\Lambda_z| = 0$. Like in the previous setting we define $\hat{X} = AY$ to be the best estimator of X out of Y. So that now $A = QH^*(I+\Lambda_z)^{-1}$, and we have

$$X = \hat{X} + \hat{N}, \tag{147}$$

where $\hat{X}$ and $\hat{N}$ are independent and using the matrix inversion Lemma $\mathrm{E}[\hat{N}\hat{N}^*] = (Q^{-1} + H^*H)^{-1} = Q(I + QH^*H)^{-1}$. Again we use the entropy power inequality:

$$2^{\frac{1}{nt}h(\boldsymbol{X}|V_{\mathcal{S}},\boldsymbol{H}=\boldsymbol{h})} \geq 2^{\frac{1}{nt}h(\hat{\boldsymbol{X}}|V\mathcal{S},\boldsymbol{H}=\boldsymbol{h})} + \pi e |Q|^{\frac{1}{t}} \prod_{k=1}^{n}\left(\frac{1}{|I+\Lambda_z(k)|}\right)^{\frac{1}{nt}}. \tag{148}$$

Using the same argument as the one used for (139), the left hand side of (148) becomes

$$2^{\frac{1}{nt}h(\boldsymbol{X}|V_{\mathcal{S}},\boldsymbol{H}=\boldsymbol{h})} = \pi e |Q|^{\frac{1}{t}} 2^{-\frac{\lambda}{t}}. \tag{149}$$

The left expression in the right hand side of (148) can be written as the sum of two arguments, as in (140), where the right-most mutual information (like (141)) is

$$\frac{1}{n} I(\boldsymbol{X};\hat{\boldsymbol{X}}|V_{\mathcal{S}},\boldsymbol{H}=\boldsymbol{h}) = \frac{1}{n} I(\boldsymbol{X};\hat{\boldsymbol{X}}|\boldsymbol{H}=\boldsymbol{h}) - \lambda = \frac{1}{n}\log_2\left(\prod^{n}|I+\Lambda_z(k)|\right) - \lambda. \tag{150}$$

The difference between the case where $|\mathcal{S}| < t$ and $|\mathcal{S}| > t$ is evident in the derivation of (142), which for $|\mathcal{S}| > t$ requires the double use of the entropy power inequality. So we want to lower bound $h(\hat{\boldsymbol{X}}|V_{\mathcal{S}},\boldsymbol{X},\boldsymbol{H}=\boldsymbol{h})$. First, let us decompose A using the singular value decomposition into $A = U_1 D U_2$, where $U_1 \in \mathbb{C}^{t\times t}$ and $U_2 \in \mathbb{C}^{|\mathcal{S}|\times|\mathcal{S}|}$ are two unitary matrices and $D \in \mathrm{R}^{t\times|\mathcal{S}|}$ is diagonal matrix. So we have that:

$$\begin{aligned} h(\hat{\boldsymbol{X}}|V_{\mathcal{S}},\boldsymbol{X},\boldsymbol{H}=\boldsymbol{h}) &= h(\boldsymbol{U}_1\boldsymbol{D}\boldsymbol{U}_2\boldsymbol{Y}|V_{\mathcal{S}},\boldsymbol{X},\boldsymbol{H}=\boldsymbol{h}) \\ &= \log_2\prod^{n}|U_1(k)|^2 + \sum_{j=1}^{t}[\log_2\prod^{n}|D_{j,j}(k)|^2 + h((\boldsymbol{U}_2)_j\boldsymbol{Y}|V_{\mathcal{S}},\boldsymbol{X},\boldsymbol{H}=\boldsymbol{h})], \end{aligned} \tag{151}$$

since U_2 is unitary matrix. Next we employ the entropy power inequality to lower bound $h((\boldsymbol{U}_2)_j\boldsymbol{Y}|V_{\mathcal{S}},\boldsymbol{X},\boldsymbol{H}=\boldsymbol{h})$:

$$2^{h((\boldsymbol{U}_2)_j\boldsymbol{Y}|V_{\mathcal{S}},\boldsymbol{X},\boldsymbol{H}=\boldsymbol{h})} \geq \sum_{i\in\mathcal{S}} 2^{h(\boldsymbol{Y}_i|V_{\mathcal{S}},\boldsymbol{X},\boldsymbol{H}=\boldsymbol{h})} \prod^{n}|(U_2(k))_{j,i}|^2. \tag{152}$$

This inequality is achieved with equality for Gaussian variables. A lower bound on (151) is given by

$$h(\hat{\boldsymbol{X}}|V_{\mathcal{S}},\boldsymbol{X},\boldsymbol{H}=\boldsymbol{h}) \geq \log_2\left(\prod^{n}|U_1(k)D(k)U_2(k)\mathrm{diag}(2^{-q_i(\boldsymbol{h})})_{i\in\mathcal{S}}U_2(k)^*D(k)^*U_1(k)^*|\right)+nt\log_2(\pi e)$$
$$=\log_2\left(\prod^{n}|QH(k)^*(I+\Lambda_z(k))^{-1}\mathrm{diag}(2^{-q_i(\boldsymbol{h})})_{i\in\mathcal{S}}(I+\Lambda_z(k))^{-1}H(k)Q|\right)+nt\log_2(\pi e)$$
$$=\log_2\left(\prod^{n}\left(\frac{|QH(k)^*\mathrm{diag}(2^{-q_i(\boldsymbol{h})})_{i\in\mathcal{S}}H(k)Q|}{|I+\Lambda_z(k)|^2}\right)\right)+nt\log_2(\pi e) \quad (153)$$

since

$$|QH^*(I+\Lambda_z)^{-1}D(I+\Lambda_z)^{-1}HQ|=\frac{|(I+QH^*H)QH^*(I+\Lambda_z)^{-1}D(I+\Lambda_z)^{-1}HQ(I+H^*HQ)|}{|I+\Lambda_z|^2}=\frac{|QH^*DHQ|}{|I+\Lambda_z|^2}. \quad (154)$$

To conclude, we use (149), (150) and (153):

$$\pi e|Q|^{\frac{1}{t}}2^{-\frac{\lambda}{t}} \geq \pi e 2^{-\frac{-\lambda}{t}}\left(\prod^{n}|I+\Lambda_z(k)|\frac{|Q|^2|H(k)^*\mathrm{diag}(2^{-q_i(\boldsymbol{h})})_{i\in\mathcal{S}}H(k)|}{|I+\Lambda_z(k)|^2}\right)^{\frac{1}{nt}}+\pi e|Q|^{\frac{1}{t}}\prod_{k=1}^{n}\left(\frac{1}{|I+\Lambda_z(k)|}\right)^{\frac{1}{nt}} \quad (155)$$

which by taking expectation with respect to $\boldsymbol{H}$, together with (146) proves Lemma 1. ■

APPENDIX VI
PROOF OF PROPOSITION 15

The proof of Proposition 15 is based on the proof of Theorem 3 from [1].

A. Code construction:

For every channel realization $\boldsymbol{h}$, determine the maximizing π. Fix $\delta > 0$ and then

1) For the broadcast transmissions, for every $i=\pi_1,\ldots,\pi_r$:
   - Randomly generate $2^{n[I(W_i;W_{\tilde{\mathcal{T}}(\pi,i)}|H)+\delta]}$ vectors $\boldsymbol{w}_i$, according to $P_{\boldsymbol{W}_i|\boldsymbol{H}}(\boldsymbol{w}_i|\boldsymbol{h})=\prod_{k=1}^{n}P_{W_i|H}(w_i(k)|h(k))$.
   - For every $\boldsymbol{w}_{\tilde{\mathcal{T}}(\pi,i)}$ generated in the previous iteration, find at least one $\boldsymbol{w}_i$ within the generated set which is jointly typical. Joint typicality means that

$$(\boldsymbol{w}_i,\boldsymbol{w}_{\tilde{\mathcal{T}}(\pi,i)})\in\mathbf{T}_{\epsilon}^{BC,i}(\boldsymbol{h}), \quad (156)$$

   where

$$\mathbf{T}_{\epsilon}^{BC,i}(\boldsymbol{h}) \triangleq \Bigg\{\boldsymbol{w}_{i,\tilde{\mathcal{T}}(\pi,i)}:\ \forall\mathcal{S}\subseteq\{i,\tilde{\mathcal{T}}(\pi,i)\},\ \forall w_{\mathcal{S}}\in\mathcal{W}_{\mathcal{S}}, h\in\mathcal{H}$$
$$\frac{1}{n}\left|N(w_{\mathcal{S}},h|\boldsymbol{w}_{\mathcal{S}},\boldsymbol{h})-P_{W_{\mathcal{S}}|H}(w_{\mathcal{S}}|\boldsymbol{h})N(h|\boldsymbol{h})\right|<\frac{\epsilon}{|\mathcal{W}_{\mathcal{S}}|}\Bigg\}. \quad (157)$$

   - In case no such vector exists, declare error event E_1.
   - Repeat the last steps for $2^{n[I(W_i;Y_i|H)-I(W_i;W_{\tilde{\mathcal{T}}(\pi,i)}|H)-\delta]}$ times.

Label the resulting vectors of each repetition, which were jointly typical, by M_i, where $M_i \in [1, 2^{n[I(W_i;Y_i|H) - I(W_i;W_{\tilde{\mathcal{T}}(\pi,i)}|H) - \delta]}]$. Then $M^r = \{M_1, \ldots, M_r\}$ and further define $\mathcal{M}_{M_i}$ as the set labeled by M_i. So that $\boldsymbol{w}^r(M^r, \boldsymbol{h})$ are the r vectors which were selected in the last stage and are jointly typical.

2) For the message which is decoded at the final destination, for every $\boldsymbol{w}^r$ defined by some M^r, and for every random encoding realization f
   - Randomly choose $2^{nR_{CEO}}$ vectors $\boldsymbol{x}$, with probability $P_{\boldsymbol{X}|\boldsymbol{W}^r,\boldsymbol{H}}(\boldsymbol{x}|\boldsymbol{w}^r, \boldsymbol{h}) = \prod_k P_{X|W^r,H}(x(k)|w^r(k), h(k))$.
   - Index these vectors by M_{CEO} where $M_{CEO} \in [1, 2^{nR_{CEO}}]$.
   - So we have $2^{n[\sum_{i=1}^{r} I(W_i;Y_i|H) - I(W_i;W_{\tilde{\mathcal{T}}(\pi,i)}|H) - \delta]}$ different mappings between indices M_{CEO} and vectors $\boldsymbol{x}$, where the one used is determined by M^r. We will therefore denote $\boldsymbol{x}(M_{CEO}, M^r, \boldsymbol{h})$ as the vector indexed by M_{CEO}, M^r. We leave out the notation of f in the sequel, for the sake of brevity, since for decoding agents, the chosen f is known at the agents, so the achievable rate is valid for every realization of f, with high probability.

3) For the compressor at the agents

   For all $\boldsymbol{w}^r$ indicated by M^r,
   - Randomly generate $2^{n[\hat{R}_i - (C_i - \{I(W_i;Y_i|H) - I(W_i;W_{\tilde{\mathcal{T}}(\pi,i)}|H) - \delta\})]}$ vectors $\boldsymbol{u}_i$ of length n according to $\prod_k P_{U_i|W_i,H}(u_i(k)|w_i(k), h(k))$.
   - Repeat the last step for $s_i = 1, \ldots, 2^{n(C_i - \{I(W_i;Y_i|H) - I(W_i;W_{\tilde{\mathcal{T}}(\pi,i)}|H) - \delta\})}$, define the resulting set of $\boldsymbol{u}_i$ of each repetition by S_{s_i}.
   - Index all the generated $\boldsymbol{u}_i$ with $z_i \in [1, 2^{n\hat{R}_i}]$. We will interchangeably use the notation S_{s_i} for the set of vectors $\boldsymbol{u}_i$ as well as for the set of the corresponding z_i.
   - Notice that the mapping between the indices z_i and the vectors $\boldsymbol{u}_i$ depends on $\boldsymbol{w}_i, \boldsymbol{h}$. So we will write $\boldsymbol{u}_i(z_i, \boldsymbol{w}_i, \boldsymbol{h})$ to denote $\boldsymbol{u}_i$ which is indexed by z_i for some specific $\boldsymbol{w}_i, \boldsymbol{h}$.

B. Encoding:

Let $M = (M^r, M_{CEO})$ be the message to be sent (M^r is defined at the previous subsection), and the channel realizations be $\boldsymbol{h}$. The transmitter then sends $\boldsymbol{x}(M_{CEO}, M^r, \boldsymbol{h})$ to the channel.

C. Processing at the agents:

1) Decoding: The i^{th} agent knows $\boldsymbol{h}$ and receives $\boldsymbol{y}_i$ from the channel. It looks for $\hat{\boldsymbol{w}}_i$ so that

$$(\boldsymbol{y}_i, \hat{\boldsymbol{w}}_i) \in \mathbf{T}_\epsilon^{i,1}(\boldsymbol{h}), \tag{158}$$

where

$$\mathbf{T}_\epsilon^{i,1}(\boldsymbol{h}) \triangleq \left\{ \boldsymbol{w}_i, \boldsymbol{y}_i : \begin{array}{c} \forall w \in \mathcal{W}_i, h \in \mathcal{H} : \ \frac{1}{n}\left|N(w,h|\boldsymbol{w}_i,\boldsymbol{h}) - P_{W_i|H}(w|h)N(h|\boldsymbol{h})\right| < \frac{\epsilon}{|\mathcal{W}_i|} \\ \forall y \in \mathcal{Y}_i, h \in \mathcal{H} : \ \frac{1}{n}\left|N(y,h|\boldsymbol{y}_i,\boldsymbol{h}) - P_{Y_i|H}(y|h)N(h|\boldsymbol{h})\right| < \frac{\epsilon}{|\mathcal{Y}_i|} \\ w \in \mathcal{W}_i, y \in \mathcal{Y}_i, h \in \mathcal{H} : \ \frac{1}{n}\left|N(w,y,h|\boldsymbol{w}_i,\boldsymbol{y}_i,\boldsymbol{h}) - P_{W_i,Y_i|H}(w,y|h)N(h|\boldsymbol{h})\right| < \frac{\epsilon}{|\mathcal{Y}_i||\mathcal{W}_i|} \end{array} \right\}. \tag{159}$$

If no such $\hat{\boldsymbol{w}}_i$ exists, chose arbitrary $\hat{\boldsymbol{w}}_i$, and if more than one is found, select one of them arbitrarily. Denote by E_2 the error event where the chosen vector $\hat{\boldsymbol{w}}_i \neq \boldsymbol{w}_i(M^r, \boldsymbol{h})$.

2) Compression: The i^{th} agent chooses any of the z_i such that

$$\left(\boldsymbol{u}_i(z_i, \hat{\boldsymbol{w}}_i, \boldsymbol{h}), \boldsymbol{y}_i, \hat{\boldsymbol{w}}_i\right) \in \mathbf{T}_\epsilon^{t,2}(\boldsymbol{h}). \tag{160}$$

The event where no such z_i is found is defined as the error event E_3.

After deciding on z_i the agent transmits s_i, which fulfills $z_i \in S_{s_i}$, and $\hat{M}_i$ to the final destination through the lossless link, where $\hat{M}_i$ corresponds to $\hat{\boldsymbol{w}}_t$.

D. Decoding (at the destination):

The destination retrieves $\hat{M}^r$ and $s^r \triangleq (s_1, \ldots, s_r)$ from the lossless links.

The destination then finds the set of indices $\hat{z}^r \triangleq \{\hat{z}_1, \ldots, \hat{z}_r\}$ of the compressed vectors $\hat{\boldsymbol{u}}^r$ and $\hat{M_{CEO}}$ which satisfy

$$\begin{cases} \left(\boldsymbol{x}(\hat{M}_{CEO}, \hat{M}^r, \boldsymbol{h}, f), \hat{\boldsymbol{u}}^r(\hat{z}^r, \hat{M}^r, \boldsymbol{h}), \hat{\boldsymbol{w}}^r(\hat{M}^r, \boldsymbol{h})\right) \in \mathbf{T}_\epsilon^3(\boldsymbol{h}) \\ \hat{z}^r \in S_{s_1} \times \cdots \times S_{s_r}. \end{cases} \tag{161}$$

Where $\mathbf{T}_\epsilon^3$ is defined in the standard way, as (103). If there is no such $\hat{z}^r, \hat{M}_{CEO}$, or if there is more than one, the destination chooses one arbitrarily. Define error E_4 as the event where $\hat{M}_{CEO} \neq M_{CEO}$.

Correct decoding means that the destination decides $\hat{M} = M$. An achievable rate R was defined as when the final destination receives the transmitted message with an error probability which is made arbitrarily small for sufficiently large block length n.

E. Error analysis

The error probability is upper bounded by:

$$\Pr\{\text{error}\} = \Pr\left(\cup_{i=1}^4 E_i\right) \leq \sum_{i=1}^4 \Pr(E_i). \tag{162}$$

Where:

1) E_1: No r-tuple $\boldsymbol{w}^r$ jointly typical is found.
2) E_2: A different $\hat{\boldsymbol{w}}_i \neq \boldsymbol{w}_i$ is selected by the i^{th} agent.
3) E_3: No $\boldsymbol{u}_i(z_i, \hat{\boldsymbol{w}}_i, \boldsymbol{h})$ is jointly typical with $(\boldsymbol{y}_i, \hat{\boldsymbol{w}}_i)$.
4) E_4: Decoding error $\boldsymbol{x}(\hat{M}_{CEO}, \hat{M}_{\mathcal{T}}, f) \neq \boldsymbol{x}(M, f)$.

Next, we will upper bound the probabilities of the individual error events by arbitrarily small ϵ.

1) E_1: From Lemma 4, it is evident that $\Pr(E_1)$ can be made as small as desired, when n is increased, as long as $\delta > 0$.

2) E_2: By Lemma 2, the probability of jointly distributed variables not to be ϵ-typical is as small as desired for n sufficiently large. According to Lemma 3, the probability that another $\hat{\boldsymbol{w}}_i$ belongs to $\mathbf{T}_\epsilon^{i,1}$ is upper bounded by $2^{-n[I(W_i;Y_i|H)-\epsilon]}$. Since there are no more than $2^{n[I(W_i;Y_i|H)-I(W_i;W_{\tilde{\mathcal{T}}(\pi,i)}|H)-\delta]}$ such $\hat{\boldsymbol{w}}_i$, the probability of E_2 can be made arbitrarily small as n goes to infinity as long as $I(W_i;W_{\tilde{\mathcal{T}}(\pi,i)}|H) + \delta > \epsilon$.

3) E_3: According to Lemma 4, the probability $\Pr\{E_3\}$ can be made as small as desired, for n sufficiently large, as long as

$$\hat{R}_i > I(U_i;Y_i|W_i,H). \tag{163}$$

4) E_4: Consider the case where $\hat{M}_{CEO} \neq M_{CEO}$ and $\hat{z}_{\mathcal{S}} \neq z_{\mathcal{S}}$. There are

$$2^{n[R_{CEO}+\sum_{i\in\mathcal{S}}[\hat{R}_i-(C_i-\{I(W_i;Y_i|H)-I(W_i;W_{\tilde{\mathcal{T}}(\pi,i)}|H)-\delta\})]]}$$

such vectors, and the probability of $(\boldsymbol{x}(\hat{M}),\boldsymbol{u}_{\mathcal{S}}(\hat{z}_{\mathcal{S}}),\boldsymbol{u}_{\mathcal{S}^C}(\hat{z}_{\mathcal{S}^C}))$ to be jointly typical is upper bounded by (Lemma 3) $2^{n[H(X,U^r|W^r,H)-H(X|W^r,H)-H(U_{\mathcal{S}^C}|W^r,H)-\sum_{i\in\mathcal{S}}H(U_i|W_i,H)+\epsilon]}$. Thus the rate R_{CEO} is achievable if:

$$\begin{aligned} R_{CEO} &< \sum_{i\in\mathcal{S}}[C_i-\{I(W_i;Y_i|H)-I(W_i;W_{\tilde{\mathcal{T}}(\pi,i)}|H)-\delta\}-\hat{R}_i+H(U_i|W_i,H)]-H(U_{\mathcal{S}}|X,W^r,H)-H(U_{\mathcal{S}^C}|X,U_{\mathcal{S}},W^r,H) \\ &< \sum_{i\in\mathcal{S}}[C_i-\{I(W_i;Y_i|H)-I(W_i;W_{\tilde{\mathcal{T}}(\pi,i)}|H)-\delta\}-I(Y_i;U_i|X,W_i,H)]+I(U_{\mathcal{S}^C};X|W^r,H), \end{aligned} \tag{164}$$

where the second inequality is due to (163) and because of the Markov chain $U_i-(W^r,X,H)-U_{1,\dots,i-1,i+1,\dots,r}$. Finally, the overall achievable rate is equal to

$$R_{CEO}+\sum_{i=1}^{r}\{I(W_i;Y_i|H)-I(W_i;W_{\tilde{\mathcal{T}}(\pi,i)}|H)-\delta\}, \tag{165}$$

which proves Proposition 15. ■

References

[1] A. Sanderovich, S. Shamai, Y. Steinberg, and G. Kramer, "Communication via decentralized processing," *Submitted to IEEE trans. on info. Theory*, Nov. 2005, partially presented at [24].

[2] E. Telatar, "Capacity of multi-antenna Gaussian channels," *European Transactions on Telecommunications, ETT*, vol. 10, no. 6, pp. 585–596, Nov. 1999.

[3] D. N. C. Tse and L. Zheng, "Diversity and multiplexing: A fundamental tradeoff in multiple-antenna channels," *IEEE Trans. Inform. Theory*, vol. 49, no. 50, pp. 1073–1096, May 2003.

[4] A. S. Avestimehr and D. N. Tse, "Outage capacity of the fading relay channel in the low SNR regime," *Submitted to IEEE trans. on info. Theory*, 2006.

[5] M. Yuksel and E. Erkip, "Cooperative wireless systems: A diversity-multiplexing tradeoff perspective," *IEEE Trans. Inform. Theory*, 2006, under review.

[6] H. Weingarten, Y. Steinberg, and S. Shamai, "The capacity region of the Gaussian multiple-input multiple-output broadcast channel," *IEEE Trans. Inform. Theory*, vol. 52, no. 9, pp. 3936–3964, Sep. 2006.

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

[8] H. Bölcskei and R. U. Nabar, "Realizing MIMO gains without user cooperation in large single-antenna adhoc wireless networks," in *Proc. of IEEE Int. Symp. Info. Theory (ISIT2004)*, Chicago, IL, June 2004, p. 18.

[9] P. Gupta and P. R. Kumar, "Towards an information theory of large networks: an achievable rate region," *IEEE Trans. Inform. Theory*, vol. 49, no. 8, pp. 1877–1894, Aug. 2003.

[10] I. Maric and R. D. Yates, "Forwarding strategies for Gaussian parallel-relay networks," in *Proc. of IEEE Int. Symp. Info. Theory (ISIT2004)*, Chicago, IL, June 2004, p. 269.

[11] O. Somekh, B. M. Zaidel, and S. Shamai, "Spectral efficiency of joint multiple cell-site processors for randomly spread DS-CDMA systems," in *Proc. of IEEE Int. Symp. Info. Theory (ISIT2004)*, Chicago, IL, June 2004, p. 278.

[12] Y. Oohama, "The rate-distortion function for the quadratic Gaussian CEO problem," *IEEE Trans. Inform. Theory*, vol. 44, no. 3, pp. 1057–1070, May 1998.

[13] A. D. Wyner and J. Ziv, "The rate-distortion function for source coding with side information at the decoder," *IEEE Trans. Inform. Theory*, vol. 22, no. 1, pp. 1–10, Jan 1976.

[14] A. B. Wagner, S. Tavildar, and P. Viswanath, "The rate region of the quadratic Gaussian two-terminal source-coding problem," *submitted to IEEE trans. on IT*, Feb 2006. [Online]. Available: http://www.citebase.org/cgi-bin/citations?id=oai:arXiv.org:cs/0510095

[15] Y. Oohama, "Rate-distortion theory for Gaussian multiterminal source coding systems with several side informations at the decoder," *IEEE Trans. Inform. Theory*, vol. 51, no. 7, pp. 2577–2593, July 2005.

[16] S. C. Draper and G. W. Wornell, "Side information aware coding strategies for sensor networks," *IEEE J. Select. Areas Commun.*, vol. 22, no. 6, pp. 966–976, Aug. 2004.

[17] T. M. Cover and A. A. El-Gamal, "Capacity theorems for the relay channel," *IEEE Trans. Inform. Theory*, vol. 25, no. 5, pp. 572–584, Jan 1979.

[18] G. Kramer, M. Gastpar, and P. Gupta, "Information-theoretic multi-hopping for relay networks," in *International Zürich seminar on communication*, Swizerland, Feb. 2004.

[19] T. Liu and P. Viswanath, "An extremal inequality motivated by multiterminal information theoretic problems," *Submitted to IEEE trans. on info. Theory*, 2006.

[20] H. S. Witsenhausen and A. D. Wyner, "a canditional entropy bound for a pair of discrete random variables," *IEEE Trans. Inform. Theory*, vol. IT-21, no. 5, pp. 493–501, Sep 1975.

[21] N. Chayat and S. Shamai, "Extension of an entropy property for binary input memoryless symmetric channels," *IEEE Trans. Inform. Theory*, vol. 35, no. 5, pp. 1077–1079, Sep. 1989.

[22] A. Tulino, A. Lozano, and S. Verdú, "Capacity-achieving input covariance for single-user multi-antenna channels," *IEEE Trans. Inform. Theory*, vol. 5, no. 1, Jan. 2006.

[23] T. M. Cover and J. A. Thomas, *Elements of Information theory*. John Wiley & Sons, Inc., 1991.

[24] A. Sanderovich, S. Shamai, Y. Steinberg, and G. Kramer, "Communication via decentralized processing," in *Proc. of IEEE Int. Symp. Info. Theory (ISIT2005)*, Adelaide, Australia, Sep. 2005, pp. 1201–1205.

[25] L. Li and A. J. Goldsmith, "Capacity and optimal resource allocation for fading broadcast channelspart I: Ergodic capacity," *IEEE Trans. Inform. Theory*, vol. 47, no. 3, pp. 1083–1102, Mar. 2001.

[26] E. Biglieri, J. Proakis, and S. Shamai, "Fading channels: Information-theoretic and communications aspects," *IEEE Trans. Inform. Theory*, vol. 44, no. 6, pp. 2619–2692, Oct. 1998.

[27] D. Krithivasan and S. S. Pradhan, "Lattices for distributed source coding: Jointly Gaussian sources and reconstruction of a linear function," *IEEE Trans. Inform. Theory*, July 2007, submitted.

[28] G. Caire and S. Shamai, "On the achievable throughput of a multiantenna Gaussian broadcast channel," *IEEE Trans. Inform. Theory*, vol. 49, no. 7, pp. 1691–1706, July 2003.

[29] S. Vishwanath, N. Jindal, and A. Goldsmith, "The "Z" channel," in *Proceedings of IEEE Global Telecommunications Conference (GlobeCom)*, San Francisco, CA, Dec.

[30] I. Csiszár and J. Körner, *Information Theory: Coding Theorems for Discrete Memoryless Systems*. New York: Academic, 1981.

[31] T. S. Han and K. Kobayashi, "A unified achievable rate region for a general class of multiterminal source coding systems," *IEEE Trans. Inform. Theory*, vol. IT-26, no. 3, pp. 277–288, May 1980.