Distributed Compression for MIMO Coordinated Networks with a Backhaul Constraint

Aitor del Coso, *Member, IEEE*, and Sébastien Simoens

Abstract—We consider the uplink of a backhaul-constrained, MIMO coordinated network. That is, a single-frequency network with $N + 1$ multi-antenna base stations (BSs) that cooperate in order to decode the users' data, and that are linked by means of a common lossless backhaul, of limited capacity R. To implement the receive cooperation, we propose distributed compression: N BSs, upon receiving their signals, compress them using a multi-source lossy compression code. Then, they send the compressed vectors to a central BS, which performs users' decoding. Distributed Wyner-Ziv coding is proposed to be used, and is designed in this work. The first part of the paper is devoted to a network with a unique multi-antenna user, that transmits a predefined Gaussian space-time codeword. For such a scenario, the "compression noise" covariance at the BSs is optimized, considering the user's achievable rate as the performance metric. In particular, for $N = 1$ the optimum covariance is derived in closed form, while for $N > 1$ an iterative algorithm is devised. The second part of the contribution focusses on the multi-user scenario. For it, the achievable rate region is obtained by means of the optimum "compression noise" covariances for sum-rate and weighted sum-rate, respectively.

Index Terms—MAC, multiple relay channel, decode-and-forward.

I. INTRODUCTION

THE current trend to reduce the frequency reuse factor of cellular networks makes inter-cell interference a critical problem. A wide range of multi-antenna techniques are reviewed in [1] to overcome it, including coordinated scheduling and interference cancelation. However, a more complex but spectrally efficient solution can be proposed: coordinated cellular networks [2]. They consist of single-frequency networks where base stations (BSs) cooperate to: *i*) beamform towards the mobile terminals in the downlink, and *ii*) coherently detect them in the uplink [3]. Hereafter, we restrain ourselves to the uplink channel.

Preliminary research on the uplink performance of coordinated networks consider all BSs connected via a lossless backhaul with unlimited capacity [4] [5]. Accordingly, the capacity region of the network equals that of a multiple-input, multiple-output (MIMO) multi-access channel, with a supra-receiver containing all the antennas of all cooperative BSs [6].

Such an assumption seems optimistic in short-mid term, as operators are currently worried about the costs of upgrading their backhaul to support *e.g.*, High-speed Packet Access (HSPA). To deal with a realistic backhaul constraint, two approaches have been proposed: *i*) *Distributed decoding* [7], consisting on a demodulating scheme distributively carried out among BSs, based on local decisions and belief propagation. Decoding delay appears to be its main drawback. *ii*) *Quantization* [8], where BSs quantize their observations and forward them to the decoding unit. Its main limitation relies on its inability to take profit of signal correlation between BSs, which introduces redundancy into the backhaul.

This paper considers a new approach for the network: *distributed compression*. The cooperative BSs, upon receiving their signals, distributively compress them using a multi-source lossy compression code [9]. Then, via the lossless backhaul, they transmit the compressed signals to the central unit (also a BS), which decompresses them using its own received signal as side information. Then it estimates the users' messages. Distributed compression has been previously proposed for coordinated networks in [10], [11]. In those works, authors consider single-antenna BSs with block and ergodic fading. We build upon these results to extend the analysis to the MIMO case with constant channel gains.

The compression of signals with side information at the decoder was introduced by Wyner and Ziv in [12], [13]. They showed that side information at the encoder is useless (*i.e.*, the rate-distortion tradeoff remains unchanged) to compress a single, Gaussian source when the side information is available at the decoder [13, Section 3]. Unfortunately, when considering multiple (correlated) signals, independently compressed at different BSs and to be recovered at a central unit with side information, no conclusive results are available to date. Indeed, this is an open problem for which, to the best of authors knowledge, the tightest lower bound (in a rate-distortion sense) is obtained with Distributed Wyner-Ziv (D-WZ) compression [14]. Such a compression is the direct extension of Berger-Tung coding to the decoding side information case [15]. In turn, Berger-Tung compression is the lossy counterpart of the Slepian-Wolf lossless coding [16]. D-WZ coding is thus the compression scheme proposed to be used, as detailed below.

Summary of Contributions. This paper considers a single-frequency network with $N + 1$ multi-antenna BSs. The first base station, denoted BS_0, is the central unit and centralizes the users' decoding. The rest, $BS_1, \cdots, BS_N$, are cooperative BSs, which distributively compress their received signals using a D-WZ code, and independently transmit them to BS_0 via the common backhaul of aggregate capacity R. In the network, constant, *frequency-flat* channels are assumed as well as receive channel state information (CSI) at the central unit.

The first part of the paper is devoted to a network with a single user, equipped with multiple antennas and transmitting a pre-defined Gaussian space-time codeword. The contributions are organized as follows:

- The system model and the user's achievable rate are presented in Sec. II. The latest is derived modelling the compression step by means of Gaussian "compression noise", added by the BSs on their observations before retransmitting them to the central unit.
- Considering a unique cooperative BS (*i.e.,* $N = 1$), Sec. III derives, in closed form, the optimum "compression noise" covariance for which the user's rate is maximized. We also show that conditional Karhunen-Loève transform plus independent Wyner-Ziv coding of scalar streams is optimal.
- The analysis is extended in Sec. IV to arbitrary N BSs. In particular, the optimum "compression noise" covariances are obtained by means of an iterative algorithm, constructed using dual decomposition and a non-linear block coordinate approach [17], [18].

The second part of the paper extends the analysis to multiple users transmitting simultaneously:

- First, the sum-rate of the network is derived in Sec. V, adapting single-user results. Later, the weighted sum-rate, and its associated optimum "compression noise" covariances, is obtained by means of an iterative algorithm, constructed using dual decomposition and Gradient Projection [18].

Notation. $\mathbf{E}\{\cdot\}$ denotes expectation. $\mathbf{A}^T$, $\mathbf{A}^\dagger$ and a^* stand for the transpose of $\mathbf{A}$, conjugate transpose of $\mathbf{A}$ and complex conjugate of a, respectively. $[a]^+ = \max\{a, 0\}$. $I(\cdot; \cdot)$ denotes mutual information, $H(\cdot)$ entropy. The derivative of scalar $f(\cdot)$ with respect to matrix $\mathbf{X}$ is defined as in [19], *i.e.,* $\left[\frac{\partial f}{\partial \mathbf{X}}\right]_{i,j} = \frac{\partial f}{\partial [\mathbf{X}]_{i,j}}$. In such a way, *e.g.,* $\frac{\partial \mathrm{tr}\{\mathbf{A}\mathbf{X}\}}{\partial \mathbf{X}} = \mathbf{A}^T$. We compactly write $\mathbf{Y}_{1:N} = \{\mathbf{Y}_1, \cdots, \mathbf{Y}_N\}$, $\mathbf{Y}_\mathcal{G} = \{\mathbf{Y}_i | i \in \mathcal{G}\}$ and $\mathbf{Y}_j^c = \{\mathbf{Y}_i | i \neq j\}$. A sequence $\{\mathbf{Y}_i^t\}_{t=1}^n$ is compactly denoted by $\mathbf{Y}_i^n$. $\mathrm{diag}(\mathbf{A}_1, \cdots, \mathbf{A}_n)$ is a block-diagonal matrix with $\mathbf{A}_i$ square. $\mathrm{coh}(\cdot)$ stands for convex hull. Finally, the covariance of $\mathbf{X}$ conditioned on $\mathbf{Y}$ is denoted by $\mathbf{R}_{\mathbf{X}|\mathbf{Y}}$ and computed $\mathbf{R}_{\mathbf{X}|\mathbf{Y}} = \mathbf{E}\left\{(\mathbf{X} - \mathbf{E}\{\mathbf{X}|\mathbf{Y}\})(\mathbf{X} - \mathbf{E}\{\mathbf{X}|\mathbf{Y}\})^\dagger | \mathbf{Y}\right\}$.

II. SYSTEM MODEL

Let a single source s, equipped with M_t antennas, transmit data to base stations $\mathrm{BS}_0, \cdots, \mathrm{BS}_N$, each equipped with M_i, $i = 0, \cdots, N$ antennas. The BSs are connected to a common lossless backhaul of aggregate capacity R, and BS_0 is selected to be the decoding unit. Such an aggregate sum-rate constraint aims at modeling 3G scenarios where BSs are connected to a common backhaul via radio network controllers; however, further scenarios can be identified with *e.g.* per-link constraints. Likewise, we assume the user-to-BSs assignment to be given by upper layers and out of the scope of the paper (see *e.g.* [5] for assignment algorithms and selection criteria).

The source transmits a message $\omega \in \{1, \cdots, 2^{n\theta}\}$ mapped onto codeword $\mathbf{X}_s^n$, drawn *i.i.d.* from vector $\mathbf{X}_s \sim \mathcal{CN}(\mathbf{0}, \mathbf{Q})$ and not subject to optimization. n is the number of transmitted

symbols. The BSs receive:

$$\mathbf{Y}_i^n = \mathbf{H}_{s,i} \cdot \mathbf{X}_s^n + \mathbf{N}_i^n, \quad i = 0, \cdots, N. \quad (1)$$

$\mathbf{H}_{s,i}$ is the MIMO channel matrix between user s and BS_i, and $\mathbf{N}_i \sim \mathcal{CN}(0, \sigma_r^2 \mathbf{I})$ is additive white Gaussian noise (AWGN). Channel coefficients are all known at BS_0.

$\mathrm{BS}_1, \cdots, \mathrm{BS}_N$ then apply a D-WZ code to their received signals and send them to BS_0 via the common backhaul. In turn, BS_0 decodes the user's message in two consecutive steps: first, it decompresses the BSs signals using $\mathbf{Y}_0^n$ as side information. Next, it coherently combines them (along with $\mathbf{Y}_0^n$) to decode the user message. Such a coding/decoding scheme is equivalent to that presented in [10, Theorem 1] and we don't claim it is optimal. Indeed, two weaknesses can be identified: *i*) as pointed out in [10], BS_0 is assumed to decompress without errors, which is unnecessarily restrictive. In fact, it would be enough forcing the user's message to be decoded without errors, and *ii*) D-WZ is not shown to be optimal. However, despite its sub-optimality, we use this approach as a first application of compression to MIMO coordinated networks.

A. The Achievable Rate

Proposition 1: Let $\mathbf{X}_s \sim \mathcal{CN}(\mathbf{0}, \mathbf{Q})$. The MIMO coordinated network achieves the rate (2) with D-WZ compression, where the conditional covariance $\mathbf{R}_{\mathbf{Y}_{1:N}|\mathbf{Y}_0}$ follows (42) and $\mathbf{\Phi}_n$ is the spatial covariance of the independent, Gaussian, "compression noise" at BS_n.

Remark 1: The maximization in (2) is not concave in standard form: although the feasible set is regular and convex, the objective function is not concave on $\mathbf{\Phi}_1, \cdots, \mathbf{\Phi}_N$.

Proof: The Proposition is proven by merely applying[1] [10, Theorem 1]. In particular, considering D-WZ coding with compression rates $\rho_1, \cdots, \rho_N$ at the cooperative BSs, the user's transmission rate θ is achievable if there exists a set of random vectors $\hat{\mathbf{Y}}_{1:N}$ such that:

i) $\left(\mathbf{Y}_0, \mathbf{Y}_i^c, \hat{\mathbf{Y}}_i^c\right) \leftrightarrow \mathbf{Y}_i \leftrightarrow \hat{\mathbf{Y}}_i$ form a Markov chain,

ii) $\theta \leq I\left(\mathbf{X}_s; \mathbf{Y}_0, \hat{\mathbf{Y}}_{1:N}\right)$,

iii) $\forall \mathcal{G} \subseteq \{1, \cdots, N\} : I\left(\mathbf{Y}_\mathcal{G}; \hat{\mathbf{Y}}_\mathcal{G} | \mathbf{Y}_0, \hat{\mathbf{Y}}_\mathcal{G}^c\right) \leq \sum_{i \in \mathcal{G}} \rho_i$.

The statement is proven for discrete channels by Sanderovich *et. al.* in [10, Appendix III] and extended to the Gaussian case in [10, Section VI]. In the framework of distributed compression, $\hat{\mathbf{Y}}_i$ represents the observation of BS_i reconstructed by the decoding unit.

Let us notice now that, in our setup, there is only an aggregate backhaul rate constraint R, *i.e.,* $\sum_{i \in \mathcal{G}} \rho_i \leq$ R, $\forall \mathcal{G} \subseteq \{1, \cdots, N\}$. Therefore, the set of constraints in *iii*) can all be re-stated as: $\forall \mathcal{G} \subseteq \{1, \cdots, N\} : I\left(\mathbf{Y}_\mathcal{G}; \hat{\mathbf{Y}}_\mathcal{G} | \mathbf{Y}_0, \hat{\mathbf{Y}}_\mathcal{G}^c\right) \leq$ R. However, from the Markov chain in *i*), the following inequality can be shown: $I\left(\mathbf{Y}_\mathcal{G}; \hat{\mathbf{Y}}_\mathcal{G} | \mathbf{Y}_0, \hat{\mathbf{Y}}_\mathcal{G}^c\right) \leq I\left(\mathbf{Y}_{1:N}; \hat{\mathbf{Y}}_{1:N} | \mathbf{Y}_0\right)$, $\forall \mathcal{G} \subseteq \{1, \cdots, N\}$. Accordingly, forcing the constraint $I\left(\mathbf{Y}_{1:N}; \hat{\mathbf{Y}}_{1:N} | \mathbf{Y}_0\right) \leq$ R to hold makes

---

[1]It is necessary to take into account that, unlike [10], in our case the central unit uses its own received signal as side information to decompress.

$$R_{\text{D-WZ}} = \max_{\boldsymbol{\Phi}_1, \cdots, \boldsymbol{\Phi}_N \succeq 0} \log \det \left(\mathbf{I} + \frac{\mathbf{Q}}{\sigma_r^2} \mathbf{H}_{s,0}^\dagger \mathbf{H}_{s,0} + \mathbf{Q} \sum_{n=1}^N \mathbf{H}_{s,n}^\dagger \left(\sigma_r^2 \mathbf{I} + \boldsymbol{\Phi}_n \right)^{-1} \mathbf{H}_{s,n} \right), \qquad (2)$$

$$\text{s.t.} \quad \log \det \left(\mathbf{I} + \text{diag} \left(\boldsymbol{\Phi}_1^{-1}, \cdots, \boldsymbol{\Phi}_N^{-1} \right) \mathbf{R}_{\mathbf{Y}_{1:N}|\mathbf{Y}_0} \right) \leq \mathrm{R}$$

all constraints in *iii*) hold too. The converse is, obviously, true by noting that $I\left(\mathbf{Y}_{1:N}; \hat{\mathbf{Y}}_{1:N}|\mathbf{Y}_0\right)$ is equal to $I\left(\mathbf{Y}_{\mathcal{G}}; \hat{\mathbf{Y}}_{\mathcal{G}}|\mathbf{Y}_0, \hat{\mathbf{Y}}_{\mathcal{G}}^c\right)$ with $\mathcal{G} = \{1, \cdots, N\}$. Therefore,

$$iii) \ I\left(\mathbf{Y}_{1:N}; \hat{\mathbf{Y}}_{1:N}|\mathbf{Y}_0\right) \leq \mathrm{R} \Leftrightarrow \qquad (3)$$
$$I\left(\mathbf{Y}_{\mathcal{G}}; \hat{\mathbf{Y}}_{\mathcal{G}}|\mathbf{Y}_0, \hat{\mathbf{Y}}_{\mathcal{G}}^c\right) \leq \mathrm{R}, \forall \mathcal{G} \subseteq \{1, \cdots, N\}.$$

Then, consider Gaussian random vectors of the form $\hat{\mathbf{Y}}_i = \mathbf{Y}_i + \mathbf{Z}_i$, where $\mathbf{Z}_i \sim \mathcal{CN}(\mathbf{0}, \boldsymbol{\Phi}_i)$ is independent of $\mathbf{Y}_i$ and is referred to as "compression noise". With such vectors, we evaluate that, if

$$ii) \theta \leq \log \det \left(\mathbf{I} + \frac{\mathbf{Q}}{\sigma_r^2} \mathbf{H}_{s,0}^\dagger \mathbf{H}_{s,0} \right. \qquad (4)$$
$$\left. + \mathbf{Q} \sum_{n=1}^N \mathbf{H}_{s,n}^\dagger \left(\sigma_r^2 \mathbf{I} + \boldsymbol{\Phi}_n \right)^{-1} \mathbf{H}_{s,n} \right)$$

$$iii) \log \det \left(\mathbf{I} + \text{diag} \left(\boldsymbol{\Phi}_1^{-1}, \cdots, \boldsymbol{\Phi}_N^{-1} \right) \mathbf{R}_{\mathbf{Y}_{1:N}|\mathbf{Y}_0} \right) \leq \mathrm{R},$$

then θ is achievable. ∎

The main goal of this paper is to optimize, by means of iterative algorithms, the spatial covariances matrices $\boldsymbol{\Phi}_1, \cdots, \boldsymbol{\Phi}_N$ so as to maximize the coordinated network achievable rate.

B. Useful Upper Bounds

Upper Bound 1: The achievable rate $R_{\text{D-WZ}}$ in (2) is upper bounded by

$$\begin{aligned} R_{\text{D-WZ}} &\leq I\left(\mathbf{X}_s; \mathbf{Y}_0, \mathbf{Y}_{1:N}\right) \\ &= \log \det \left(\mathbf{I} + \frac{\mathbf{Q}}{\sigma_r^2} \sum_{n=0}^N \mathbf{H}_{s,n}^\dagger \mathbf{H}_{s,n} \right). \end{aligned} \qquad (5)$$

Upper Bound 2: The achievable rate $R_{\text{D-WZ}}$ in (2) satisfies

$$\begin{aligned} R_{\text{D-WZ}} &\leq I\left(\mathbf{X}_s; \mathbf{Y}_0\right) + \mathrm{R} \\ &= \log \det \left(\mathbf{I} + \frac{1}{\sigma_r^2} \mathbf{H}_{s,0} \mathbf{Q} \mathbf{H}_{s,0}^\dagger \right) + \mathrm{R}. \end{aligned} \qquad (6)$$

Proof: It follows directly from the *max-flow-min-cut* upper bound [20, Theorem 14.10.1] ∎

Remark 2: Notice that, independently of the number of BSs, the achievable rate is bounded above by the capacity with BS_0 plus the backhaul rate.

III. THE TWO-BASE STATIONS CASE

We first solve (2) for $N = 1$. As mentioned, the objective function, which has to be maximized, is convex on $\boldsymbol{\Phi}_1 \succeq 0$. In order to make it concave, we change the variables $\boldsymbol{\Phi}_1 = \mathbf{A}_1^{-1}$, so that

$$R_{\text{D-WZ}} = \max_{\mathbf{A}_1 \succeq 0} \log \det \left(\mathbf{I} + \frac{\mathbf{Q}}{\sigma_r^2} \mathbf{H}_{s,0}^\dagger \mathbf{H}_{s,0} \right. \qquad (7)$$
$$\left. + \mathbf{Q} \mathbf{H}_{s,1}^\dagger \left(\mathbf{A}_1 \sigma_r^2 + \mathbf{I} \right)^{-1} \mathbf{A}_1 \mathbf{H}_{s,1} \right)$$
$$\text{s.t.} \quad \log \det \left(\mathbf{I} + \mathbf{A}_1 \mathbf{R}_{\mathbf{Y}_1|\mathbf{Y}_0} \right) \leq \mathrm{R}.$$

The objective has turned into concave. However, the constraint now does not define a convex feasible set. Therefore, Karush-Kuhn-Tucker (KKT) conditions become necessary but not sufficient for optimality. In order to solve the problem, we thus need to resort to the general sufficiency condition [18, Proposition 3.3.4]. The solution is presented in the next Theorem.

Theorem 1: Let $\mathbf{X}_s \sim \mathcal{CN}(\mathbf{0}, \mathbf{Q})$ and let the conditional covariance be (see Appendix A-A):

$$\mathbf{R}_{\mathbf{Y}_1|\mathbf{Y}_0} = \mathbf{H}_{s,1} \left(\mathbf{I} + \frac{\mathbf{Q}}{\sigma_r^2} \mathbf{H}_{s,0}^\dagger \mathbf{H}_{s,0} \right)^{-1} \mathbf{Q} \mathbf{H}_{s,1}^\dagger + \sigma_r^2 \mathbf{I} \qquad (8)$$

with eigen decomposition $\mathbf{R}_{\mathbf{Y}_1|\mathbf{Y}_0} = \mathbf{U} \text{diag}(s_1, \cdots, s_{M_1}) \mathbf{U}^\dagger$. $R_{\text{D-WZ}}$ is attained with a "compression noise" covariance at BS_1 equal to $\boldsymbol{\Phi}_1^* = \mathbf{U} \left(\text{diag}(\eta_1, \cdots, \eta_{M_1}) \right)^{-1} \mathbf{U}^\dagger$, where

$$\eta_j = \left[\frac{1}{\lambda} \left(\frac{1}{\sigma_r^2} - \frac{1}{s_j} \right) - \frac{1}{\sigma_r^2} \right]^+, \qquad (9)$$

and λ is such that $\sum_{j=1}^{M_1} \log(1 + \eta_j s_j) = \mathrm{R}$.

Proof: See Appendix B. The result can be viewed as a Wyner-Ziv rate allocation, equivalent to that in [21]. ∎

A. Practical Implementation

In this subsection, we show that the compression derived in Theorem 1 can be practically carried out using a Transform Coding (TC) approach. TC consists of BS_1 first transforming its received vector using an invertible linear function and then separately compressing the resulting scalar streams [22]. In particular, we show that the conditional Karhunen-Loève transform (CKLT) is an optimal linear transformation [22]. First, recall that multiplying a vector by a non-singular matrix does not change the mutual information [20], *i.e.*, $I\left(\mathbf{X}_s; \mathbf{Y}_0, \hat{\mathbf{Y}}_1\right) = I\left(\mathbf{X}_s; \mathbf{Y}_0, \mathbf{U}^\dagger \hat{\mathbf{Y}}_1\right)$ and $I\left(\mathbf{Y}_1; \hat{\mathbf{Y}}_1|\mathbf{Y}_0\right) = I\left(\mathbf{Y}_1; \mathbf{U}^\dagger \hat{\mathbf{Y}}_1|\mathbf{Y}_0\right)$. From Theorem 1, the optimum compressed vector satisfies $\hat{\mathbf{Y}}_1^* = \mathbf{Y}_1 + \mathbf{Z}_1^*$, with $\mathbf{Z}_1^* \sim \mathcal{CN}\left(0, \mathbf{U}\boldsymbol{\eta}^{-1}\mathbf{U}^\dagger\right)$ and $\mathbf{R}_{\mathbf{Y}_1|\mathbf{Y}_0} = \mathbf{U}\mathbf{S}\mathbf{U}^\dagger$. Therefore, the following compressed vectors are also optimal

$$\hat{\mathbf{Y}}_1 = \mathbf{U}^\dagger \mathbf{Y}_1 + \mathbf{U}^\dagger \mathbf{Z}_1^*, \qquad (10)$$

where vector $\mathbf{U}^\dagger \mathbf{Y}_1$ is referred to as the CKLT of vector $\mathbf{Y}_1$. Notice now that $\mathbf{R}_{\hat{\mathbf{Y}}_1|\mathbf{Y}_0} = \mathbf{R}_{\mathbf{U}^\dagger \mathbf{Y}_1|\mathbf{Y}_0} + \mathbf{R}_{\mathbf{U}^\dagger \mathbf{Z}_1^*} = \mathbf{S} + \boldsymbol{\eta}^{-1}$ is diagonal. Therefore, the elements of the compressed vector $\hat{\mathbf{Y}}_1$ are conditionally uncorrelated given $\mathbf{Y}_0$. Likewise, so are the elements of vector $\mathbf{U}^\dagger \mathbf{Y}_1$. Due to this uncorrelation, each element $j = 1, \cdots, M_1$ of vector $\mathbf{U}^\dagger \mathbf{Y}_1$ can be compressed, without loss of optimality, independently of the compression of the others elements, at a compression rate $r_j = \log(1 + \eta_j s_j)$, $j = 1, \cdots, M_1$ [13]. In fact, from Theorem 1 we validate that $\sum_{j=1}^{M_1} r_j = \mathrm{R}$. This demonstrates that CKLT plus independent coding of streams is optimal, not only for minimizing distortion as shown in [22], but also for maximizing the achievable rate of coordinated networks.

IV. THE MULTIPLE-BASE STATIONS CASE

Consider now BS_0 assisted by $N > 1$ cooperative BSs. The achievable rate follows (2) where, as mentioned, the objective function is not concave over $\boldsymbol{\Phi}_n$, $n = 1, \cdots, N$. To make it concave, we again change the variables: $\boldsymbol{\Phi}_n = \mathbf{A}_n^{-1}$, $n = 1, \cdots, N$, so that:

$$R_{\text{D-WZ}} = \max_{\mathbf{A}_1, \cdots, \mathbf{A}_N \succeq 0} \log \det \left(\mathbf{I} + \frac{\mathbf{Q}}{\sigma_r^2} \mathbf{H}_{s,0}^\dagger \mathbf{H}_{s,0} \right. \tag{11}$$

$$\left. + \mathbf{Q} \sum_{n=1}^{N} \mathbf{H}_{s,n}^\dagger \left(\mathbf{A}_n \sigma_r^2 + \mathbf{I} \right)^{-1} \mathbf{A}_n \mathbf{H}_{s,n} \right)$$

$$\text{s.t.} \quad \log \det \left(\mathbf{I} + \text{diag} \left(\mathbf{A}_1, \cdots, \mathbf{A}_N \right) \mathbf{R}_{\mathbf{Y}_{1:N} | \mathbf{Y}_0} \right) \leq \text{R}.$$

As previously, the feasible set does not define a convex set. Our strategy to solve the optimization is the following: first, we show that (although not convex) the duality gap for the problem is zero. Later, we propose an iterative algorithm that solves the dual problem, thus solving the primal problem too. The key property of the dual problem is that the coupling constraint in (11) is decoupled [18, Chapter 5].

A. The dual problem

Let the Lagrangian of (11) be defined on $\mathbf{A}_n \succeq 0$, $n = 1, \cdots, N$ and $\lambda \geq 0$ as in (12). The dual function $g(\lambda)$ is then computed as [17, Section 5.1]:

$$g(\lambda) = \max_{\mathbf{A}_1, \cdots, \mathbf{A}_N \succeq 0} \mathcal{L}(\mathbf{A}_1, \cdots, \mathbf{A}_N, \lambda), \tag{14}$$

while the solution of the dual problem is obtained from

$$\mathcal{C}' = \min_{\lambda \geq 0} g(\lambda). \tag{15}$$

Lemma 1: The duality gap for optimization (11) is zero, *i.e.,* the primal problem (11) and the dual problem (15) have the same solution.

Proof: The duality gap for problems of the form of (11), and satisfying the time-sharing property, is zero [23, Theorem 1]. Time-sharing property is defined as follows: let $\mathcal{C}_x, \mathcal{C}_y, \mathcal{C}_z$ be the solution of (11) for backhaul rates $\text{R}_x, \text{R}_y, \text{R}_z$, respectively. Consider $\text{R}_z = \nu \text{R}_x + (1 - \nu) \text{R}_y$ for some $0 \leq \nu \leq 1$. Then, the property is satisfied if and only if $\mathcal{C}_z \geq \nu \mathcal{C}_x + (1 - \nu) \mathcal{C}_y, \forall \nu \in [0, 1]$. That is, if the solution of (11) is concave with respect to the backhaul rate R. It is well known that time-sharing of compressions cannot decrease the resulting distortion [20, Lemma 13.4.1], neither improve the mutual information obtained from the reconstructed vectors[2]. Hence, the property holds for (11), and the duality gap is zero. ∎

We then solve the dual problem in order to obtain the solution of the primal. First, consider maximization (14). As expected, the maximization can not be solved in closed form. However, as the feasible set (*i.e.,* $\mathbf{A}_1, \cdots, \mathbf{A}_N \succeq 0$) is the cartesian product of convex sets, then a *block coordinate ascent* algorithm[3] can be used to search for the maximum [18, Section 2.7]. The algorithm iteratively optimizes the function with respect to one $\mathbf{A}_n$ while keeping the others fixed. It has

been previously used to *e.g.,* solve the sum-capacity problem of MIMO multiple access channels with individual and sum-power constraint [25] [26]. We define it for our problem as:

$$\mathbf{A}_n^{t+1} = \arg \max_{\mathbf{A}_n \succeq 0} \mathcal{L}\left(\mathbf{A}_1^{t+1}, \cdots, \mathbf{A}_{n-1}^{t+1}, \mathbf{A}_n, \right.$$

$$\left. \mathbf{A}_{n+1}^t, \cdots, \mathbf{A}_N^t, \lambda \right), \tag{16}$$

where t is the iteration index. As shown in Theorem 2, the maximization (16) is uniquely attained.

Theorem 2: Let the optimization $\mathbf{A}_n^* = \arg \max_{\mathbf{A}_n \succeq 0} \mathcal{L}(\mathbf{A}_1, \cdots, \mathbf{A}_N, \lambda)$ and the conditional covariance matrix (See Appendix A-A)

$$\mathbf{R}_{\mathbf{Y}_n | \mathbf{Y}_0, \hat{\mathbf{Y}}_n^c} = \mathbf{H}_{s,n} \left(\mathbf{I} + \mathbf{Q} \left(\frac{1}{\sigma_r^2} \mathbf{H}_{s,0}^\dagger \mathbf{H}_{s,0} \right. \right.$$

$$\left. \left. + \sum_{p \neq n} \mathbf{H}_{s,p}^\dagger \left(\mathbf{A}_p \sigma_r^2 \mathbf{I} + \mathbf{I} \right)^{-1} \mathbf{A}_p \mathbf{H}_{s,p} \right) \right)^{-1} \mathbf{Q} \mathbf{H}_{s,n}^\dagger + \sigma_r^2 \mathbf{I}$$

with eigen-decomposition $\mathbf{R}_{\mathbf{Y}_n | \mathbf{Y}_0, \hat{\mathbf{Y}}_n^c} = \mathbf{U}_n \mathbf{S} \mathbf{U}_n^\dagger$. The optimum is attained at $\mathbf{A}_n^* = \mathbf{U}_n \boldsymbol{\eta} \mathbf{U}_n^\dagger$, where

$$\eta_j = \left[\frac{1}{\lambda} \left(\frac{1}{\sigma_r^2} - \frac{1}{s_j} \right) - \frac{1}{\sigma_r^2} \right]^+, \quad j = 1, \cdots, M_n. \tag{17}$$

Proof: See Appendix C-A for the proof. ∎

Function $\mathcal{L}(\mathbf{A}_1, \cdots, \mathbf{A}_N, \lambda)$ is continuously differentiable and the maximization (16) is uniquely attained. Hence, the limit point of the sequence $\{\mathbf{A}_1^t, \cdots, \mathbf{A}_N^t\}$ is proven to converge to a stationary point [18, Proposition 2.7.1]. To demonstrate convergence to the global maximum, though, it would be necessary to show that the mapping $T(\mathbf{A}_1, \cdots, \mathbf{A}_N) = [\mathbf{A}_1 + \gamma \nabla_{\mathbf{A}_1} \mathcal{L}, \cdots, \mathbf{A}_N + \gamma \nabla_{\mathbf{A}_N} \mathcal{L}]$ is a block-contraction[4] for some γ [27, Proposition 3.10]. Unfortunately, we were not able to demonstrate the contraction property, although simulation results suggest global convergence of our algorithm always.

Once obtained $g(\lambda)$ through the Gauss-Seidel Algorithm[5], it remains to minimize it on $\lambda \geq 0$. First, recall that $g(\lambda)$ is a convex function by definition, since it is defined as the pointwise maximum of a family of affine functions [17]. Hence, to minimize it, we may use a subgradient approach as *e.g.,* that proposed by Yu in [26]. The subgradient search consists on following search direction $-h(\lambda)$ such that

$$\frac{g(\lambda') - g(\lambda)}{\lambda' - \lambda} \geq h(\lambda) \qquad \forall \lambda'. \tag{18}$$

Such a search is proven to converge to the global minimum for diminishing step-size rules [24, Section II-B]. Considering the definition of $g(\lambda)$, the following $h(\lambda)$ satisfies (18):

$$h(\lambda) = \text{R} - \log \det \left(\mathbf{I} + \text{diag} \left(\mathbf{A}_{1:N}(\lambda) \right) \mathbf{R}_{\mathbf{Y}_{1:N} | \mathbf{Y}_0} \right). \tag{19}$$

where $\mathbf{A}_{1:N}(\lambda)$ is the limit point of (16). Therefore, it is used to search for the optimum λ as:

$$\text{increase } \lambda \text{ if } h(\lambda) \leq 0 \quad \text{or} \quad \text{decrease } \lambda \text{ if } h(\lambda) > 0. \tag{20}$$

Consider now $\lambda^0 = 1$ as the initial value of the Lagrange multiplier. For such a multiplier, the optimum solution of (14)

---

[2]In the proof of [20, Lemma 13.4.1], optimal source coding is assumed. However, time-sharing distortion deterioration also holds when using suboptimal codes as ours.

[3]Also known as *Non-Linear Gauss-Seidel* algorithm [24, Section II-C].

[4]See [27, Section 3.1.2] for the definition of block-contraction.

[5]Assume hereafter that the algorithm has converged to the global maximum of $\mathcal{L}(\mathbf{A}_1, \cdots, \mathbf{A}_N, \lambda)$.

$$\mathcal{L}\left(\mathbf{A}_1, \cdots, \mathbf{A}_N, \lambda\right) = \log\det\left(\mathbf{I} + \frac{\mathbf{Q}}{\sigma_r^2}\mathbf{H}_{s,0}^\dagger\mathbf{H}_{s,0} + \mathbf{Q}\sum_{n=1}^N \mathbf{H}_{s,n}^\dagger\left(\mathbf{A}_n\sigma_r^2 + \mathbf{I}\right)^{-1}\mathbf{A}_n\mathbf{H}_{s,n}\right) \tag{12}$$
$$- \lambda \cdot \left(\log\det\left(\mathbf{I} + \mathrm{diag}\left(\mathbf{A}_1, \cdots, \mathbf{A}_N\right)\mathbf{R}_{\mathbf{Y}_{1:N}|\mathbf{Y}_0}\right) - \mathrm{R}\right).$$

$$\mathcal{R}_{\mathrm{D-WZ}} = \mathrm{coh}\left(\bigcup_{\substack{\mathbf{\Phi}_1, \cdots, \mathbf{\Phi}_N \\ \in c(\mathrm{R})}}\left\{R_{1,2}: \begin{matrix} R_1 \le \log\det\left(\mathbf{I} + \frac{\mathbf{Q}_1}{\sigma_r^2}\mathbf{H}_{1,0}^\dagger\mathbf{H}_{1,0} + \mathbf{Q}_1\sum_{n=1}^N \mathbf{H}_{1,n}^\dagger\left(\sigma_r^2\mathbf{I} + \mathbf{\Phi}_n\right)^{-1}\mathbf{H}_{1,n}\right) \\ R_2 \le \log\det\left(\mathbf{I} + \frac{\mathbf{Q}_2}{\sigma_r^2}\mathbf{H}_{2,0}^\dagger\mathbf{H}_{2,0} + \mathbf{Q}_2\sum_{n=1}^N \mathbf{H}_{2,n}^\dagger\left(\sigma_r^2\mathbf{I} + \mathbf{\Phi}_n\right)^{-1}\mathbf{H}_{2,n}\right) \\ R_1 + R_2 \le \log\det\left(\mathbf{I} + \frac{\mathbf{Q}}{\sigma_r^2}\mathbf{H}_{s,0}^\dagger\mathbf{H}_{s,0} + \mathbf{Q}\sum_{n=1}^N \mathbf{H}_{s,n}^\dagger\left(\sigma_r^2\mathbf{I} + \mathbf{\Phi}_n\right)^{-1}\mathbf{H}_{s,n}\right) \end{matrix}\right\}\right) \tag{13}$$

is $\{\mathbf{A}_1^*\cdots, \mathbf{A}_N^*\} = \mathbf{0}$ and the subgradient (19) is $h\left(\lambda^0\right) = $ R (See Appendix C-B). Hence, following (20), the optimum value of λ is strictly lower than one. Algorithm 1 takes all this into account in order to solve the dual problem, hence solving the primal too. As mentioned, we can only claim local convergence of the algorithm.

Algorithm 1 Multiple-BSs dual problem

1: Initialize $\lambda_{\min} = 0$ and $\lambda_{\max} = 1$
2: **repeat**
3:     $\lambda = \frac{\lambda_{\max} - \lambda_{\min}}{2}$
4:     Obtain $\{\hat{\mathbf{A}}_1^*, \cdots, \mathbf{A}_N^*\} = \arg\max\mathcal{L}\left(\mathbf{A}_1, \cdots, \mathbf{A}_N, \lambda\right)$ from Algorithm 2
5:     Evaluate $h\left(\lambda\right)$ as in (19).
6:     if $h\left(\lambda\right) \le 0$, then $\lambda_{\min} = \lambda$, else $\lambda_{\max} = \lambda$
7: **until** $\lambda_{\max} - \lambda_{\min} \le \epsilon$
8: $\{\mathbf{\Phi}_1^*, \cdots, \mathbf{\Phi}_N^*\} = \left\{\left(\mathbf{A}_1^*\right)^{-1}, \cdots, \left(\mathbf{A}_N^*\right)^{-1}\right\}$

Algorithm 2 Block-coordinate algorithm to obtain $g\left(\lambda\right)$

1: Initialize $\mathbf{A}_n^0 = \mathbf{0}$, $n = 1, \cdots, N$ and $t = 0$
2: **repeat**
3:     **for** $n = 1$ to N **do**
4:       Take $\mathbf{R}_{\mathbf{Y}_n|\mathbf{Y}_0, \hat{\mathbf{Y}}_n^c}\left(\mathbf{A}_1^{t+1}, \cdots, \mathbf{A}_{n-1}^{t+1}, \mathbf{A}_{n+1}^t, \cdots, \mathbf{A}_N^t\right)$ from (17).
5:       Compute its eigen-decomposition $\mathbf{U}_n\mathbf{S}\mathbf{U}_n^\dagger$ and eval-   uate $\boldsymbol{\eta}$ as in (17).
6:       Update $\mathbf{A}_n^{t+1} = \mathbf{U}_n\boldsymbol{\eta}\mathbf{U}_n^\dagger$.
7:     **end for**
8:     $t = t + 1$
9: **until** The sequence converges $\{\mathbf{A}_1^t, \cdots, \mathbf{A}_N^t\} \rightarrow$ $\{\mathbf{A}_1^*, \cdots, \mathbf{A}_N^*\}$
10: Return $\{\mathbf{A}_1^*, \cdots, \mathbf{A}_N^*\}$

V. THE MULTIPLE USER SCENARIO

In previous sections, we considered a single user within the network. To complement the analysis, we study hereafter multiple senders transmitting simultaneously. For simplicity, we consider two users, s_1 and s_2, transmitting two independent messages $\omega_u \in \{1, \cdots, 2^{nR_u}\}$, $u = 1, 2$, mapped onto codewords $\mathbf{X}_u^n$, $u = 1, 2$, respectively. Codewords are drawn *i.i.d.* from random vectors $\mathbf{X}_u \sim \mathcal{CN}\left(\mathbf{0}, \mathbf{Q_u}\right)$, $u = 1, 2$ and not subject to optimization. Hence, now, the BSs receive:

$$\mathbf{Y}_i^n = \sum_{u=1}^2 \mathbf{H}_{u,i}\mathbf{X}_u^n + \mathbf{N}_i^n, \quad i = 0, \cdots, N, \tag{21}$$

where $\mathbf{H}_{u,i}$ is the MIMO channel between user s_u and BS_i, and $\mathbf{N}_i \sim \mathcal{CN}\left(0, \sigma_r^2\mathbf{I}\right)$. As previously, signals at $\mathrm{BS}_1, \cdots, \mathrm{BS}_N$ are compressed using a D-WZ code and later sent to BS_0, which centralizes decoding.

Using previous arguments and considering the MIMO-MAC capacity region [20, Theorem 14.3.1], the set $\mathcal{R}_{\mathrm{D-WZ}}$ of transmission rate-duples $\left(R_1, R_2\right)$ that can be reliably decoded at the BS_0 is (13), where $c\left(\mathrm{R}\right) = \left\{\mathbf{\Phi}_{1:N} : \log\det\left(\mathbf{I} + \mathrm{diag}\left(\mathbf{\Phi}_1^{-1}, \cdots, \mathbf{\Phi}_N^{-1}\right)\mathbf{R}_{\mathbf{Y}_{1:N}|\mathbf{Y}_0}\right) \le \mathrm{R}\right\}$, $\mathbf{Q} = \mathrm{diag}\left(\mathbf{Q}_1, \mathbf{Q_2}\right)$ and $\mathbf{H}_{s,n} = [\mathbf{H}_{1,n}, \mathbf{H}_{2,n}]$, for $n = 0, \cdots, N$. Covariance $\mathbf{R}_{\mathbf{Y}_{1:N}|\mathbf{Y}_0}$ is calculated in Appendix A-B. The union in (13) is explained by the fact that compression codebooks might be arbitrarily chosen at the BSs. To evaluate such a region, we resort to the weighted sum-rate (WSR) optimization [28, Sec. III-C]. That is, we express

$$\mathcal{R}_{\mathrm{D-WZ}} = \{(R_1, R_2) : \alpha R_1 + (1 - \alpha)R_2 \le \tag{22}$$
$$\mathcal{R}\left(\alpha\right), \forall \alpha \in [0, 1]\},$$

with $\mathcal{R}\left(\alpha\right)$ the maximum WSR, given weights α and $(1 - \alpha)$ for user s_1 and s_2, respectively. Such a WSR is attained at the boundary of the region. It is easily shown that the boundary points of (13) can be achieved using successive interference cancellation (SIC) at the BS_0, and (optionally) *time-sharing* (TS). SIC consists of first decoding the user with lowest weight (*i.e.*, priority) considering the second user as interference. Later, once decoded the first user, the decoder subtracts its contribution to the received signal, and then decodes the second user without interference.

A. Useful Outer Regions

Prior to solving the WSR optimization, we present two outer regions on (13).

Outer Region 1: If $\left(R_1, R_2\right) \in \mathcal{R}_{\mathrm{D-WZ}}$, then

$$R_1 \le \log\det\left(\mathbf{I} + \frac{\mathbf{Q}_1}{\sigma_r^2}\sum_{n=0}^N \mathbf{H}_{1,n}^\dagger\mathbf{H}_{1,n}\right)$$
$$R_2 \le \log\det\left(\mathbf{I} + \frac{\mathbf{Q}_2}{\sigma_r^2}\sum_{n=0}^N \mathbf{H}_{2,n}^\dagger\mathbf{H}_{2,n}\right)$$
$$R_1 + R_2 \le \log\det\left(\mathbf{I} + \frac{\mathbf{Q}}{\sigma_r^2}\sum_{n=0}^N \mathbf{H}_{s,n}^\dagger\mathbf{H}_{s,n}\right)$$

Remark 3: This is the capacity region when $\mathbf{Y}_i$, $i = 1, \cdots, N$ are available at BS_0.

Outer Region 2: If $\left(R_1, R_2\right) \in \mathcal{R}_{\mathrm{D-WZ}}$ then

$$R_1 + R_2 \le \log\det\left(\mathbf{I} + \frac{1}{\sigma_r^2}\mathbf{H}_{s,0}\mathbf{Q}\mathbf{H}_{s,0}^\dagger\right) + \mathrm{R}.$$

Proof: It is equivalent to the proof of upper bound 2. ∎

B. Sum Rate Maximization

The maximum sum-rate of (13) is identical to the maximum transmission rate of a single user s transmitting a vector $\mathbf{X}_s = \left[\mathbf{X}_1^T, \mathbf{X}_2^T\right]^T$ over an equivalent channel $\mathbf{H}_{s,n} = [\mathbf{H}_{1,n}, \ \mathbf{H}_{2,n}]$, $n = 0, \cdots, N$. Hence, to obtain it we resort to Algorithm 1.

C. Weighted Sum Rate Maximization

Let consider the WSR optimization with $\alpha > \frac{1}{2}$ (*i.e.,* higher priority to user 1, which is decoded last at the SIC). With such a decoding, the rate of user 1 is then

$$R_1 = \log\det\left(\mathbf{I} + \frac{\mathbf{Q}_1}{\sigma_r^2}\mathbf{H}_{1,0}^\dagger\mathbf{H}_{1,0} + \right. \tag{23}$$
$$\left. \mathbf{Q}_1\sum_{n=1}^N \mathbf{H}_{1,n}^\dagger\left(\sigma_r^2\mathbf{I} + \mathbf{\Phi}_n\right)^{-1}\mathbf{H}_{1,n}\right).$$

On the other hand, the rate of user 2, which is decoded first, follows:

$$R_2 = \log\det\left(\mathbf{I} + \frac{\mathbf{Q}}{\sigma_r^2}\mathbf{H}_{s,0}^\dagger\mathbf{H}_{s,0} + \right. \tag{24}$$
$$\left. \mathbf{Q}\sum_{n=1}^N \mathbf{H}_{s,n}^\dagger\left(\sigma_r^2\mathbf{I} + \mathbf{\Phi}_n\right)^{-1}\mathbf{H}_{s,n}\right) - R_1,$$

where $\mathbf{Q} = \mathrm{diag}\left(\mathbf{Q}_1, \mathbf{Q}_2\right)$ and $\mathbf{H}_{s,n} = [\mathbf{H}_{1,n}, \ \mathbf{H}_{2,n}]$. The WSR, $\alpha R_1 + (1-\alpha)R_2$, which has to be maximized is convex on $\mathbf{\Phi}_1, \cdots, \mathbf{\Phi}_N$. To make it concave, we change the variables $\mathbf{\Phi}_n = \mathbf{A}_n^{-1}$, $n = 1, \cdots, N$. Then, plugging (23) and (24) into (22), the WSR optimization turns into

$$\mathcal{R}\left(\alpha\right) = \max_{\mathbf{A}_1, \cdots, \mathbf{A}_N} \quad \alpha \cdot R_1 + (1-\alpha) \cdot R_2 \tag{25}$$
$$\text{s.t.} \quad \log\det\left(\mathbf{I} + \mathrm{diag}\left(\mathbf{A}_{1:N}\right)\mathbf{R}_{\mathbf{Y}_{1:N}|\mathbf{Y}_0}\right) \le \mathrm{R}$$

As previously, the constraint does not define a convex feasible set. To solve the optimization, we follow the same strategy presented previously: first, we show that the optimization has zero duality gap. Later, we propose an iterative algorithm that solves the dual problem, thus solving the primal problem too.

Lemma 2: The duality gap for the WSR optimization (25) is zero.

Proof: Applying the time-sharing property in [23, Theorem 1] the zero-duality gap is demonstrated. ∎

Let us then solve the dual problem. The Lagrangian for optimization (25) is defined as:

$$\mathcal{L}_\alpha\left(\mathbf{A}_1, \cdots, \mathbf{A}_n, \lambda\right) = \alpha \cdot R_1 + (1-\alpha) \cdot R_2 - \tag{26}$$
$$\lambda \cdot \left(\log\det\left(\mathbf{I} + \mathrm{diag}\left(\mathbf{A}_{1:N}\right)\mathbf{R}_{\mathbf{Y}_{1:N}|\mathbf{Y}_0}\right) - \mathrm{R}\right)$$

The first step is to find the dual function [18, Section 5]

$$g_\alpha\left(\lambda\right) = \max_{\mathbf{A}_1, \cdots, \mathbf{A}_n \succeq 0} \mathcal{L}_\alpha\left(\mathbf{A}_1, \cdots, \mathbf{A}_n, \lambda\right) \tag{27}$$

In previous sections, we showed that such an optimization can be tackled using a block-coordinate algorithm. Unfortunately, now, the maximization with respect to a single $\mathbf{A}_n$ cannot be solved in closed-form and is not clear to be uniquely attained. Hence, to solve (27), we propose another algorithm: the *gradient projection method* (GP) [18, Section 2.3]. GP has been used to *e.g.,* compute transmit covariances for MIMO interference channels and the WSR of MIMO broadcast channels [29, Section IV-C] [30]. It is defined as follows: let (27),

and consider the initial point $\left\{\mathbf{A}_1^0, \cdots, \mathbf{A}_n^0\right\} \succeq 0$. It iteratively updates [18, Section 2.3.1]:

$$\mathbf{A}_n^{t+1} = \mathbf{A}_n^t + \gamma_t\left(\bar{\mathbf{A}}_n^t - \mathbf{A}_n^t\right), \quad n = 1, \cdots, N \tag{28}$$

where t is the iteration index and $0 < \gamma_t \le 1$ is the step size. Also,

$$\bar{\mathbf{A}}_n^t = \left[\mathbf{A}_n^t + s_t \cdot \nabla_{\mathbf{A}_n}\mathcal{L}_\alpha\left(\lambda, \mathbf{A}_1^t, \cdots, \mathbf{A}_N^t\right)\right]_{\succeq 0}, \tag{29}$$

with $s_t \ge 0$ an scalar and $\nabla_{\mathbf{A}_n}\mathcal{L}_\alpha\left(\lambda, \mathbf{A}_1^t, \cdots, \mathbf{A}_N^t\right)$ the gradient of $\mathcal{L}_\alpha\left(\cdot\right)$ with respect to $\mathbf{A}_n$, evaluated at $\mathbf{A}_1^t, \cdots, \mathbf{A}_N^t$. Finally, $[\cdot]_{\succeq 0}$ denotes the projection (with respect to the Frobenius norm) onto the cone of positive semidefinite matrices. Whenever γ_t and s_t are chosen appropriately, the sequence $\left\{\mathbf{A}_1^t, \cdots, \mathbf{A}_n^t\right\}$ is proven to converge to a stationary point of (27) [18, Proposition 2.2.1]. For global convergence to hold, the contraction property must be satisfied. Unfortunately, we were not able to prove this property for our optimization.

In order to make the algorithm work for the problem, we need to: *i*) compute the projection of a Hermitian matrix $\mathbf{S}$, with eigen-decomposition $\mathbf{S} = \mathbf{U}\boldsymbol{\eta}\mathbf{U}^\dagger$, onto the cone of positive semidefinite matrices. It is equal to [31, Theorem 2.1]:

$$[\mathbf{S}]_{\succeq 0} = \mathbf{U}\mathrm{diag}\left(\max\left\{\eta_1, 0\right\}, \cdots, \max\left\{\eta_m, 0\right\}\right)\mathbf{U}^\dagger. \tag{30}$$

ii) Obtain the gradient of $\mathcal{L}_\alpha\left(\cdot\right)$ with respect to a single $\mathbf{A}_n$, which is twice the conjugate of the partial derivative of the function with respect to such a matrix [19]:

$$\nabla_{\mathbf{A}_n}\mathcal{L}_\alpha\left(\mathbf{A}_{1:N}, \lambda\right) = 2\left(\left[\frac{\partial\mathcal{L}_\alpha\left(\mathbf{A}_{1:N}, \lambda\right)}{\partial\mathbf{A}_n}\right]^T\right)^\dagger \tag{31}$$

The Lagrangian is defined in (26). To obtain its partial derivative, we make use of (57):

$$\left[\frac{\partial\log\det\left(\mathbf{I} + \mathrm{diag}\left(\mathbf{A}_{1:N}\right)\mathbf{R}_{\mathbf{Y}_{1:N}|\mathbf{Y}_0}\right)}{\partial\mathbf{A}_n}\right]^T \tag{32}$$
$$= \left[\frac{\partial\log\det\left(\mathbf{I} + \mathbf{A}_n\mathbf{R}_{\mathbf{Y}_n|\mathbf{Y}_0, \hat{\mathbf{Y}}_n^c}\right)}{\partial\mathbf{A}_n}\right]^T$$
$$= \mathbf{R}_{\mathbf{Y}_n|\mathbf{Y}_0, \hat{\mathbf{Y}}_n^c}\left(\mathbf{I} + \mathbf{A}_n\mathbf{R}_{\mathbf{Y}_n|\mathbf{Y}_0, \hat{\mathbf{Y}}_n^c}\right)^{-1}.$$

The conditional covariance is computed in Appendix A-B. Furthermore, we can also derive that

$$\frac{\partial R_1}{\partial\mathbf{A}_n} = \frac{\partial I\left(\mathbf{X}_1; \mathbf{Y}_0, \hat{\mathbf{Y}}_{1:N}|\mathbf{X}_2\right)}{\partial\mathbf{A}_n} \tag{33}$$
$$= \frac{\partial I\left(\mathbf{X}_1; \hat{\mathbf{Y}}_n|\mathbf{X}_2, \mathbf{Y}_0, \hat{\mathbf{Y}}_n^c\right)}{\partial\mathbf{A}_n}$$

where second equality follows from the chain rule for mutual information and noting that $I\left(\mathbf{X}_1; \mathbf{Y}_0, \hat{\mathbf{Y}}_n^c|\mathbf{X}_2\right)$ does not depend on $\mathbf{A}_n$. The mutual information above is evaluated as:

$$I\left(\mathbf{X}_1; \hat{\mathbf{Y}}_n|\mathbf{X}_2, \mathbf{Y}_0, \hat{\mathbf{Y}}_n^c\right) \tag{34}$$
$$= H\left(\hat{\mathbf{Y}}_n|\mathbf{X}_2, \mathbf{Y}_0, \hat{\mathbf{Y}}_n^c\right) - H\left(\hat{\mathbf{Y}}_n|\mathbf{X}_1, \mathbf{X}_2, \mathbf{Y}_0, \hat{\mathbf{Y}}_n^c\right)$$
$$= \log\det\left(\mathbf{R}_{\mathbf{Y}_n|\mathbf{X}_2, \mathbf{Y}_0, \hat{\mathbf{Y}}_n^c} + \mathbf{\Phi}_n\right) - \log\det\left(\sigma_r^2\mathbf{I} + \mathbf{\Phi}_n\right)$$
$$= \log\det\left(\mathbf{A}_n\mathbf{R}_{\mathbf{Y}_n|\mathbf{X}_2, \mathbf{Y}_0, \hat{\mathbf{Y}}_n^c} + \mathbf{I}\right) - \log\det\left(\mathbf{A}_n\sigma_r^2 + \mathbf{I}\right)$$

Last equality follows from $\boldsymbol{\Phi}_n = \mathbf{A}_n^{-1}$, and $\mathbf{R}_{\mathbf{Y}_n|\mathbf{X}_2,\mathbf{Y}_0,\hat{\mathbf{Y}}_n^c}$ is computed in Appendix A-B. Therefore, the derivative of R_1 remains [19]

$$
\begin{aligned}
\left[\frac{\partial R_1}{\partial \mathbf{A}_n}\right]^T &= \mathbf{R}_{\mathbf{Y}_n|\mathbf{X}_2,\mathbf{Y}_0,\hat{\mathbf{Y}}_n^c}\left(\mathbf{A}_n \mathbf{R}_{\mathbf{Y}_n|\mathbf{X}_2,\mathbf{Y}_0,\hat{\mathbf{Y}}_n^c}+\mathbf{I}\right)^{-1} \\
&\quad - \sigma_r^2\left(\mathbf{A}_n\sigma_r^2+\mathbf{I}\right)^{-1}.
\end{aligned}
\tag{35}
$$

Equivalently, we can obtain for the derivative of R_2 that

$$
\begin{aligned}
\frac{\partial R_2}{\partial \mathbf{A}_n} &= \frac{\partial I\left(\mathbf{X}_2;\mathbf{Y}_0,\hat{\mathbf{Y}}_{1:N}\right)}{\partial \mathbf{A}_n} \tag{36} \\
&= \frac{\partial I\left(\mathbf{X}_2;\hat{\mathbf{Y}}_n|\mathbf{Y}_0,\hat{\mathbf{Y}}_n^c\right)}{\partial \mathbf{A}_n}.
\end{aligned}
$$

Where we evaluate:

$$
\begin{aligned}
I&\left(\mathbf{X}_2;\hat{\mathbf{Y}}_n|\mathbf{Y}_0,\hat{\mathbf{Y}}_n^c\right) \tag{37} \\
&= H\left(\hat{\mathbf{Y}}_n|\mathbf{Y}_0,\hat{\mathbf{Y}}_n^c\right) - H\left(\hat{\mathbf{Y}}_n|\mathbf{X}_2,\mathbf{Y}_0,\hat{\mathbf{Y}}_n^c\right) \\
&= \log\det\left(\mathbf{A}_n\mathbf{R}_{\mathbf{Y}_n|\mathbf{Y}_0,\hat{\mathbf{Y}}_n^c}+\mathbf{I}\right) \\
&\quad - \log\det\left(\mathbf{A}_n\mathbf{R}_{\mathbf{Y}_n|\mathbf{X}_2,\mathbf{Y}_0,\hat{\mathbf{Y}}_n^c}+\mathbf{I}\right)
\end{aligned}
$$

Conditional covariances are obtained in Appendix A-B. The derivative of R_2 thus remains:

$$
\begin{aligned}
\left[\frac{\partial R_2}{\partial \mathbf{A}_n}\right]^T &= \mathbf{R}_{\mathbf{Y}_n|\mathbf{Y}_0,\hat{\mathbf{Y}}_n^c}\left(\mathbf{A}_n\mathbf{R}_{\mathbf{Y}_n|\mathbf{Y}_0,\hat{\mathbf{Y}}_n^c}+\mathbf{I}\right)^{-1} \tag{38} \\
&\quad - \mathbf{R}_{\mathbf{Y}_n|\mathbf{X}_2,\mathbf{Y}_0,\hat{\mathbf{Y}}_n^c}\left(\mathbf{A}_n\mathbf{R}_{\mathbf{Y}_n|\mathbf{X}_2,\mathbf{Y}_0,\hat{\mathbf{Y}}_n^c}+\mathbf{I}\right)^{-1}.
\end{aligned}
$$

Plugging (32), (35) and (38) into (31) we obtain the gradient of the function. The gradient can be shown to be Hermitian by a straightforward application of Lemma 3 below and, therefore, the projection defined by (30) holds.

Lemma 3: Let $\mathbf{A}, \mathbf{B}$ be Hermitian matrices, with $\mathbf{B}$ non-singular. Then $\mathbf{B}(\mathbf{I}+\mathbf{A}\mathbf{B})^{-1}$ is Hermitian.

Proof: It is straightforward since $\mathbf{B}(\mathbf{I}+\mathbf{A}\mathbf{B})^{-1} = \left(\mathbf{B}^{-1}+\mathbf{A}\right)^{-1}$ and the sum and inverse of Hermitian matrices is an Hermitian matrix. ∎

The gradient is then used in the GP algorithm to obtain $g_\alpha(\lambda)$. Notice that for $\alpha \leq \frac{1}{2}$, the roles of users s_1 and s_2 are interchanged, being user 1 decoded first. These roles would also need to be interchanged in the computation of the gradients of R_1 and R_2. Once obtained the dual function, we minimize it:

$$
\mathcal{R}(\alpha) = \min_{\lambda \geq 0} g_\alpha(\lambda). \tag{39}
$$

To solve this minimization, we use the subgradient approach as in Section IV. Taking all this into account we build up Algorithm 3. As for the previous section, we can only claim local convergence.

VI. NUMERICAL RESULTS

We evaluate the performance of D-WZ coding within a realistic single-frequency network, composed of a central base station BS_0 plus its first tier of six cells. The radius of each cell is 700 m, and BSs have all three receive antennas. On the transmit side, users have two antennas and are located at the

---

Algorithm 3 Two-user WSR dual problem

1: Initialize $\lambda_{\min} = 0$ and $\lambda_{\max}$
2: **repeat**
3: $\lambda = \frac{\lambda_{\max}-\lambda_{\min}}{2}$
4: Obtain $\{\hat{\mathbf{A}}_{1:N}^*\} = \arg\max \mathcal{L}_\alpha(\mathbf{A}_{1:N},\lambda)$ from Algorithm 4
5: Evaluate $h(\lambda)$ as in (19), where $\mathbf{R}_{\mathbf{Y}_{1:N}|\mathbf{Y}_0}$ follows Appendix A-B.
6: if $h(\lambda) \leq 0$, then $\lambda_{\min} = \lambda$, else $\lambda_{\max} = \lambda$
7: **until** $\lambda_{\max} - \lambda_{\min} \leq \epsilon$
8: $\mathcal{R}(\alpha) = \alpha R_1(\mathbf{A}_{1:N}^*) + (1-\alpha)R_2(\mathbf{A}_{1:N}^*)$.

Algorithm 4 GP to obtain $g_\alpha(\lambda)$

1: Initialize $\mathbf{A}_n^0 = \mathbf{0}$, $n = 1,\cdots,N$ and $t = 0$
2: **repeat**
3: Compute the gradient $\mathbf{G}_n^t = \nabla_{\mathbf{A}_n}\mathcal{L}_\alpha(\lambda,\mathbf{A}_{1:N}^t)$, $n = 1,\cdots,N$ from (31).
4: Choose appropriate s_t
5: Set $\hat{\mathbf{A}}_n^t = \mathbf{A}_n^t + s_t \cdot \mathbf{G}_n^t$.
6: Eigen-decompose $\hat{\mathbf{A}}_n^t = \mathbf{U}_n\boldsymbol{\eta}\mathbf{U}_n^\dagger$, and set $\bar{\mathbf{A}}_n^t = \mathbf{U}_n \max\{\boldsymbol{\eta},0\}\mathbf{U}_n^\dagger$, $n = 1,\cdots,N$.
7: Choose appropriate γ_t
8: Update $\mathbf{A}_n^{t+1} = \mathbf{A}_n^t + \gamma_t\left(\bar{\mathbf{A}}_n^t - \mathbf{A}_n^t\right)$, $n = 1,\cdots,N$
9: $t = t + 1$
10: **until** The sequence converges $\{\mathbf{A}_{1:N}^t\} \to \mathbf{A}_{1:N}^*$
11: Return $\{\mathbf{A}_1^*,\cdots,\mathbf{A}_N^*\}$

---

edge of the central cell. Wireless channels are simulated taking into account path loss, log-normal shadowing and Rayleigh fading. Specifically, fading is assumed *i.i.d.* among antennas, and shadowing uncorrelated among BSs. Two propagation scenarios are studied: *i)* Line-of-sight (LOS), with path-loss exponent $\alpha = 2.6$ and shadowing standard deviation $\beta = 4$ dB, and *ii)* Non Line-of-sight (N-LOS), with $\alpha = 4.05$ and $\beta = 10$ dB. Channel matrices thus follow

$$
\mathbf{H}_{s,i} = \left(\frac{1}{d_{s,i}}\right)^{\frac{\alpha}{2}} \cdot \sqrt{h_{sh}} \cdot \mathbf{H}_{mp}^i. \tag{40}
$$

where $10\log_{10}(h_{sh}) \sim \mathcal{N}(0,\beta^2)$ and $\left[\mathbf{H}_{mp}^i\right]_{r,c} \sim \mathcal{CN}(0,1)$. Users transmit isotropically (*i.e.* $\mathbf{Q}_i = \frac{\mathrm{P}_{TX}}{2}\mathbf{I}$) with a transmitted power $\mathrm{P}_{TX} = 23$ dBm. Their symbol rate is set to 1 Msymb/s, occupying a total bandwidth $B = 1$ MHz. Base stations have a Noise Figure $F = 4$ dB. Therefore, the noise power at the receivers is $\sigma_r^2 = k \cdot T_o \cdot B \cdot F$, with k the Boltzmann constant and $T_0 = 290$ K.

Fig. 1 plots the cumulative density function (cdf) of the uplink rate for a single-user network, considering different values of the backhaul rate R [Mbit/s][6]. In particular, Fig. 1(a) depicts results for LOS propagation, and shows gains up to 6 Mbit/s @ 5% probability, with R = 15 Mbit/s. It is clearly shown that BSs cooperation becomes more remarkable for lower probabilities. On the other hand, Fig. 1(b) shows results for N-LOS propagation, where rate gains are reduced.

---

[6]Throughout the paper, the backhaul rate has been measured in [bits/symbol]. Its translation into [bit/s] is trivial by noting that the symbol rate is 1 Msymb/s.
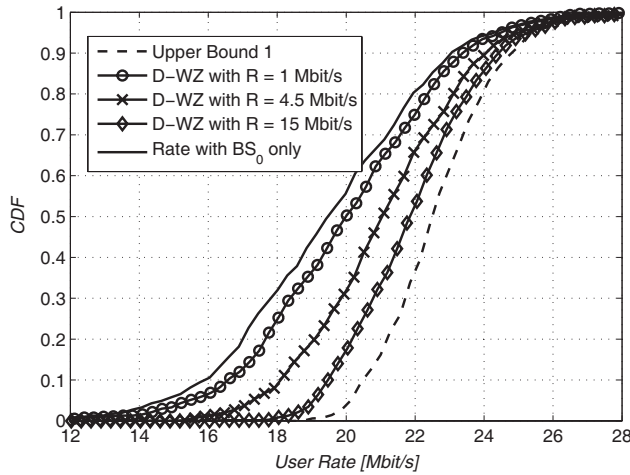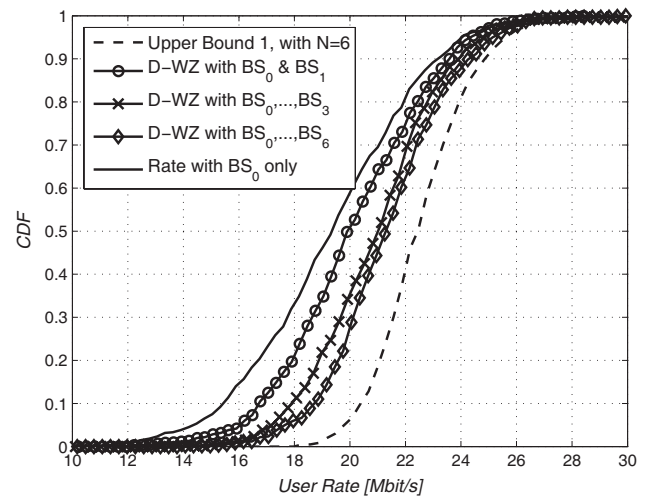
(a) CDF of the user rate. LOS conditions
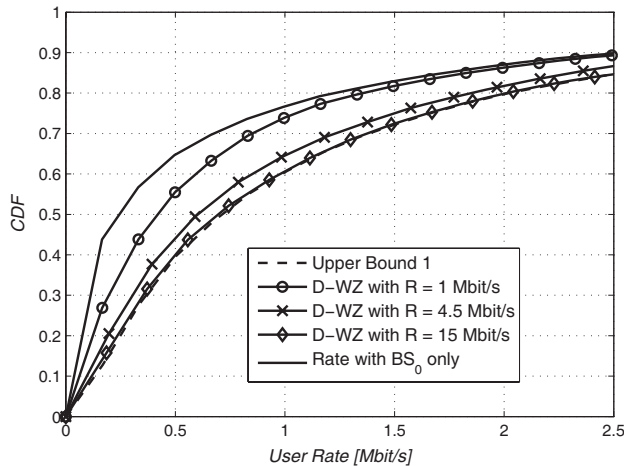
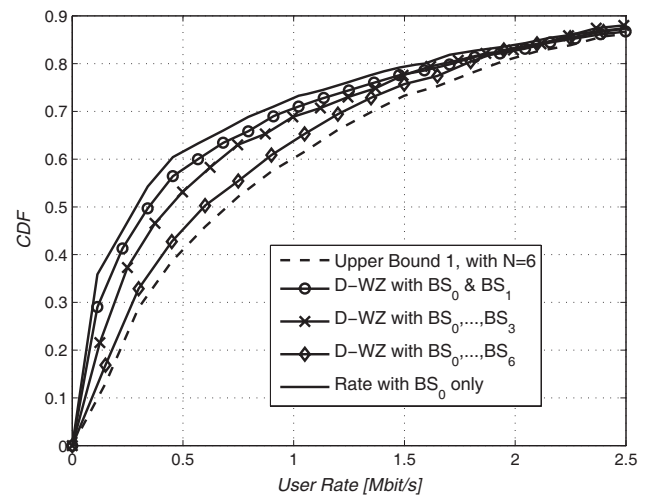(b) CDF of the user rate. N-LOS conditions

Fig. 1. Single user capacity results for different values of the backhaul rate R. $BS_1, \cdots, BS_6$ cooperate with BS_0.

(a) CDF of the user rate. LOS conditions

(b) CDF of the user rate. N-LOS conditions

Fig. 2. Single user capacity results, for different number of Cooperative BS, N. Backhaul rate R = 7 Mbit/s.

In this case, cooperation becomes more convenient for higher probabilities, showing that @ 50%, three-fold gains arise with 15 Mbit/s of backhaul.

Fig 2 plots the uplink rate of a single-user network with R = 7 Mbit/s, for different number N of cooperative BSs. First, Fig. 2(a) depicts the cdf of the user's rate under LOS propagation conditions. We notice that @ 5%, with only 1 cooperative BS, a rate gain of 2 Mbit/s is obtained with respect to the non-cooperative case. However, when increasing the number of cooperative BSs to 6, only an additional rate gain of 2 Mbit/s is obtained. That is, the impact of introducing new cooperative BSs in the system diminishes as the network grows. Again, cooperation is more useful for low probabilities. On the other hand, Fig. 2(b) depicts results for N-LOS propagation. It can be shown that, @ 50%, the rate is doubled from 1 cooperative BS to 6 cooperative BS. This fact highlights the relevant role of macro-diversity on N-LOS conditions, which are most common ones on urban cellular networks. Next, Fig. 3 compares the rate performance of our D-WZ approach with respect to that of *Quantization* [8], assuming LOS propagation. We consider a simple network with two BSs: BS_0 and BS_1, and plot its outage capacity with D-WZ and with uniform quantization, respectively. Both are normalized with respect to

the outage capacity with infinite backhaul and computed at a probability of outage of 10^{-2}. Results show significant gains, of up to 12%, for low backhaul rates, and highlights the fact that D-WZ requires half of backhaul rate than *Quantization* to converge to the ∞ backhaul capacity.

Fig 4 depicts the expected sum-rate of the multi-user setup versus the total number of users. Expectation is taken over the joint channel distribution via Monte-Carlo. Results are shown for different values of the backhaul rate. Although the sum-rate analysis (see Sec. V-B) was carried out for two users only, the extension to $U > 2$ is straightforward. Fig 4(a) depicts the sum-rate for LOS propagation. We first notice that the sum rate with ∞ backhaul (*i.e.,* outer region 1) is far lower than that with D-WZ compression. This is explained by means of outer region 2: the sum-rate of the system is constrained by the available rate at the backhaul network. On the other hand, for N-LOS propagation (Fig. 4(b)), upper bound 2 is not reached. Indeed, for less than 5 users, the expected sum-rate with only R = 15 Mbit/s of backhaul is almost identical to that of R = ∞. Therefore, for practical number of transmitters, the full rate gain due to macro-diversity is obtained via D-WZ compression. Finally, Fig. 5(a) and Fig. 5(b) depict the rate
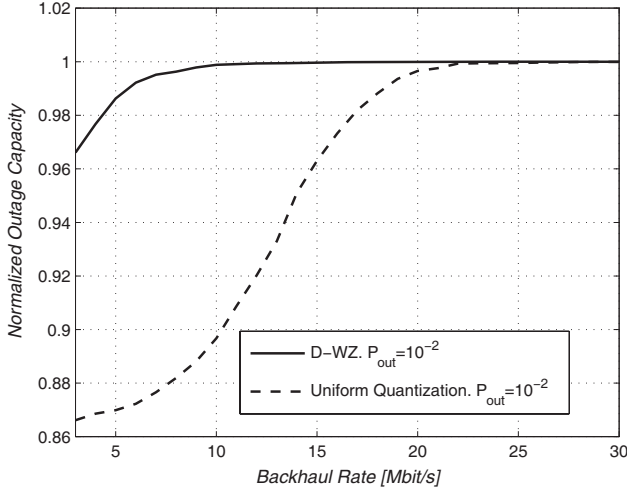
Fig. 3. Outage Capacity with D-WZ and with Quantization, respectively, versus the backhaul rate R. LOS.

region of a 2-user network, with and without LOS respectively, for different values of the backhaul rate R. It is shown that the region is significantly enlarged.

VII. CONCLUSIONS

In this paper, Distributed Wyner-Ziv coding has been proposed as an effective means to deploy receive cooperation among the BSs of a single-frequency network. This aims at both increasing the network capacity and the achievable sum-rate/backhaul rate tradeoff of coordinated networks. Considering MIMO BSs, the application of such a coding gave rise to a compression noise covariance optimization. The optimization was solved in this paper for a single user and multiple users, respectively. Significant capacity gains were shown, much greater than those of the distributed quantization which does not exploit signal correlation [8].

APPENDIX A

We present conditional covariances used along the paper.
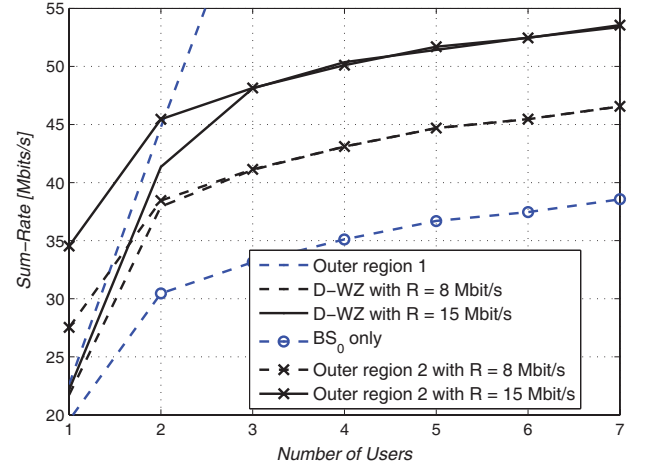
A. The single user case

$$\mathbf{R}_{\mathbf{Y}_n|\mathbf{Y}_0} = \mathbf{H}_{s,n} \left(\mathbf{I} + \frac{\mathbf{Q}}{\sigma_r^2} \mathbf{H}_{s,0}^\dagger \mathbf{H}_{s,0} \right)^{-1} \mathbf{Q} \mathbf{H}_{s,n}^\dagger + \sigma_r^2 \mathbf{I} \quad (41)$$

$$\mathbf{R}_{\mathbf{Y}_{1:N}|\mathbf{Y}_0} = \begin{bmatrix} \mathbf{H}_{s,1} \\ \vdots \\ \mathbf{H}_{s,N} \end{bmatrix} \left(\mathbf{I} + \frac{\mathbf{Q}}{\sigma_r^2} \mathbf{H}_{s,0}^\da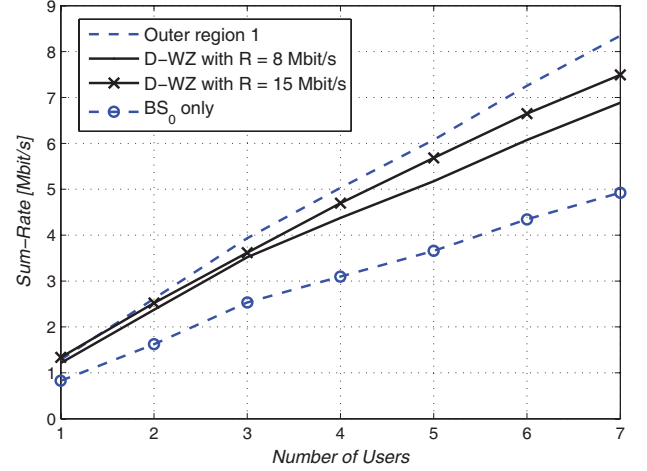gger \mathbf{H}_{s,0} \right)^{-1} \\ \times \mathbf{Q} \begin{bmatrix} \mathbf{H}_{s,1} \\ \vdots \\ \mathbf{H}_{s,N} \end{bmatrix}^\dagger + \sigma_r^2 \mathbf{I} \quad (42)$$

$$\mathbf{R}_{\mathbf{Y}_n|\mathbf{Y}_0,\hat{\mathbf{Y}}_n^c} = \mathbf{H}_{s,n} \left(\mathbf{I} + \frac{\mathbf{Q}}{\sigma_r^2} \mathbf{H}_{s,0}^\dagger \mathbf{H}_{s,0} \right. \quad (43)$$
$$\left. + \sum_{j \neq n} \mathbf{Q} \mathbf{H}_{s,j}^\dagger \left(\sigma_r^2 \mathbf{I} + \mathbf{\Phi}_j \right)^{-1} \mathbf{H}_{s,j} \right)^{-1} \mathbf{Q} \mathbf{H}_{s,n}^\dagger + \sigma_r^2 \mathbf{I}.$$

(a) LOS propagation

(b) N-LOS propagation

Fig. 4. Sum-rate versus number of users. $\mathrm{BS}_1, \cdots, \mathrm{BS}_6$ cooperate.

$$\mathbf{R}_{\mathbf{Y}_n|\mathbf{Y}_0,\hat{\mathbf{Y}}_\mathcal{G}} = \mathbf{H}_{s,n} \left(\mathbf{I} + \frac{\mathbf{Q}}{\sigma_r^2} \mathbf{H}_{s,0}^\dagger \mathbf{H}_{s,0} \right. \quad (44)$$
$$\left. + \sum_{j \in \mathcal{G}} \mathbf{Q} \mathbf{H}_{s,j}^\dagger \left(\sigma_r^2 \mathbf{I} + \mathbf{\Phi}_j \right)^{-1} \mathbf{H}_{s,j} \right)^{-1} \mathbf{Q} \mathbf{H}_{s,n}^\dagger + \sigma_r^2 \mathbf{I}$$

B. The multiuser case

Define $\mathbf{H}_{s,n} = [\mathbf{H}_{1,n}, \mathbf{H}_{2,n}]$ and $\mathbf{Q} = \mathrm{diag}(\mathbf{Q}_1, \mathbf{Q}_2)$. Then, Conditional covariances $\mathbf{R}_{\mathbf{Y}_n|\mathbf{Y}_0}$, $\mathbf{R}_{\mathbf{Y}_{1:N}|\mathbf{Y}_0}$ $\mathbf{R}_{\mathbf{Y}_n|\mathbf{Y}_0,\hat{\mathbf{Y}}_n^c}$ and $\mathbf{R}_{\mathbf{Y}_n|\mathbf{Y}_0,\hat{\mathbf{Y}}_\mathcal{G}}$ follow Subsection A-A. Furthermore, let $i, j \in \{1, 2\}$ with $j \neq i$, then:

$$\mathbf{R}_{\mathbf{Y}_n|\mathbf{X}_i,\mathbf{Y}_0,\hat{\mathbf{Y}}_n^c} = \mathbf{H}_{j,n} \left(\mathbf{I} + \frac{\mathbf{Q}_j}{\sigma_r^2} \mathbf{H}_{j,0}^\dagger \mathbf{H}_{j,0} \right. \quad (45)$$
$$\left. + \sum_{p \neq n} \mathbf{Q}_j \mathbf{H}_{j,p}^\dagger \left(\sigma_r^2 \mathbf{I} + \mathbf{\Phi}_p \right)^{-1} \mathbf{H}_{j,p} \right)^{-1} \mathbf{Q}_j \mathbf{H}_{j,n}^\dagger + \sigma_r^2 \mathbf{I}$$

APPENDIX B

In this Appendix, we solve the non-convex optimization (7). Let us first expand:
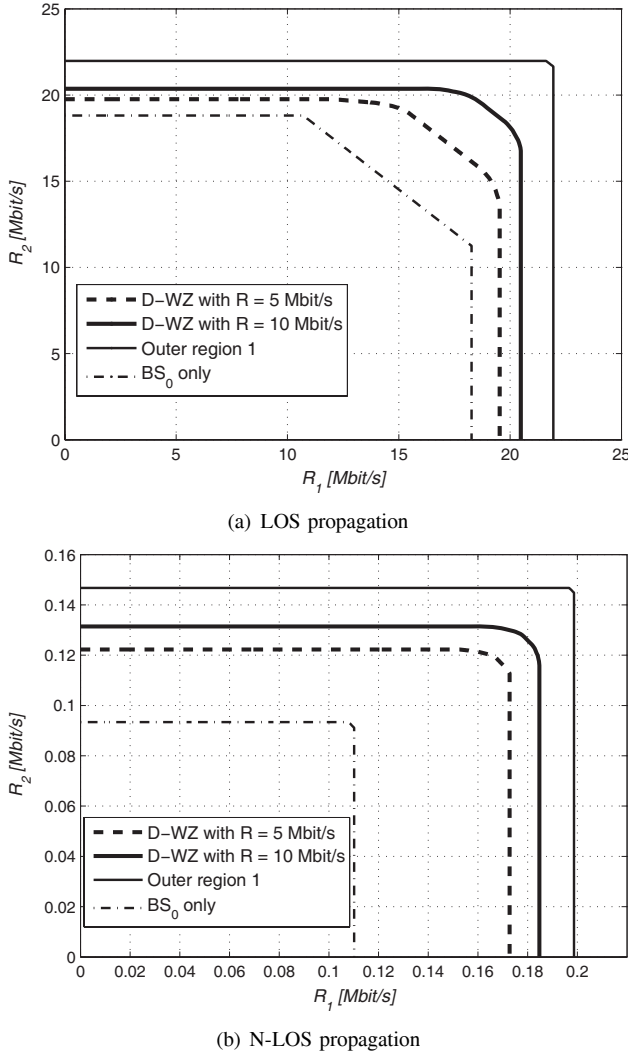
(a) LOS propagation

(b) N-LOS propagation

Fig. 5. Rate region for different values of R. $BS_1, \cdots, BS_6$ cooperate with BS_0.

$$\log \det \left(\mathbf{I} + \frac{\mathbf{Q}}{\sigma_r^2} \mathbf{H}_{s,0}^\dagger \mathbf{H}_{s,0} + \mathbf{Q} \mathbf{H}_{s,1}^\dagger \left(\mathbf{A}_1 \sigma_r^2 + \mathbf{I} \right)^{-1} \mathbf{A}_1 \mathbf{H}_{s,1} \right) \quad (46)$$

$$= \log \det \left(\mathbf{I} + \frac{\mathbf{Q}}{\sigma_r^2} \mathbf{H}_{s,0}^\dagger \mathbf{H}_{s,0} \right)$$
$$+ \log \det \left(\mathbf{I} + \left(\mathbf{A}_1 \sigma_r^2 + \mathbf{I} \right)^{-1} \mathbf{A}_1 \left(\mathbf{R}_{\mathbf{Y}_1|\mathbf{Y}_0} - \sigma_r^2 \mathbf{I} \right) \right)$$
$$= \log \det \left(\mathbf{I} + \frac{\mathbf{Q}}{\sigma_r^2} \mathbf{H}_{s,0}^\dagger \mathbf{H}_{s,0} \right)$$
$$+ \log \det \left(\mathbf{I} + \mathbf{A}_1 \mathbf{R}_{\mathbf{Y}_1|\mathbf{Y}_0} \right) - \log \det \left(\mathbf{I} + \mathbf{A}_1 \sigma_r^2 \right).$$

First equality follows from the value of $\mathbf{R}_{\mathbf{Y}_1|\mathbf{Y}_0}$ in (41). Notice that $\log \det \left(\mathbf{I} + \frac{\mathbf{Q}}{\sigma_r^2} \mathbf{H}_{s,0}^\dagger \mathbf{H}_{s,0} \right)$ does not depend on $\mathbf{A}_1$. Therefore, the Lagrangian for the problem can be written as

$$\mathcal{L} \left(\mathbf{A}_1, \lambda, \mathbf{\Upsilon} \right) = (1 - \lambda) \log \det \left(\mathbf{I} + \mathbf{A}_1 \mathbf{R}_{\mathbf{Y}_1|\mathbf{Y}_0} \right) \quad (47)$$
$$- \log \det \left(\mathbf{I} + \mathbf{A}_1 \sigma_r^2 \right) + \lambda \mathbf{R} - \mathrm{tr} \left\{ \mathbf{\Upsilon} \mathbf{A}_1 \right\},$$

where λ is the Lagrange multiplier for the explicit constraint and $\mathbf{\Upsilon} \preceq 0$ for the semidefinite positiveness constraint. The derivative of the Lagrangian with respect to $\mathbf{A}_1$ thus reads

[19]:

$$\left[\frac{\partial \mathcal{L}}{\partial \mathbf{A}_1} \right]^T = (1 - \lambda) \mathbf{R}_{\mathbf{Y}_1|\mathbf{Y}_0} \left(\mathbf{I} + \mathbf{A}_1 \mathbf{R}_{\mathbf{Y}_1|\mathbf{Y}_0} \right)^{-1} \quad (48)$$
$$- \sigma_r^2 \left(\mathbf{I} + \mathbf{A}_1 \sigma_r^2 \right)^{-1} - \mathbf{\Upsilon}.$$

Accordingly, the KKT conditions for the problem, which are necessary but not sufficient, are:

i) $\quad \left[\frac{\partial \mathcal{L}}{\partial \mathbf{A}_1} \right]^T = \mathbf{0}$ \hfill (49)

ii) $\quad \lambda \left(\log \det \left(\mathbf{I} + \mathbf{A}_1 \mathbf{R}_{\mathbf{Y}_1|\mathbf{Y}_0} \right) - \mathbf{R} \right) = 0$

iii) $\quad \mathrm{tr} \left\{ \mathbf{\Upsilon} \mathbf{A}_1 \right\} = 0.$

Let now the eigen-decomposition $\mathbf{R}_{\mathbf{Y}_1|\mathbf{Y}_0} = \mathbf{U} \mathbf{S} \mathbf{U}^\dagger$. Then, it can be readily shown that matrix $\mathbf{A}_1^* = \mathbf{U} \mathrm{diag} \left(\eta_1, \cdots, \eta_{M_1} \right) \mathbf{U}^\dagger$, where

$$\eta_j = \left[\frac{1}{\lambda^*} \left(\frac{1}{\sigma_r^2} - \frac{1}{s_j} \right) - \frac{1}{\sigma_r^2} \right]^+, \quad (50)$$

satisfies the KKT conditions, with multiplier λ^* such that $\sum_{j=1}^{M_1} \log \left(1 + \eta_j s_j \right) = \mathbf{R}$ (therefore, $\lambda^* < 1$), and multiplier $\mathbf{\Upsilon}^* \preceq 0$ computed from *i)*. Let now show that $\mathbf{A}_1^*$ satisfies also the general sufficiency condition for optimality, which is presented in the next Lemma.

Lemma 4: [18, Proposition 3.3.4] Let the maximization (7). Consider a pair $\left(\mathbf{A}_1^*, \lambda^* \right)$ for which $\lambda^* \left(\log \det \left(\mathbf{I} + \mathbf{A}_1^* \mathbf{R}_{\mathbf{Y}_1|\mathbf{Y}_0} \right) - \mathbf{R} \right) = 0$. Then, $\mathbf{A}_1^*$ is the global maximum of (7) if:

$$\mathbf{A}_1^* \in \arg \max_{\mathbf{A}_1 \succeq 0} \mathcal{L} \left(\mathbf{A}_1, \lambda^* \right), \quad (51)$$

where the Lagrangian[7] has been defined in (47).

Lemma 5: Let $\mathbf{A}, \mathbf{B} \succeq 0$, with ordered eigenvalues $\mathbf{\Gamma}_A, \mathbf{\Gamma}_B$ respectively. Then,

$$\log \det \left(\mathbf{I} + \mathbf{A} \mathbf{B} \right) \leq \log \det \left(\mathbf{I} + \mathbf{\Gamma}_A \mathbf{\Gamma}_B \right), \quad (52)$$

with equality whenever $\mathbf{A}$ and $\mathbf{B}$ have conjugate transpose eigenvectors.

Proof: It is known that $\log \det \left(\mathbf{I} + \mathbf{A} \mathbf{B} \right) = \log \det \left(\mathbf{I} + \mathbf{\Gamma}_{AB} \right)$, where $\mathbf{\Gamma}_{AB}$ are the ordered eigenvalues of $\mathbf{A} \mathbf{B}$. Those eigenvalues are logarithmically majorized [32, Definition 1.4] by the product of the separate eigenvalues of $\mathbf{A}$ and $\mathbf{B}$, i.e., $\mathbf{\Gamma}_{AB} \prec_\times \mathbf{\Gamma}_A \mathbf{\Gamma}_B$ [33, Theorem 9.H.1.d]. Let now the function $f(\mathbf{X}) = \log \det \left(\mathbf{I} + \mathbf{X} \right)$ be defined on the set of semi-definite positive diagonal matrices, *i.e.,* $f(\mathbf{X}) = \sum \log \left(1 + x_i \right)$. We may apply [32, Theorem 1.6] to prove that $f(\mathbf{X})$ is a Schur-geometrically-convex function. Accordingly, provided that $\mathbf{\Gamma}_{AB} \prec_\times \mathbf{\Gamma}_A \mathbf{\Gamma}_B$, then $\log \det \left(\mathbf{I} + \mathbf{\Gamma}_{AB} \right) \leq \log \det \left(\mathbf{I} + \mathbf{\Gamma}_A \mathbf{\Gamma}_B \right)$, which concludes the proof. ∎

Let us prove now that our pair $\left(\mathbf{A}_1^*, \lambda^* \right)$ satisfies (51). The Lagrangian is defined for the problem as

$$\mathcal{L} \left(\mathbf{A}_1, \lambda^* \right) = (1 - \lambda^*) \log \det \left(\mathbf{I} + \mathbf{A}_1 \mathbf{R}_{\mathbf{Y}_1|\mathbf{Y}_0} \right) \quad (53)$$
$$- \log \det \left(\mathbf{I} + \mathbf{A}_1 \sigma_r^2 \right) + \lambda^* \mathbf{R}.$$

---

[7]Notice that multiplier $\mathbf{\Upsilon}$ has been removed of the Lagrangian by constraining the maximization (51) to $\mathbf{A}_1 \succeq 0$.

Recall that $\lambda^* < 1$ and $\mathbf{R}_{\mathbf{Y}_1|\mathbf{Y}_0} = \mathbf{U}\mathbf{S}\mathbf{U}^\dagger$. Then, using Lemma 5 we can bound:

$$
\max_{\mathbf{A}_1 \succeq 0} \mathcal{L}(\mathbf{A}_1, \lambda^*) \leq \max_{\boldsymbol{\eta} \succeq 0} (1 - \lambda^*) \log \det (\mathbf{I} + \boldsymbol{\eta}\mathbf{S})
$$
$$
- \log \det (\mathbf{I} + \boldsymbol{\eta}\sigma_r^2) + \lambda^*\mathbf{R}
$$
$$
= \lambda^*\mathbf{R} + \sum_{j=1}^{M_1} \max_{\eta_j \geq 0} (1 - \lambda^*) \log (1 + \eta_j s_j)
$$
$$
- \log (1 + \eta_j \sigma_r^2) \tag{54}
$$

where $\boldsymbol{\eta}$ is the diagonal matrix of ordered eigenvalues of $\mathbf{A}_1$. The individual maximizations on η_j in (54) are not concave. However, the continuously differentiable functions $f_j(\eta_j) = (1 - \lambda^*) \log (1 + \eta_j s_j) - \log (1 + \eta_j \sigma_r^2)$ have only one stationary point, namely:

$$
\frac{df_j}{d\eta_j} = 0 \rightarrow \eta_j^* = \frac{1}{\lambda^*} \left(\frac{1}{\sigma_r^2} - \frac{1}{s_j} \right) - \frac{1}{\sigma_r^2}. \tag{55}
$$

For the stationary point, we can prove that its second derivative exists and is lower than zero; accordingly, it is a local maximum of the function, unique because there is no other. Moreover, it is easy to obtain that: i) $f_j(0) = 0$, and ii) since $\lambda < 1$, then $\lim_{\eta_j \to \infty} f_j(\eta_j) = -\infty$. That is, $\eta_j = \infty$ is the global minimum of the problem. Making use of i) and ii), we can claim that the local maximum η_j^* is the global maximum. However, we restricted the optimization to the values $\eta_j \geq 0$. Hence, functions $f_j(\eta_j)$ take maximum at:

$$
\eta_j^* = \left[\frac{1}{\lambda^*} \left(\frac{1}{\sigma_r^2} - \frac{1}{s_j} \right) - \frac{1}{\sigma_r^2} \right]^+. \tag{56}
$$

Plugging these optimal values into (54), we bound

$$
\max_{\mathbf{A}_1 \succeq 0} \mathcal{L}(\mathbf{A}_1, \lambda^*) \leq \lambda^*\mathbf{R} + (1 - \lambda^*) \sum_{j=1}^{M_1} \log (1 + \eta_j^* s_j)
$$
$$
- \sum_{i=1}^{M_1} \log (1 + \eta_j^* \sigma_r^2)
$$

Nevertheless, notice that for $\mathbf{A}_1^* = \mathbf{U}\boldsymbol{\eta}^*\mathbf{U}^\dagger$:

$$
\mathcal{L}(\mathbf{A}_1^*, \lambda^*) = \lambda^*\mathbf{R} + (1 - \lambda^*) \sum_{j=1}^{M_1} \log (1 + \eta_j^* s_j)
$$
$$
- \sum_{i=1}^{M_1} \log (1 + \eta_j^* \sigma_r^2).
$$

Then, it is demonstrated that $\mathbf{A}_1^* = \arg\max_{\mathbf{A}_1 \succeq 0} \mathcal{L}(\mathbf{A}_1, \lambda^*)$. Hence, the general sufficient condition holds, and it is optimum. Finally, $\boldsymbol{\Phi}_1^* = (\mathbf{A}_1^*)^{-1}$, which concludes the proof.

APPENDIX C

A. Proof of Theorem 2

In this Appendix, we solve the non-convex optimization $\mathbf{A}_n^* = \arg\max_{\mathbf{A}_n \succeq 0} \mathcal{L}(\mathbf{A}_1, \cdots, \mathbf{A}_N, \lambda)$. First, recall that $\log \det (\mathbf{I} + \mathrm{diag}(\mathbf{A}_1, \cdots, \mathbf{A}_N) \mathbf{R}_{\mathbf{Y}_{1:N}|\mathbf{Y}_0})$ is equal to $I(\mathbf{Y}_{1:N}; \hat{\mathbf{Y}}_{1:N}|\mathbf{Y}_0)$ (as shown in the proof of Proposition 1, changing $\boldsymbol{\Phi}_n = \mathbf{A}_n^{-1} \forall n$). Then:

$$
\log \det (\mathbf{I} + \mathrm{diag}(\mathbf{A}_1, \cdots, \mathbf{A}_N) \mathbf{R}_{\mathbf{Y}_{1:N}|\mathbf{Y}_0}) \tag{57}
$$
$$
= I(\mathbf{Y}_{1:N}; \hat{\mathbf{Y}}_{1:N}|\mathbf{Y}_0)
$$
$$
= I(\mathbf{Y}_{1:N}; \hat{\mathbf{Y}}_n^c|\mathbf{Y}_0) + I(\mathbf{Y}_{1:N}; \hat{\mathbf{Y}}_n|\mathbf{Y}_0, \hat{\mathbf{Y}}_n^c)
$$
$$
= I(\mathbf{Y}_n^c; \hat{\mathbf{Y}}_n^c|\mathbf{Y}_0) + I(\mathbf{Y}_n; \hat{\mathbf{Y}}_n|\mathbf{Y}_0, \hat{\mathbf{Y}}_n^c)
$$
$$
= \log \det (\mathbf{I} + \mathrm{diag}(\mathbf{A}_1, \cdots, \mathbf{A}_{n-1}, \mathbf{A}_{n+1}, \cdots, \mathbf{A}_N) \mathbf{R}_{\mathbf{Y}_n^c|\mathbf{Y}_0})
$$
$$
+ \log \det (\mathbf{I} + \mathbf{A}_n \mathbf{R}_{\mathbf{Y}_n|\mathbf{Y}_0, \hat{\mathbf{Y}}_n^c})
$$

where second equality follows from the chain rule for mutual information, and the third from the Markov chain in the proof Proposition 1. Finally, the fourth equality evaluates the mutual information as in (5), with $\boldsymbol{\Phi}_n = \mathbf{A}_n^{-1}$. The conditional covariances are computed in Appendix A. Later, using (43) and equivalently to (46):

$$
\log \det \left(\mathbf{I} + \frac{\mathbf{Q}}{\sigma_r^2} \mathbf{H}_{s,0}^\dagger \mathbf{H}_{s,0} + \mathbf{Q} \sum_{n=1}^N \mathbf{H}_{s,n}^\dagger (\mathbf{A}_n \sigma_r^2 + \mathbf{I})^{-1} \mathbf{A}_n \mathbf{H}_{s,n} \right)
$$
$$
= \log \det \left(\mathbf{I} + \frac{\mathbf{Q}}{\sigma_r^2} \mathbf{H}_{s,0}^\dagger \mathbf{H}_{s,0} + \mathbf{Q} \sum_{j \neq n} \mathbf{H}_{s,j}^\dagger (\mathbf{A}_j \sigma_r^2 + \mathbf{I})^{-1} \mathbf{A}_j \mathbf{H}_{s,j} \right)
$$
$$
+ \log \det (\mathbf{I} + \mathbf{A}_n \mathbf{R}_{\mathbf{Y}_n|\hat{\mathbf{Y}}_n^c, \mathbf{Y}_0}) - \log \det (\mathbf{I} + \mathbf{A}_n \sigma_r^2).
$$

Therefore, plugging last equality along with (58) into (12), we can expand the function under study as:

$$
\mathcal{L}(\mathbf{A}_1, \cdots, \mathbf{A}_N, \lambda) = \log \det \left(\mathbf{I} + \frac{\mathbf{Q}}{\sigma_r^2} \mathbf{H}_{s,0}^\dagger \mathbf{H}_{s,0} + \right.
$$
$$
\left. \mathbf{Q} \sum_{j \neq n}^N \mathbf{H}_{s,j}^\dagger (\mathbf{A}_j \sigma_r^2 + \mathbf{I})^{-1} \mathbf{A}_j \mathbf{H}_{s,j} \right)
$$
$$
+ \log \det (\mathbf{I} + \mathbf{A}_n \mathbf{R}_{\mathbf{Y}_n|\hat{\mathbf{Y}}_n^c, \mathbf{Y}_0}) - \log \det (\mathbf{I} + \mathbf{A}_n \sigma_r^2)
$$
$$
- \lambda (\log \det (\mathbf{I} + \mathrm{diag}(\mathbf{A}_1, \cdots, \mathbf{A}_{n-1}, \mathbf{A}_{n+1}, \cdots, \mathbf{A}_N) \mathbf{R}_{\mathbf{Y}_n^c|\mathbf{Y}_0})
$$
$$
+ \log \det (\mathbf{I} + \mathbf{A}_n \mathbf{R}_{\mathbf{Y}_n|\hat{\mathbf{Y}}_n^c, \mathbf{Y}_0}) - \mathbf{R})
$$

In order to obtain $\mathbf{A}_n^* = \arg\max_{\mathbf{A}_n \succeq 0} \mathcal{L}(\mathbf{A}_1, \cdots, \mathbf{A}_N, \lambda)$, we first notice that the following Lagrangian

$$
\bar{\mathcal{L}}(\mathbf{A}_n, \lambda) = (1 - \lambda) \log \det (\mathbf{I} + \mathbf{A}_n \mathbf{R}_{\mathbf{Y}_n|\mathbf{Y}_0, \hat{\mathbf{Y}}_n^c})
$$
$$
- \log \det (\mathbf{I} + \mathbf{A}_n \sigma_r^2) + \lambda\mathbf{R}
$$

satisfies $\arg\max_{\mathbf{A}_n \succeq 0} \bar{\mathcal{L}}(\mathbf{A}_n, \lambda) = \arg\max_{\mathbf{A}_n \succeq 0} \mathcal{L}(\mathbf{A}_{1:N}, \lambda)$, and it is identical to the Lagrangian in (53). Therefore, we can directly apply derivation (53)-(57) to solve it:

Consider first $\lambda \geq 1$. For it, and $\forall \mathbf{A}_n \succeq 0$:

$$
(1 - \lambda) \log \det (\mathbf{I} + \mathbf{A}_n \mathbf{R}_{\mathbf{Y}_n|\mathbf{Y}_0, \hat{\mathbf{Y}}_n^c})
$$
$$
- \log \det (\mathbf{I} + \mathbf{A}_n \sigma_r^2) \leq 0 \tag{58}
$$

Therefore, it is readily shown that:

$$
\mathbf{0} = \arg\max_{\mathbf{A}_n \succeq 0} \mathcal{L}(\mathbf{A}_1, \cdots, \mathbf{A}_N, \lambda) \quad \text{for } \lambda \geq 1. \tag{59}
$$

Let now $\lambda < 1$. Applying (53)-(57) we show that

$$
\mathbf{U}_n \boldsymbol{\eta} \mathbf{U}_n^\dagger = \arg\max_{\mathbf{A}_n \succeq 0} \mathcal{L}(\mathbf{A}_1, \cdots, \mathbf{A}_N, \lambda) \tag{60}
$$

with $\mathbf{R}_{\mathbf{Y}_n|\mathbf{Y}_0,\hat{\mathbf{Y}}_n^c} = \mathbf{U}_n \mathbf{S} \mathbf{U}_n^\dagger$, and

$$\eta_j = \left[\frac{1}{\lambda}\left(\frac{1}{\sigma_r^2} - \frac{1}{s_j} \right) - \frac{1}{\sigma_r^2} \right]^+, \quad j = 1, \cdots, M_n. \quad (61)$$

B. Solution of (14) with $\lambda \geq 1$

Applying equivalent arguments to those in (46), we can rewrite the Lagrangian in (14) as:

$$\mathcal{L}\left(\mathbf{A}_{1:N}, \lambda\right) = (1-\lambda)\log\det\left(\mathbf{I} + \operatorname{diag}\left(\mathbf{A}_{1:N}\right)\mathbf{R}_{\mathbf{Y}_{1:N}|\mathbf{Y}_0}\right)$$
$$- \log\det\left(\mathbf{I} + \operatorname{diag}\left(\mathbf{A}_{1:N}\right)\sigma_r^2\right) - \lambda \mathrm{R},$$

It is clear that, for $\lambda \geq 1$, the Lagrangian takes its optimal value at $\{\mathbf{A}_1^*, \cdots, \mathbf{A}_N^*\} = \mathbf{0}$.

REFERENCES

[1] J. G. Andrews, W. Choi, and R. W. Heath, "Overcoming interference in spatial multiplexing MIMO cellular networks," *IEEE Wireless Commun.*, vol. 14, no. 6, pp. 95–104, December 2007.

[2] G. J. Foschini, K. Karakayali, and R. A. Valenzuela, "Coordinating multiple antenna cellular networks to achieve enormous spectral efficiency," *IEE Proc. Commun.*, vol. 153, no. 4, pp. 548–555, Aug. 2006.

[3] K. Karakayali, G. J. Foschini, R. A. Valenzuela, and R. D. Yates, "On the maximum common rate achievable in a coordinated network," in *Proc. IEEE International Conference on Communications (ICC)*, Turkey, June 2006.

[4] O. Somekh, O. Simeone, Y. Bar-ness, A. Haimovich, U. Spagnolini, and S. Shamai, *An Information Theoretic View of Distributed Antenna Processing in Cellular Systems*. Auerbach Publication, CRC Press, 2007.

[5] M. Kamoun and L. Mazet, "Base-station selection in cooperative single frequency cellular network," in *Proc. IEEE Workshop on Signal Processing Advances in Wireless Communications*, Helsinki, Finland, June 2007.

[6] I. Telatar, "Capacity of multi-antenna Gaussian channel," *European Trans. Telecommun.*, vol. 10, no. 6, pp. 585–595, Nov. 1999.

[7] E. Aktas, J. Evans, and S. Hanly, "Distributed decoding in a cellular multiple-access channel," in *Proc. IEEE International Symposium on Infomation Theory*, Chicago, IL, June 2004, p. 484.

[8] P. Marsch and G. Fettweis, "A framework for optimizing the uplink performance of distributed antenna systems under a constrained backhaul," in *Proc. IEEE International Conference on Communications (ICC)*, Glasgow, UK, June 2007.

[9] Y. Oohama, "Gaussian multiterminal source coding," *IEEE Trans. Inform. Theory*, vol. 43, no. 6, pp. 1912–1923, Nov. 1997.

[10] A. Sanderovich, S. Shamai (Shitz), Y. Steinberg, and G. Kramer, "Communication via decentralized processing," *IEEE Trans. Inform. Theory*, vol. 54, no. 7, pp. 3008–3023, July 2008.

[11] A. Sanderovich, S. Shamai (Shitz), and Y. Steinberg, "Distributed MIMO receiver: achievable rates and upper bounds," submitted to *IEEE Trans. Inform. Theory*, 2008.

[12] 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.

[13] A. D. Wyner, "The rate-distortion function for source coding with side information at the decoder—II: general sources," *Inform. and Control*, pp. 60–80, 1978.

[14] M. Gastpar, "The Wyner-Ziv problem with multiple sources," *IEEE Trans. Inform. Theory*, vol. 50, no. 11, 2004.

[15] J. Chen and T. Berger, "Successive Wyner-Ziv coding scheme and its implications to the quadratic Gaussian CEO problem," *IEEE Trans. Inform. Theory*, vol. 4, no. 54, pp. 1586–1603, Apr. 2008.

[16] D. Slepian and J. K. Wolf, "Noiseless coding of correlated information sources," *IEEE Trans. Inform. Theory*, vol. 19, no. 4, pp. 471–481, July 1973.

[17] S. Boyd and L. Vandenberghe, *Convex Optimization*, 1st ed. Cambridge University Press, 2004.

[18] D.P. Bertsekas, *Nonlinear Programming*. Athena Scientific, Belmont, MA, 1995.

[19] K. B. Petersen and M. S. Pedersen, *The Matrix Cookbook*, 2007.

[20] T. Cover and J. Thomas, *Elements of Information Theory*. Wiley Series in Telecommunications, 1st ed., 1991.

[21] S. Simoens, O. Muoz, and J.Vidal, "Achievable rates of compress-and-forward cooperative relaying on Gaussian vector channels," in *Proc. International Conference on Communications (ICC)*, Glasgow, UK, June 2007.

[22] M. Gastpar, P. L. Dragotti, and M. Vetterli, "The distributed Karhunen-Loeve transform," *IEEE Trans. Inform. Theory*, vol. 52, no. 12, pp. 5177–5196, Dec. 2006.

[23] W. Yu and R. Lui, "Dual methods for nonconvex spectrum optimization of multicarrier systems," *IEEE Trans. Commun.*, vol. 54, no. 7, pp. 1310–1322, July 2006.

[24] D. P. Palomar and M. Chiang, "A tutorial on decomposition methods for network utility maximization," *IEEE J. Select. Areas Commun.*, vol. 24, no. 8, pp. 1439–1451, Aug. 2006.

[25] W. Yu, W. Rhee, S. Boyd, and J. M. Cioffi, "Iterative water-filling for Gaussian multiple-access channels," *IEEE Trans. Inform. Theory*, vol. 50, no. 1, pp. 145–152, Jan. 2004.

[26] W. Yu, "A dual decomposition approach to the sum power Gaussian vector multiple-access channel sum capacity problem," in *Proc. Conference on Information Sciences and Systems*, The Johns Hopkins University, Mar. 2003.

[27] D. P. Bertsekas and J. N. Tsitsiklis, *Parallel and Distributed Computation: Numerical Methods*. Athena Scientific, 1997.

[28] R. G. Cheng and S. Verdú, "Gaussian multiple-access channels with ISI: capacity region and multi-user water-filling," *IEEE Trans. Inform. Theory*, vol. 39, no. 3, pp. 773–785, May 1993.

[29] S. Ye and R. S. Blum, "Optimized signaling for MIMO interference systems with feedback," *IEEE Trans. Signal Processing*, vol. 51, no. 11, pp. 2839–2847, Nov. 2003.

[30] J. Liu, Y. T. Hou, and H. D. Sherali, "Conjugate gradient projection approach for multi-antenna Gaussian broadcast channels," in *Proc. IEEE International Symposium on Information Theory*, Nice, France, June 2007.

[31] J. Malick and H. S. Sendov, "Clarke generalized Jacobian of the projection onto the cone of positive semidefinite matrices," *Springer Set-Valued Analysis*, vol. 14, no. 3, pp. 273–293, Sept. 2006.

[32] K. Guan, "Some properties of a class of symmetric functions," *J. Math. Anal. and Appl.*, vol. 336, pp. 70–80, 2007.

[33] A. W. Marshall and I. Olkin, *Inequalities: Theory of Majorization and Its Applications*. Academic Press, 1979.

Aitor del Coso (Madrid, 1980) received M.Sc. degree in Telecommunications from Universidad Politécnica de Madrid (UPM) and Ph.D. degree in Signal Theory and Communications from Universitat Politècnica de Catalunya (UPC), in 2003 and 2008, respectively. He conducted his Ph.D. studies at the Access Technologies area of the Centre Tecnològic de Telecomunicacions de Catalunya (CTTC), Barcelona, from 2004 to 2008. While completing his Ph.D. degree, he also held visiting positions at the Politecnico de Milano, Italy, in 2005; New Jersey Institute of Tecnology (NJIT), USA, in 2006; and Motorola Research Labs, France, in 2007.

Currently, he is with the Multimedia Telecommunications Systems group of Thales Alenia Space Espana, Madrid. His research interests lie within the fields of wireless and satellite communications, communication theory, signal processing and information theory.

Sebastien Simoens graduated from Ecole Nationale Supérieure des Télécommunications (ENST) Paris in 1998. Until 2008 he worked with Motorola Labs Paris where he conducted research on signal processing for broadband wireless communications. He was involved in several European projects including IST-FIREWORKS and ICT-ROCKET. Since September 2005 he has been pursuing a PhD thesis with the Signal Theory and Communications Department of the Technical University of Catalonia (UPC), Barcelona. Since Nov. 2008, he has been with Thales Aerospace Division (Valence, France), working on signal processing for inertial navigation sensors.