

DISTRIBUTED COMPRESSION FOR THE UPLINK CHANNEL OF A COORDINATED CELLULAR NETWORK WITH A BACKHAUL CONSTRAINT

Aitor del Coso[†] and Sebastien Simoens^{*}

[†] Centre Tecnològic de Telecomunicacions de Catalunya (CTTC), Castelldefels, Spain

^{*} Motorola Labs, Paris, France

e-mail: aitor.delcoso@cttc.es, simoens@motorola.com

ABSTRACT

We propose distributed compression for the uplink channel of a backhaul-constrained coordinated cellular network. In the network, N multiple-antenna base stations compress their received signals using a Distributed Wyner-Ziv code, and then send the compressed vectors to a central BS, which centralizes decoding. For a single-user network, the compression codes at the BSs are optimized in this paper, considering the user's achievable rate as the performance metric.

1. INTRODUCTION

Inter-cell interference is one of the most limiting factors of current cellular networks. To overcome it, designers have resorted to orthogonal frequency channels, sectorized antennas and fractional frequency reuse. However, a more spectrally efficient solution has been recently proposed: coordinated cellular networks [1]. They consist of single-frequency networks with base stations (BSs) cooperating in order to receive from the mobile terminals. Inter-cell interference thus plays a constructive role, providing spectral efficiency gains, as well as macro and micro-diversity.

Preliminary results on the uplink capacity of coordinated networks consider all BSs connected via a lossless backhaul of unlimited capacity [2]. Accordingly, the capacity region of the network equals that of the MIMO multi-access channel, with a supra-receiver containing all the antennas of all cooperative BSs. Such an assumption is optimistic for current cellular networks. To deal with a more realistic backhaul constraint, two approaches have been already proposed. 1) *distributed decoding* [3], consisting on a demodulating scheme distributely carried out, based on local decisions and belief propagation. Decoding delay appears to be its main problem. 2) *Quantization* [4]. The 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; thus, introducing redundancy into the backhaul.

In this paper, we study a new approach for the network: *distributed compression*. The cooperative BSs, upon receiving their signals, distributely compress them using a multi-source lossy compression code [5]. Then, via the constrained backhaul, they transmit the compressed signals to the central unit (also a BS); this decompresses them using its own received signal as side information, and uses them to estimate the users' messages. The optimum compression of multiple, correlated, sources to be decompressed at single central unit is still unknown. To the best of authors knowledge, the scheme that better performs (in a rate-distortion sense) for this problem is Distributed Wyner-Ziv (D-WZ) compression [6]. Such a

compression is the direct extension of Berger-Tung coding to the decoding side information case [5]. In turn, Berger-Tung compression can be thought as the lossy counterpart of the Slepian-Wolf lossless coding [7]. D-WZ coding is thus the compression scheme proposed to be used in the network.

Our study focusses on the uplink achievable rate of a coordinated network with $N + 1$ multi-antenna BSs. The first BS, denoted BS₀, is the central unit and centralizes the user's decoding. The rest, BS₁, ..., BS_N, are cooperative BSs which compress their signals and transmit them to BS₀ using the common backhaul, of aggregate rate R . Each BS has N_i receive antennas, $i = 0, \dots, N$. In the network, we assume a unique multiple-antenna user s transmitting a pre-defined, Gaussian, space-time codeword over a *time-invariant, frequency-flat* channel. Receive channel state information is assumed at the decoding unit.

The reminder of the paper is organized as follows: Sec. 2 briefly introduces distributed compression. In Sec. 3 we state the problem and upper bound the user's achievable rate. Sec. 4 considers the case $N = 1$, while Sec. 5 generalizes results for $N > 1$. Sec. 6 depicts numerical results, and finally, Sec. 7 summarizes conclusions.

Notation. We write $\mathbf{Y}_{1:N} = \{\mathbf{Y}_1, \dots, \mathbf{Y}_N\}$, $\mathbf{Y}_{\mathcal{G}} = \{\mathbf{Y}_i | i \in \mathcal{G}\}$ and $\mathbf{Y}_n^c = \{\mathbf{Y}_i | i \neq n\}$. Block-diagonal matrices are defined as $\text{diag}(\mathbf{A}_1, \dots, \mathbf{A}_n)$, with \mathbf{A}_i square matrices. A sequence of vectors $\{\mathbf{Y}_i^t\}_{t=1}^n$ is compactly denoted by \mathbf{Y}_i^n . Finally, we define $\mathbf{R}_{\mathbf{X}|\mathbf{Y}} = \mathbf{E} \{(\mathbf{X} - \mathbf{E} \{\mathbf{X}|\mathbf{Y}\})(\mathbf{X} - \mathbf{E} \{\mathbf{X}|\mathbf{Y}\})^\dagger | \mathbf{Y}\}$.

2. MULTI-SOURCE DISTRIBUTED COMPRESSION

Let \mathbf{Y}_i^n , $i = 1, \dots, N$ be N zero-mean, temporally memoryless, Gaussian vectors compressed independently at BS₁, ..., BS_N, respectively. Assume that they are the observations at the BSs of the signal transmitted by user s , i.e., \mathbf{X}_s^n . The compressed vectors are sent to BS₀, which decompresses them using its side information \mathbf{Y}_0^n , and uses them to estimate the user's message. Notice that such an architecture imposes source-channel separation at the compression step, which is not shown to be optimal. However, it includes the coding scheme with best known performance: Distributed Wyner-Ziv coding [6]. It applies to the setup as follows.

Definition 1 (Multiple-source Compression Code) A compression code $(n, 2^{n\rho_1}, \dots, 2^{n\rho_N})$ with side information at the decoder \mathbf{Y}_0 is defined by $N + 1$ mappings, $f_n^i(\cdot)$, $i = 1, \dots, N$, and $g_n(\cdot)$, and $2N + 1$ spaces $\mathcal{Y}_i, \hat{\mathcal{Y}}_i$, $i = 1, \dots, N$ and \mathcal{Y}_0 , where

$$f_n^i : \mathcal{Y}_i^n \rightarrow \{1, \dots, 2^{n\rho_i}\}, \quad i = 1, \dots, N$$

$$g_n : \{1, \dots, 2^{n\rho_1}\} \times \dots \times \{1, \dots, 2^{n\rho_N}\} \times \mathcal{Y}_0^n \rightarrow \hat{\mathcal{Y}}_1^n \times \dots \times \hat{\mathcal{Y}}_N^n.$$

This work was carried out during A. del Coso's intern in Motorola Labs.

Proposition 1 (Distributed Wyner-Ziv Coding [6]) Let the random vectors $\hat{\mathbf{Y}}_i$, $i = 1, \dots, N$, have conditional probability $p(\hat{\mathbf{Y}}_i | \mathbf{Y}_i)$ and satisfy the Markov chain $(\mathbf{Y}_0, \mathbf{Y}_i^c, \hat{\mathbf{Y}}_i^c) \rightarrow \mathbf{Y}_i \rightarrow \hat{\mathbf{Y}}_i$. Let \mathbf{Y}_0 and \mathbf{Y}_i , $i = 1, \dots, N$ be jointly Gaussian. Considering a sequence of compression codes $(n, 2^{n\rho_1}, \dots, 2^{n\rho_N})$ with side information \mathbf{Y}_0 at the decoder, then

$$\frac{1}{n} I(\mathbf{X}_s^n; \mathbf{Y}_0^n, g_n(\mathbf{Y}_0^n, f_n^1(\mathbf{Y}_1^n), \dots, f_n^N(\mathbf{Y}_N^n))) = I(\mathbf{X}_s; \mathbf{Y}_0, \hat{\mathbf{Y}}_{1:N})$$

as $n \rightarrow \infty$ if:

- the compression rates ρ_1, \dots, ρ_N satisfy

$$I(\mathbf{Y}_G; \hat{\mathbf{Y}}_G | \mathbf{Y}_0, \hat{\mathbf{Y}}_G^c) \leq \sum_{i \in G} \rho_i \quad \forall G \subseteq \{1, \dots, N\}, \quad (1)$$

- each compression codebook \mathcal{C}_i , $i = 1, \dots, N$ consists of $2^{n\rho_i}$ random sequences $\hat{\mathbf{Y}}_i^n$ drawn i.i.d. from $\prod_{t=1}^n p(\hat{\mathbf{Y}}_i)$, where $p(\hat{\mathbf{Y}}_i) = \sum_{\mathbf{Y}_i} p(\mathbf{Y}_i) p(\hat{\mathbf{Y}}_i | \mathbf{Y}_i)$.
- for every $i = 1, \dots, N$, the encoding $f_i^n(\cdot)$ outputs the bin-index of codewords $\hat{\mathbf{Y}}_i^n$ that are jointly typical with the source sequence \mathbf{Y}_i^n . In turn, $g_n(\cdot)$ outputs the codewords $\hat{\mathbf{Y}}_i^n$, $i = 1, \dots, N$ that, belonging to the bins selected by the encoders, are all jointly typical with \mathbf{Y}_0^n .

Proof: The proposition is proven for discrete sources and discrete side information in [6, Theorem 2] and, also, conjectured therein for the Gaussian case. See [8, Proposition 1] for a complete proof.

3. PROBLEM STATEMENT

Let a single user s transmit a message $\omega \in \{1, \dots, 2^{nR_s}\}$ mapped onto a zero-mean, Gaussian codeword \mathbf{X}_s^n ; this is drawn i.i.d. from random vector $\mathbf{X}_s \sim \mathcal{CN}(\mathbf{0}, \mathbf{Q})$ and not subject to optimization. The transmitted signal, affected by *time-invariant, frequency-flat* fading, is received at the $N + 1$ BSs under additive noise:

$$\mathbf{Y}_i^n = \mathbf{H}_{s,i} \cdot \mathbf{X}_s^n + \mathbf{Z}_i^n, \quad i = 0, \dots, N, \quad (2)$$

where $\mathbf{H}_{s,i}$ is the MIMO channel matrix between user s and BS $_i$, and $\mathbf{Z}_i \sim \mathcal{CN}(\mathbf{0}, \sigma_r^2 \mathbf{I})$ is AWGN. Base stations BS $_1, \dots, \text{BS}_N$, upon receiving their signals, distributively compress them using a D-WZ compression code. Later, they transmit the compressed vectors to BS $_0$, which recovers them and uses them to decode. Considering so, the user's message can be reliably decoded *iff* [9, Theorem 1]:

$$\begin{aligned} R_s &\leq \frac{1}{n} I(\mathbf{X}_s^n; \mathbf{Y}_0^n, g_n(\mathbf{Y}_0^n, f_n^1(\mathbf{Y}_1^n), \dots, f_n^N(\mathbf{Y}_N^n))) \\ &= I(\mathbf{X}_s; \mathbf{Y}_0, \hat{\mathbf{Y}}_{1:N}). \end{aligned} \quad (3)$$

Second equality follows directly from Proposition 1. However, the equality only holds for compression rates satisfying (1). Notice that in the network there is only an overall backhaul constraint, $\sum_{i=1}^N \rho_i \leq R$. Therefore, the set of constraints in (1) are all embedded into the global constraint: $I(\mathbf{Y}_{1:N}; \hat{\mathbf{Y}}_{1:N} | \mathbf{Y}_0) \leq R$. Hence, the maximum transmission rate of user follows:

$$\begin{aligned} \mathcal{C} &= \max_{\prod_{i=1}^N p(\hat{\mathbf{Y}}_i | \mathbf{Y}_i)} I(\mathbf{X}_s; \mathbf{Y}_0, \hat{\mathbf{Y}}_{1:N}) \\ \text{s.t. } &I(\mathbf{Y}_{1:N}; \hat{\mathbf{Y}}_{1:N} | \mathbf{Y}_0) \leq R. \end{aligned} \quad (4)$$

Proposition 2 Let $\mathbf{X}_s \sim \mathcal{CN}(\mathbf{0}, \mathbf{Q})$. Optimization (4) is solved for Gaussian conditional distributions $p(\hat{\mathbf{Y}}_i | \mathbf{Y}_i)$, $i = 1, \dots, N$. Thus, the compressed vectors can be modelled as $\hat{\mathbf{Y}}_i = \mathbf{Y}_i + \mathbf{Z}_i^c$, where $\mathbf{Z}_i^c \sim \mathcal{CN}(\mathbf{0}, \Phi_i)$ is independent, Gaussian, "compression" noise at BS $_i$. That is,

$$\begin{aligned} \mathcal{C} &= \max_{\Phi_1, \dots, \Phi_N \geq 0} \log \det \left(\mathbf{I} + \frac{\mathbf{Q}}{\sigma_r^2} \mathbf{H}_{s,0}^\dagger \mathbf{H}_{s,0} \right. \\ &\quad \left. + \mathbf{Q} \sum_{n=1}^N \mathbf{H}_{s,n}^\dagger (\sigma_r^2 \mathbf{I} + \Phi_n)^{-1} \mathbf{H}_{s,n} \right) \\ \text{s.t. } &\log \det (\mathbf{I} + \text{diag}(\Phi_1^{-1}, \dots, \Phi_N^{-1}) \mathbf{R}_{\mathbf{Y}_{1:N} | \mathbf{Y}_0}) \leq R. \end{aligned} \quad (5)$$

where $\mathbf{R}_{\mathbf{Y}_{1:N} | \mathbf{Y}_0}$ is computed from [8, Appendix I] as:

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

Proof: See [8, Appendix II] for the proof.

Remark 1 The maximization above is not concave in standard form. Although the feasible set is convex, the objective function is not concave on Φ_1, \dots, Φ_N .

3.1. Useful Upper Bounds

Prior to solving the optimization above, we present two upper bounds that give insight into the problem at hand.

Upper Bound 1 The achievable rate \mathcal{C} in Prop. 2 satisfies

$$\begin{aligned} \mathcal{C} &\leq I(\mathbf{X}_s; \mathbf{Y}_0, \mathbf{Y}_{1:N}) \\ &= \log \det \left(\mathbf{I} + \frac{\mathbf{Q}}{\sigma_r^2} \sum_{n=0}^N \mathbf{H}_{s,n}^\dagger \mathbf{H}_{s,n} \right). \end{aligned} \quad (6)$$

Upper Bound 2 The achievable rate \mathcal{C} in Prop. 2 is bounded above by

$$\begin{aligned} \mathcal{C} &\leq I(\mathbf{X}_s; \mathbf{Y}_0) + R \\ &= \log \det \left(\mathbf{I} + \frac{1}{\sigma_r^2} \mathbf{H}_{s,0} \mathbf{Q} \mathbf{H}_{s,0}^\dagger \right) + R. \end{aligned} \quad (7)$$

Proof: See [8, Appendix III] for the proof.

Remark: Notice that, independently of the number of cooperative BSs, the achievable rate is bounded by the capacity with BS $_0$ plus the backhaul rate.

4. THE TWO-BASE STATIONS CASE

We consider first a simple network composed of BS $_0$ plus a single cooperative base station BS $_1$. The achievable rate of the uplink is given by Prop. 2, with $N = 1$. As mentioned, the optimization problem is not concave in standard form: the objective function, which has to be maximized, is convex on $\Phi_1 \geq 0$. To make it concave, we introduce the change of variables $\Phi_1 = \mathbf{A}_1^{-1}$, so that

$$\begin{aligned} \mathcal{C} &= \max_{\mathbf{A}_1 \geq 0} \log \det \left(\mathbf{I} + \frac{\mathbf{Q}}{\sigma_r^2} \mathbf{H}_{s,0}^\dagger \mathbf{H}_{s,0} \right. \\ &\quad \left. + \mathbf{Q} \mathbf{H}_{s,1}^\dagger (\mathbf{A}_1 \sigma_r^2 + \mathbf{I})^{-1} \mathbf{A}_1 \mathbf{H}_{s,1} \right) \\ \text{s.t. } &\log \det (\mathbf{I} + \mathbf{A}_1 \mathbf{R}_{\mathbf{Y}_{1:N} | \mathbf{Y}_0}) \leq R. \end{aligned} \quad (10)$$

$$\mathcal{L}(\mathbf{A}_{1:N}, \lambda) = \log \det \left(\mathbf{I} + \frac{\mathbf{Q}}{\sigma_r^2} \mathbf{H}_{s,0}^\dagger \mathbf{H}_{s,0} + \mathbf{Q} \sum_{n=1}^N \mathbf{H}_{s,n}^\dagger (\mathbf{A}_n \sigma_r^2 + \mathbf{I})^{-1} \mathbf{A}_n \mathbf{H}_{s,n} \right) - \lambda \cdot (\log \det (\mathbf{I} + \text{diag}(\mathbf{A}_{1:N}) \mathbf{R}_{\mathbf{Y}_{1:N}|\mathbf{Y}_0}) - \mathbf{R}). \quad (8)$$

$$\mathbf{R}_{\mathbf{Y}_n|\mathbf{Y}_0, \hat{\mathbf{Y}}_n^c} = \mathbf{H}_{s,n} \left(\mathbf{I} + \mathbf{Q} \left(\frac{1}{\sigma_r^2} \mathbf{H}_{s,0}^\dagger \mathbf{H}_{s,0} + \sum_{j \neq n} \mathbf{H}_{s,j}^\dagger (\mathbf{A}_j \sigma_r^2 \mathbf{I} + \mathbf{I})^{-1} \mathbf{A}_j \mathbf{H}_{s,j} \right) \right)^{-1} \mathbf{Q} \mathbf{H}_{s,n}^\dagger + \sigma_r^2 \mathbf{I} \quad (9)$$

Even though the objective has been transformed into concave, the constraint now does not define a convex feasible set on $\mathbf{A}_1 \succeq 0$. Therefore, the optimization is still not concave on standard form, and the Karush-Kuhn-Tucker (KKT) conditions become necessary but not sufficient for optimality. To solve the problem, we need to resort to the general sufficiency condition for optimality in non-convex problems [10, Proposition 3.3.4], as follows: first, we derive a matrix \mathbf{A}_1^* for which the KKT conditions hold. Later, we demonstrate that the selected matrix also satisfies the general sufficiency condition, thus becoming the optimal solution. The optimum compression noise is finally recovered as $\Phi_1^* = (\mathbf{A}_1^*)^{-1}$. This result is presented in the following Proposition.

Proposition 3 Let $\mathbf{X}_s \sim \mathcal{CN}(\mathbf{0}, \mathbf{Q})$ and the conditional covariance (see [8, Appendix I] for its computation):

$$\mathbf{R}_{\mathbf{Y}_1|\mathbf{Y}_0} = \mathbf{H}_{s,1} \left(\mathbf{I} + \frac{\mathbf{Q}}{\sigma_r^2} \mathbf{H}_{s,0}^\dagger \mathbf{H}_{s,0} \right)^{-1} \mathbf{Q} \mathbf{H}_{s,1}^\dagger + \sigma_r^2 \mathbf{I}, \quad (11)$$

with eigen-decomposition $\mathbf{R}_{\mathbf{Y}_1|\mathbf{Y}_0} = \mathbf{U} \text{diag}(s_1, \dots, s_{N_1}) \mathbf{U}^\dagger$. The optimum compression noise at BS₁ is

$$\Phi_1^* = \mathbf{U} (\text{diag}(\eta_1, \dots, \eta_{N_1}))^{-1} \mathbf{U}^\dagger, \quad (12)$$

where

$$\eta_i = \left[\frac{1}{\lambda} \left(\frac{1}{\sigma_r^2} - \frac{1}{s_i} \right) - \frac{1}{\sigma_r^2} \right]^+, \quad i = 1, \dots, N_1, \quad (13)$$

and λ is such that $\sum_{i=1}^{N_1} \log(1 + \eta_i s_i) = \mathbf{R}$.

Proof: See [8, Appendix IV] for the complete proof.

5. THE MULTIPLE-BASE STATIONS CASE

Consider now BS₀ assisted by a set of $N > 1$ cooperative BSs. The achievable rate follows Prop. 2. As mentioned previously, the maximization is not concave over the set of "compression" noises Φ_i , $i = 1, \dots, N$. To make it concave, we introduce again the change of variables $\Phi_n = \mathbf{A}_n^{-1}$, $n = 1, \dots, N$, so that:

$$\mathcal{C} = \max_{\mathbf{A}_1, \dots, \mathbf{A}_N \succeq 0} \log \det \left(\mathbf{I} + \frac{\mathbf{Q}}{\sigma_r^2} \mathbf{H}_{s,0}^\dagger \mathbf{H}_{s,0} + \mathbf{Q} \sum_{n=1}^N \mathbf{H}_{s,n}^\dagger (\mathbf{A}_n \sigma_r^2 + \mathbf{I})^{-1} \mathbf{A}_n \mathbf{H}_{s,n} \right) \quad (14)$$

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

This is not a concave maximization either since, again, the feasible set is not convex. Our strategy to solve such an optimization is the following: first, we show that the duality gap for the problem is equal to zero. To demonstrate it, we resort to the time-sharing property in [11, Theorem 1]. Later, we propose an iterative algorithm that solves the dual problem, thus solving the primal too. An interesting property of the dual problem is that the coupling constraint in (14) is decoupled [10, Chapter 5].

5.1. The dual problem

Let the Lagrangian of the problem be defined for $\mathbf{A}_n \succeq 0$, $n = 1, \dots, N$ and $\lambda \geq 0$ as in (8) above. Then, the dual function $g(\lambda)$ is defined as [12, Section 5.1]:

$$g(\lambda) = \max_{\mathbf{A}_1, \dots, \mathbf{A}_N \succeq 0} \mathcal{L}(\mathbf{A}_1, \dots, \mathbf{A}_N, \lambda), \quad (15)$$

and the solution of the dual problem is then obtained from

$$\mathcal{C}' = \min_{\lambda \geq 0} g(\lambda). \quad (16)$$

Lemma 1 The duality gap for optimization (14) is zero, i.e., the primal problem (14) and the dual problem (16) have the same solution.

Proof: The duality gap for problems of the form of (14), and satisfying the time-sharing property, is zero [11, Theorem 1]. Time-sharing property is defined as follows: let $\mathcal{C}_x, \mathcal{C}_y, \mathcal{C}_z$ be the solution of (14) for backhaul rates R_x, R_y, R_z , respectively. Consider $R_z = \nu R_x + (1 - \nu) R_y$ 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 (14) is concave with respect to the backhaul rate R . It is well known that time-sharing of compressions cannot decrease the resulting distortion [13, Lemma 13.4.1], neither improve the mutual information obtained from the reconstructed vectors. Hence, the property holds for (14), and the duality gap is zero.

We then solve the dual problem in order to obtain the solution of the primal. We do it by means of an iterative algorithm. First, let consider the maximization (15). As expected, it can not be solved in closed form. However, as the feasible set (i.e., $\mathbf{A}_1, \dots, \mathbf{A}_N \succeq 0$) is the cartesian product of convex sets, then a block coordinate ascent algorithm [10, Section 2.7] (a.k.a. Gauss-Seidel Algorithm [14, Section II-C]) can be used in order to search for the maximum. The algorithm consists of iteratively optimize $\mathcal{L}(\mathbf{A}_1, \dots, \mathbf{A}_N, \lambda)$ with respect to one variable \mathbf{A}_n while keeping the others fixed. We define it for our problem as:

$$\mathbf{A}_n^{t+1} = \arg \max_{\mathbf{A}_n \succeq 0} \mathcal{L}(\mathbf{A}_1^{t+1}, \dots, \mathbf{A}_{n-1}^{t+1}, \mathbf{A}_n, \mathbf{A}_{n+1}^t, \dots, \mathbf{A}_N^t, \lambda), \quad (17)$$

where t is the iteration index. As shown in Proposition 4, the maximization (17) is uniquely attained.

Proposition 4 Let the optimization $\mathbf{A}_n^* = \arg \max_{\mathbf{A}_n \succeq 0} \mathcal{L}(\mathbf{A}_{1:N}, \lambda)$. Consider the conditional covariance matrix in (9), with eigen-decomposition $\mathbf{R}_{\mathbf{Y}_n|\mathbf{Y}_0, \hat{\mathbf{Y}}_n^c} = \mathbf{U}_n \mathbf{S} \mathbf{U}_n^\dagger$ (See [8, Appendix I] for its computation). The optimization is uniquely attained at $\mathbf{A}_n^* = \mathbf{U}_n \boldsymbol{\eta} \mathbf{U}_n^\dagger$, where

$$\eta_i = \left[\frac{1}{\lambda} \left(\frac{1}{\sigma_r^2} - \frac{1}{s_i} \right) - \frac{1}{\sigma_r^2} \right]^+, \quad i = 1, \dots, N_n. \quad (18)$$

Proof: See [8, Appendix V] for the proof.

Function $\mathcal{L}(\mathbf{A}_1, \dots, \mathbf{A}_N, \lambda)$ is continuously differentiable, and the maximization (17) is uniquely attained. Hence, the limit

point of the sequence $\{\mathbf{A}_1^t, \dots, \mathbf{A}_N^t\}$ is proven to converge to a local maximum of the Lagrangian [10, Proposition 2.7.1]. To demonstrate convergence to the global maximum, and therefore to $g(\lambda)$, it is necessary to prove that the mapping $T(\mathbf{A}_1, \dots, \mathbf{A}_N) = [\mathbf{A}_1 + \gamma \nabla_{\mathbf{A}_1} \mathcal{L}, \dots, \mathbf{A}_N + \gamma \nabla_{\mathbf{A}_N} \mathcal{L}]$ is a block contraction¹ for some γ [15, Proposition 3.10]. We were not able to mathematically demonstrate the contraction property on the Lagrangian. However, simulation results show convergence of our algorithm to the global maximum always.

Once obtained $g(\lambda)$ through the Gauss-Seidel Algorithm, it remains to minimize it on $\lambda \geq 0$ (assume hereafter that the algorithm has converged to the global maximum, as simulations suggest). $g(\lambda)$ is a convex function since it is defined as the pointwise maximum of a family of affine functions [12]. To minimize it, we may use a sub-gradient approach, which consists of following search direction $-h$ such that

$$\frac{g(\lambda') - g(\lambda)}{\lambda' - \lambda} \geq h(\lambda) \quad \forall \lambda'. \quad (19)$$

The subgradient search is proven to converge to the global minimum for diminishing step-size rules [14, Section II-B]. Taking into account the definition of $\mathcal{L}(\mathbf{A}_1, \dots, \mathbf{A}_N, \lambda)$ and $g(\lambda)$, the following $h(\lambda)$ is proven to satisfy the subgradient condition

$$h(\lambda) = R - \log \det(\mathbf{I} + \text{diag}(\mathbf{A}_1, \dots, \mathbf{A}_N) \mathbf{R}_{\mathbf{Y}_{1:N}|\mathbf{Y}_0}). \quad (20)$$

where $\{\mathbf{A}_1, \dots, \mathbf{A}_N\}$ are the limiting points of (17). That subgradient is then used to search the optimum λ as:

$$\begin{aligned} &\text{increase } \lambda \text{ if } h \leq 0 \\ &\text{decrease } \lambda \text{ if } h \geq 0. \end{aligned} \quad (21)$$

Consider now $\lambda^0 = 1$ as the initial value of the Lagrange multiplier. For such a multiplier, the optimum solution of (15) is $\{\mathbf{A}_1^*, \dots, \mathbf{A}_N^*\} = \mathbf{0}$ (See [8, Appendix V-B]). Accordingly, the subgradient (20) is equal to $h = R$. Therefore, following the sub-gradient search (21), the optimum value of the multiplier has to be strictly lower than one. Algorithm 1 takes all this into account in order to solve the dual problem, and thus to find the solution of (14). As mentioned, we can only claim convergence of the algorithm to a local maximum.

Algorithm 1 Multiple-BSSs Dual Problem

- 1: Initialize $\lambda_{\min} = 0$ and $\lambda_{\max} = 1$
 - 2: **repeat**
 - 3: $\lambda = \frac{\lambda_{\max} + \lambda_{\min}}{2}$
 - 4: Obtain $\{\mathbf{A}_1^*, \dots, \mathbf{A}_N^*\} = \arg \max \mathcal{L}(\mathbf{A}_1, \dots, \mathbf{A}_N, \lambda)$ from Algorithm 2
 - 5: Evaluate $h = R - \log \det(\mathbf{I} + \text{diag}(\mathbf{A}_1^*, \dots, \mathbf{A}_N^*) \mathbf{R}_{\mathbf{Y}_{1:N}|\mathbf{Y}_0})$, where $\mathbf{R}_{\mathbf{Y}_{1:N}|\mathbf{Y}_0}$ follows Prop. 1.
 - 6: if $h \leq 0$, then $\lambda_{\min} = \lambda$, else $\lambda_{\max} = \lambda$
 - 7: **until** $\lambda_{\max} - \lambda_{\min} \leq \epsilon$
 - 8: $\{\mathbf{R}_1^*, \dots, \mathbf{R}_N^*\} = \{(\mathbf{A}_1^*)^{-1}, \dots, (\mathbf{A}_N^*)^{-1}\}$
-

6. NUMERICAL RESULTS

We evaluate here the uplink capacity of a cellular network with 7 cells, composed of BS₀ and its first tier of six cells. The radius

¹See [15, Section 3.1.2] for the definition of block-contraction.

Algorithm 2 Gauss-Seidel to Obtain $g(\lambda)$

- 1: Initialize $\mathbf{A}_n^0 = \mathbf{0}$, $n = 1, \dots, N$ and $t = 0$
 - 2: **repeat**
 - 3: **for** $n = 1$ to N **do**
 - 4: Compute $\mathbf{R}_{\mathbf{Y}_n|\hat{\mathbf{Y}}_n^c, \mathbf{Y}_0}(\mathbf{A}_1^{t+1}, \dots, \mathbf{A}_{n-1}^{t+1}, \mathbf{A}_{n+1}^t, \dots, \mathbf{A}_N^t)$ as in (9).
 - 5: Take its eigen-decomposition $\mathbf{U}_n \mathbf{S} \mathbf{U}_n^{\dagger}$.
 - 6: Compute $\boldsymbol{\eta}$ as in (18).
 - 7: Update $\mathbf{A}_n^{t+1} = \mathbf{U}_n \boldsymbol{\eta} \mathbf{U}_n^{\dagger}$.
 - 8: **end for**
 - 9: $t = t + 1$
 - 10: **until** The sequence converges $\{\mathbf{A}_1^t, \dots, \mathbf{A}_N^t\} \rightarrow \{\mathbf{A}_1^*, \dots, \mathbf{A}_N^*\}$
 - 11: Return $\{\mathbf{A}_1^*, \dots, \mathbf{A}_N^*\}$
-

of each cell is 700 m. Within the network, wireless channels are simulated with path loss, shadowing and *i.i.d* Rayleigh fading among antennas. The transmission bandwidth is set to 1 MHz and the carrier frequency is 2.5 GHz. We consider a single user equipped with 2 TX antennas, while all BSs have 3 RX antennas. Finally, the transmitted power is 23 dBm, with isotropic transmission, *i.e.*, $\mathbf{Q} = \frac{P}{2} \mathbf{I}$.

Fig. 1 depicts the cdf of the user's rate considering Line-of-Sight (LOS) propagation, with path-loss exponent $\alpha = 2.6$, and shadowing standard deviation $\sigma = 4$ dB. The user is located at the edge of the central cell, and the cdf is plotted for different values of the backhaul rate R. Results show gains up to 4 and 6 Mbit/s @ 5% outage, with $R = 4.5$ Mbit/s and $R = 15$ Mbit/s, respectively. As expected, the distance with the upper bound 1 vanishes to zero for increasing backhaul rates. Fig. 2 depicts the cdf of the user's rate for a coordinate network composed of BS₀ plus N cooperative BSs, with N ranging from 1 to 6. The backhaul rate is set to 7 Mbit/s. Again, LoS is considered. It is shown that, @ 5% outage and 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 a relative gain of 2 Mbit/s is obtained with respect to the 1 BS case.

Fig. 3 plots results for non-Line-of-Sight (N-LOS) propagation, assuming a path-loss exponent $\alpha = 4.05$ and and shadowing standard deviation $\sigma = 10$ dB. Results for different number of cooperative BSs are shown, considering a backhaul rate of 7 Mbit/s. Interestingly, @ 50% outage, the capacity is doubled from 1 cooperative BS to 6 cooperative BS. Hence, it is shown the definite advantage of macro-diversity for N-LOS propagation, which is the most common condition within urban environments.

7. CONCLUSIONS

We studied distributed compression for the uplink of a coordinated cellular network composed of $N + 1$ multi-antenna BSs. Considering a constrained backhaul of limited capacity R, base stations BS₁, ..., BS_N distributively compress their received signal using a Distributed Wyner-Ziv code. The compressed vectors are sent to BS₀, which centralizes user's decoding. Considering a single user within the network, the D-WZ scheme has been optimized. Indeed, for the $N = 1$ case, the optimum compression noise has been derived in closed form. For $N > 1$ an iterative algorithm, based on dual decomposition and Gauss-Seidel optimization, has been devised. The extension to multiple-user is carried out in a further contribution [8].

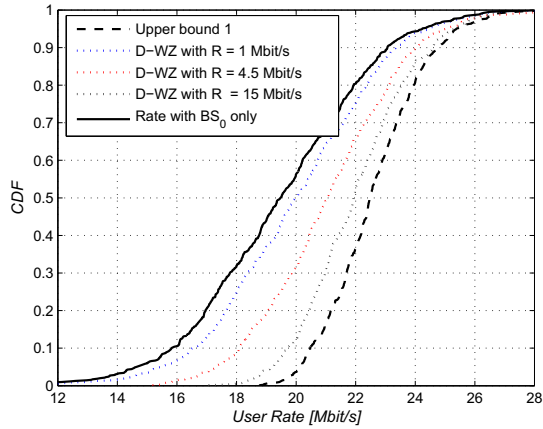


Fig. 1. CDF of the user rate for different backhaul rates R . LOS propagation. $N = 6$ cooperative BSs.

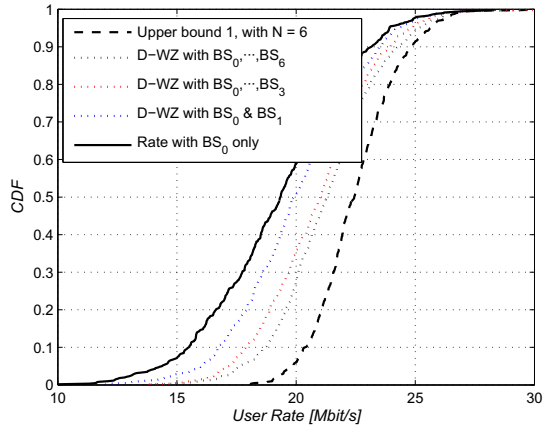


Fig. 2. CDF of the user rate for different number of cooperative BSs. LOS propagation. $R = 7$ Mbps

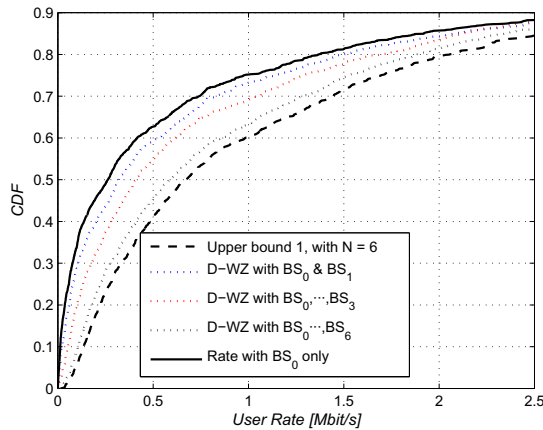


Fig. 3. CDF of the user rate for different number of cooperative BSs. N-LOS propagation. $R = 7$ Mbps

8. REFERENCES

- [1] G.J. Foschini, K. Karakayali, and R.A. Valenzuela, "Coordinating multiple antenna cellular networks to achieve enormous spectral efficiency," *IEE Proceedings Communications*, vol. 153, no. 4, pp. 548–555, Aug. 2006.
- [2] 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.
- [3] E. Aktas, J. Evans, and S. Hanly, "Distributed decoding in a cellular multiple-access channel," in *Proc. IEEE International Symposium on Information Theory*, Chicago, IL, Jun. 2004, p. 484.
- [4] 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, Jun. 2007.
- [5] S.Y. Tung, *Multiterminal source coding*, PhD Dissertation, Cornell University, 1978.
- [6] M. Gastpar, "The Wyner-Ziv problem with multiple sources," *IEEE Trans. on Information Theory*, vol. 50, no. 11, Nov. 2004.
- [7] D. Slepian and J.K. Wolf, "Noiseless coding of correlated information sources," *IEEE Trans. on Information Theory*, vol. 19, no. 4, pp. 471–481, Jul. 1973.
- [8] A. del Coso and S. Simoens, "Distributed compression for the uplink of a backhaul-constrained coordinated cellular network," *submitted to IEEE Trans. on Signal Processing*, 2008. arXiv:0802.0776.
- [9] A. Sanderovich, S. Shamai (Shitz), Y. Steinberg, and G. Kramer, "Communication via decentralized processing," in *Proc. IEEE International Symposium on Information Theory (ISIT)*, Adelaide, Australia, Jun. 2005.
- [10] D.P. Bertsekas, *Nonlinear Programming*, Athena Scientific, Belmont, MA, 1995.
- [11] W. Yu and R. Lui, "Dual methods for nonconvex spectrum optimization of multicarrier systems," *IEEE Trans. on Communications*, vol. 54, no. 7, pp. 1310–1322, Jul. 2006.
- [12] S. Boyd and L. Vandenberghe, *Convex Optimization*, 1st Edition, Cambridge University Press, 2004.
- [13] T. Cover and J. Thomas, *Elements of Information Theory*, Wiley Series in Telecommunications, 1991.
- [14] D.P. Palomar and M. Chiang, "A tutorial on decomposition methods for network utility maximization," *IEEE Journal on Selected Areas in Communications*, vol. 24, no. 8, pp. 1439–1451, Aug. 2006.
- [15] D.P. Bertsekas and J.N. Tsitsiklis, *Parallel and Distributed Computation: Numerical Methods*, Athena Scientific, Belmont, MA, 1997.