Compress-and-Forward Cooperative MIMO Relaying with Full Channel State Information

S. Simoens[*†], O. Muñoz, J. Vidal[†], [†]*Member, IEEE* and A. Del Coso, *Student Member, IEEE*

Abstract—This paper addresses cooperative Time Division Duplex (TDD) relaying in the multiple-antenna case with full Channel State Information (CSI), i.e. assuming perfect knowledge of all channels. The main focus of the paper is on the Compress-and-Forward (CF) strategy, for which an achievable rate on the Gaussian MIMO relay channel can be derived by applying distributed vector compression techniques. The processing at the CF relay consists in a Conditional Karhunen-Loève Transform (CKLT) followed by a separate Wyner-Ziv (WZ) coding of each output stream at a different rate. The paper provides a simple analytical expression for the optimum WZ coding rates, and also proposes an iterative procedure to perform this optimization jointly with that of the transmit covariance matrices at the source and relay. The Multiple Access Channel (MAC) formed by the source and relay transmitting simultaneously to the destination is considered, and it is shown that an optimal decoding order exists at least in the single-antenna case. We discuss the extension to MIMO-OFDM, as well as practical source coding implementation. The CF achievable rates are benchmarked with other upper and lower bounds on capacity. Simulation results show that CF can outperform Decode-and-Forward (DF) and approach capacity for realistic SNR values, which validates the performance of the proposed optimization procedure.

Index Terms—compress, forward, relay, CSI, MIMO

EDICS— MSP-CODR (MIMO precoder/decoder design)

Manuscript received November 23, 2008. This work was supported in part by the ICT ROCKET Project of the 7Th European Research Framework, by the Spanish/Catalan Science and Technology Commissions and FEDER funds from the EC: TEC2006-06481/TCM, TEC2004-04526, and 2005SGR-00639

Sébastien Simoens, Josep Vidal and Olga Muñoz are with the Signal Theory and Communications Department, Technical University of Catalonia, 08034 Barcelona, Spain (e-mail: sebastien.simoens@gmail.com, olga.munoz@upc.edu; josep.vidal@upc.edu).

Aitor Del Coso is with Thales Alenia Space, Madrid, Spain. e-mail: aitor.delcososanchez@thalesaleniaspace.com

I. INTRODUCTION

Cooperative Relaying (see e.g. [1] for an overview) is regarded as a promising technique for next generation cellular wireless systems and is being considered by several standardization groups (e.g. IEEE802.16 Task Groups j and m [2]). In cooperative relaying, a node on a multi-hop path from a source to a destination can receive and process the signals transmitted by several predecessors. The latter can thus transmit cooperatively to maximize the end-to-end information rate. In the above-mentioned standards, it is also envisioned that both infrastructure devices (e.g. base stations, fixed relay stations) and mobile devices (e.g. handsets) will be equipped with multiple antennas, and that MIMO Channel State Information will be exploited not only at the receiver-side (CSIR) but also at the transmitter-side (CSIT). On the point-to-point Gaussian MIMO channel, the capacity gains brought by full CSI (i.e. both CSIT and CSIR) vs. only CSIR are well-known and analyzed in e.g. §8.2.3 of [3]. In particular, the source transmit covariance that attains capacity with full CSI is derived by Telatar in [4]. On the Gaussian MIMO relay channel, it is also expected that full CSI (here defined as the perfect knowledge of all channels at each node) can increase capacity. However, although information theory provides various coding strategies for the relay channel, it remains difficult to compute their corresponding achievable rates with full MIMO CSI and to quantify the gap with respect to capacity. Indeed, the optimization of the associated MIMO signal processing is a difficult problem and this is the domain in which our paper aims at contributing.

Before introducing our contribution in more details it is useful to review coding strategies for the relay channel. In the landmark paper [5] the capacity of the full-duplex single-relay channel is studied. Although not naming them, the authors in [5] introduce DF in Theorem 1 and CF in Theorem 6. Moreover, an upper-bound on capacity (a.k.a. the cut-set bound) is provided for the general relay channel. In DF, the Relay (R) decodes at least a part of the message sent by (S) and forwards it to (D), whereas in CF the relay observation is compressed and forwarded to D as a new codeword. In [6], DF and CF coding strategies are proposed for the Gaussian TDD relay channel. This interest in half-duplex relaying is motivated by its simpler practical implementation, as the relay transmit and receive chain cannot interfere with each other. The TDD protocol in [6] consists of a relay-receive slot of duration denoted by $t \in [0;1]$ followed by a relay-transmit slot of duration $(1-t)$. According to the conventions of [7], the TDD protocol considered in [6] is "fixed-dynamic" i.e. t is not random but can be optimized for a given channel

realization. In the DF coding strategy of [6], the message transmitted by the source is split into two parts: the first part is transmitted by S and decoded by R during the first slot and the second part is transmitted directly to D during the second slot while the relay is forwarding the first part. This strategy is thus named "partial DF" in [8]. When the capacity of the source to relay link becomes infinitely high, DF becomes capacity-achieving as shown in e.g. [8].

Another well-known strategy is Amplify-and-Forward (AF), in which the relay amplifies and retransmits the samples it receives. The AF strategy can be viewed as a sub-case of Linear Relaying (LR) (see e.g. [9]) in which the relay applies a linear processing to the received signal. One advantage of LR compared to DF is the processing simplicity. However, a drawback in the single TDD relay case is that time-sharing parameter is constrained to $t = 1/2$, which severely limits the achievable rate. Indeed, when the R-D link has infinitely high capacity, the relay channel capacity tends to that of a point-to-point MIMO channel between S and (R,D) but AF cannot achieve this capacity because the relay only receives half of the time. Although this limitation of LR can be alleviated by employing multiple relays [10] or by exploiting spatial reuse [11], the scenario with infinite R-D link capacity naturally leads to introducing the CF strategy: during the second slot, instead of sending a replica of the received signal, one can think of quantizing it and sending the resulting finite sequence of bits. The destination can then reconstruct the relay observation, and the duration of the second slot can be made arbitrarily short whatever the quantization accuracy (or equivalently whatever the distortion), provided the R-D link capacity is infinite. In this case, CF becomes capacity-achieving. The CF strategy in [5] and [6] goes beyond a plain quantization of the relay observation and exploits the fact that the signals received at R and D are correlated, which improves the rate-distortion trade-off. The CF relay employs source coding with side information at the destination (a.k.a. Wyner-Ziv (WZ) coding [12]). In [6], achievable rates on the scalar Gaussian TDD relay channel are derived for a partial CF strategy, in which a first message is transmitted during the first slot and is WZ-compressed by the relay while a second message is transmitted by the source directly to the destination during the second slot. Further improvements were recently introduced in [13] by e.g. tolerating errors in the WZ decoding[1].

This paper presents an extension of the partial CF achievable rates derived in [6] to the multiple antenna case. Our contribution is the following:

[1] In [13], relays are called "agents" and are connected to a central processing unit by a fixed capacity link.

- Distributed Gaussian vector compression [16] is applied to the specific case of CF relaying (§II.B) and the effect of this compression on the achievable rate of CF is analyzed (§II.C).
- The achievable rate is maximized with full CSI. A closed-form expression of the optimum WZ coding rates is derived (§III.A) which differs from the rate-distortion trade-off in [16].
- It is shown in §III.B that during the second slot of the TDD protocol an optimum decoding order exists for the messages transmitted by S and R, and this can be used to simplify the optimization of the source and relay covariance matrices. Finally, an iterative procedure is proposed (§III.C) which jointly optimizes the compression, the transmit covariances and the time resource allocation.
- Practical implementation aspects are discussed (§IV.A) and simulations in a scenario of practical interest are performed to compare the CF achievable rate with other known capacity bounds (§IV.B.)

Note that the first two items in the above bullet-list were partially published in [17] but are more detailed here. Novel aspects include the optimization of transmit covariance matrices and the comparison with recently published capacity bounds. This comparison relies on a computation of the cut-set bound and partial DF achievable rates for the MIMO relay channel with full CSI which is performed in [18] and [19].

II. Signal Model And Achievable Rates

In the following section II.A we introduce the notations and describe the coding strategy, before explaining in details the vector source coding at the relay in II.B and computing achievable rates in II.C.

A. Signal model and coding strategy

The coding strategy assumed in this paper is an extension to the multiple antenna case of the partial CF described in [6]. The Source, Relay and Destination are equipped with respectively N_S, N_R and N_D antennas. The channels from S to D, S to R and R to D are all assumed static and are denoted respectively $\mathbf{H}_0$, $\mathbf{H}_1$ and $\mathbf{H}_2$. During the first slot of duration t, S transmits a first message ω_0 at rate R_0 using the codeword[2] $\mathbf{x}_S^{(1)}(\omega_0)$. We assume that $\mathbf{x}_S^{(1)}$ is a

[2] A complete information-theoretic treatment would not define a codeword as length-N_S random vector but as a length-n sequence of length-N_S vectors drawn i.i.d. from a certain distribution. In this paper, we are interested in finding the distribution that maximizes the achievable rate, hence the simplified notation used. We rely on [6], [16] and [13] for the rate achievability proofs, established for n going to infinity.

proper [20] complex Gaussian vector $\mathbf{x}_S^{(1)} \sim \mathbb{CN}\left(\mathbf{0}_{N_S}, \mathbf{R}_S^{(1)}\right)$, although this may not be the optimum distribution, as pointed out in section VI.B of [13]. The received signals at R and D read:

$$\mathbf{y}_R^{(1)} = \mathbf{H}_1 \mathbf{x}_S^{(1)} + \mathbf{n}_R^{(1)} \tag{1}$$

$$\mathbf{y}_D^{(1)} = \mathbf{H}_0 \mathbf{x}_S^{(1)} + \mathbf{n}_D^{(1)} \tag{2}$$

where the superscript $^{(i)}$ indicates that the signals are transmitted or received during the ith slot. The noise at R and D is also assumed (proper) complex white Gaussian of respective covariance $\mathbf{\Sigma}_R = \sigma^2 \mathbf{I}_{N_R}$ and $\mathbf{\Sigma}_D = \sigma^2 \mathbf{I}_{N_D}$. The compression at R, which will be detailed in §II.B, consists in mapping $\mathbf{y}_R^{(1)}$ onto the index ω_1 assuming side information $\mathbf{y}_D^{(1)}$ at the destination. The relay forwards ω_1 to D during the second slot of duration $1-t$ by transmitting $\mathbf{x}_R^{(2)}(\omega_1)$ at rate R_1, while S transmits a new message ω_2 at rate R_2 by means of the codeword $\mathbf{x}_S^{(2)}(\omega_2)$. The transmit power is constrained to be lower than P_S and P_R at S and R respectively during each slot, i.e. $\mathrm{tr}\left(\mathbf{R}_S^{(i)}\right) \le P_S$ i=1,2 and $\mathrm{tr}\left(\mathbf{R}_R^{(2)}\right) \le P_R$. The destination receives

$$\mathbf{y}_D^{(2)} = \mathbf{H}_0 \mathbf{x}_S^{(2)} + \mathbf{H}_2 \mathbf{x}_R^{(2)} + \mathbf{n}_D^{(2)} \tag{3}$$

from which $\hat{\omega}_1$ and $\hat{\omega}_2$ are successively decoded (the decoding order will be discussed in part III.C). The decompression at D can be modeled as a mapping of $\left(\mathbf{y}_D^{(1)}, \hat{\omega}_1\right)$ onto $\hat{\mathbf{y}}_R^{(1)}$, the reconstructed relay observation. After decompression, D decodes $\hat{\omega}_0$ from $\left(\mathbf{y}_D^{(1)}, \hat{\mathbf{y}}_R^{(1)}\right)$. Finally the achievable rate between S and D is :

$$R_{CF} = tR_0 + (1-t)R_2 \tag{4}$$

The rate R_{CF} needs to be maximized w.r.t. the time-sharing parameter, the distribution of the codewords, the compression mapping and the decoding order at the destination. Because the received signals at R and D are Gaussian, recently published results can be applied to design the compression mapping at the relay, as explained in the next section.

B. Vector source coding at the relay

In [6], Høst-Madsen et al. show that the achievable rate of partial CF depends on the variance of a "compression noise" which differs from the quadratic distortion in general. This compression noise was first introduced by Wyner, who derived in [12] the rate-distortion function for source coding with Gaussian source and Gaussian side

information. In [15] and [16], Gastpar et al. investigate distributed source coding and introduce the Distributed KLT and the Conditional KLT. They show that the rate-distortion coding of a Gaussian vector source with side information at the decoder is achieved by first applying a CKLT and then separately performing WZ coding of each CKLT output at a different rate. The results presented is this section II.B can be viewed as a special case of [16]. A compression noise vector is defined, and its relationship with distortion is clarified. The main difference between our work and [16] will arise in §II.C where it will be shown that the code which maximizes the CF achievable rate is not the same code that minimizes the quadratic distortion, which is the design criterion in [16].

Let first define the Conditional Karhunen-Loeve Transform (CKLT) in the CF relaying case. Given the knowledge of $\mathbf{y}_D^{(1)}$, the vector $\mathbf{y}_R^{(1)}$ is Gaussian distributed (see ex. 4.12.1 in [22]) of mean:

$$E\left[\mathbf{y}_R^{(1)}\middle|\mathbf{y}_D^{(1)}\right]=\mathbf{R}_{R,D}^{(1)}\left(\mathbf{R}_D^{(1)}\right)^{-1}\mathbf{y}_D^{(1)} \tag{5}$$

and covariance matrix denoted by $\mathbf{R}_{R|D}^{(1)}$ and equal to:

$$\mathbf{R}_{R|D}^{(1)}=\mathbf{R}_R^{(1)}-\mathbf{R}_{R,D}^{(1)}\left(\mathbf{R}_D^{(1)}\right)^{-1}\mathbf{R}_{D,R}^{(1)} \tag{6}$$

where $\mathbf{R}_R^{(1)}$ and $\mathbf{R}_D^{(1)}$ denote the covariance of the received signal at R and D, while $\mathbf{R}_{R,D}^{(1)}$ denotes the cross-correlation between the relay and destination observations:

$$\mathbf{R}_{R,D}^{(1)}\triangleq E\left[\mathbf{y}_R^{(1)}\left(\mathbf{y}_D^{(1)}\right)^H\right]=\mathbf{H}_1\mathbf{R}_S^{(1)}\mathbf{H}_0^H\triangleq\mathbf{R}_{D,R}^H \tag{7}$$

$$\mathbf{R}_D^{(1)}=\mathbf{H}_0\mathbf{R}_S^{(1)}\mathbf{H}_0^H+\sigma^2\mathbf{I}_{N_D} \tag{8}$$

$$\mathbf{R}_R^{(1)}=\mathbf{H}_1\mathbf{R}_S^{(1)}\mathbf{H}_1^H+\sigma^2\mathbf{I}_{N_R} \tag{9}$$

After some matrix manipulations, equations (6)-(9) yield:

$$\mathbf{R}_{R|D}^{(1)}=\mathbf{H}_1\left(\mathbf{R}_S^{(1)}\right)^{1/2}\left(\mathbf{I}_{N_S}+\mathbf{H}_0\mathbf{R}_S^{(1)}\mathbf{H}_0^H\right)^{-1}\left(\mathbf{R}_S^{(1)}\right)^{1/2}\mathbf{H}_1^H+\sigma^2\mathbf{I}_{N_R} \tag{10}$$

The CKLT can now be defined as the matrix $\mathbf{U}^H$ such that:

$$\mathbf{R}_{R|D}^{(1)}=\mathbf{U}\mathrm{diag}\left(\mathbf{s}\right)\mathbf{U}^H \tag{11}$$

The columns of $\mathbf{U}$ are the eigenvectors of the conditional covariance matrix, and the vector $\mathbf{s}$ contains the associated eigenvalues. Note that from (10), it is clear that the matrix $\mathbf{R}_{R|D}^{(1)}-\sigma^2\mathbf{I}_{N_R}$ is positive semi-definite and therefore:

$$s_i \geq \sigma^2 \quad \forall i \in \{1, 2, \ldots, N_R\} \tag{12}$$

It is shown in Appendix A that the rate-distortion coding of vector $\mathbf{y}_R^{(1)}$ with side information $\mathbf{y}_D^{(1)}$ at the decoder can be modeled by the following relationship:

$$\hat{\mathbf{y}}_R^{(1)} = \mathbf{U}\mathbf{A}\mathbf{U}^H\mathbf{y}_R^{(1)} + \mathbf{U}\mathbf{A}\boldsymbol{\psi} + \mathbf{U}\mathbf{K}\mathbf{y}_D^{(1)} \tag{13}$$

Where

- $\boldsymbol{\psi}$ is a vector of i.i.d. components ψ_i of variance η_i called the compression noise:

$$\psi_i \sim \mathbb{CN}(0, \eta_i) \tag{14}$$

- $$\mathbf{A} \triangleq \mathrm{diag}(\mathbf{a}) \text{ with } a_i \triangleq s_i / (s_i + \eta_i) \tag{15}$$

- $$\mathbf{K} \triangleq (\mathbf{I} - \mathbf{A})\mathbf{U}^H\mathbf{R}_{R,D}^{(1)}\left(\mathbf{R}_D^{(1)}\right)^{-1} \tag{16}$$

The relationship (13) is illustrated on Fig. 1:

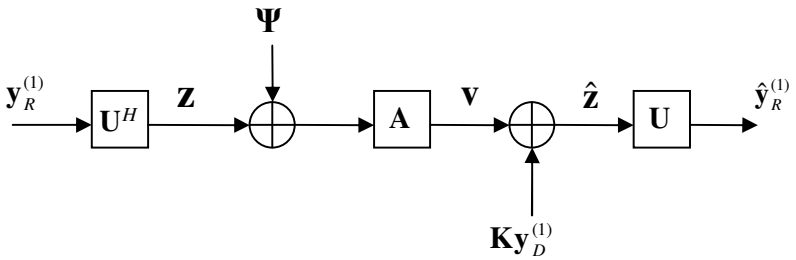

Figure 1: Source coding of Gaussian vector $\mathbf{y}_R^{(1)}$ with side information $\mathbf{y}_D^{(1)}$ at the decoder.

The coding scheme consists in applying a CKLT to each source output $\mathbf{y}_R^{(1)}$, followed by independent WZ encoding of each CKLT output sequence z_i, where $\mathbf{z} \triangleq \mathbf{U}^H\mathbf{y}_R^{(1)}$. The destination then performs WZ decoding to obtain $\hat{z}_i$ and applies the inverse CKLT to $\hat{\mathbf{z}}$. Defining r_i as the WZ coding rate for the ith CKLT output z_i, the relationship between this rate and the compression noise is:

$$r_i = \log(1 + s_i / \eta_i) \tag{17}$$

Moreover, defining d_i as the squared distortion on the ith component z_i of the transformed vector $\mathbf{z}$, the following relationship holds between the compression noise and the distortion:

$$d_i = s_i\eta_i / (s_i + \eta_i) \tag{18}$$

Inserting (18) into (17) yields:

$$r_i = \log(s_i / d_i) \tag{19}$$

It is shown in Appendix A that the total quadratic distortion is

$$\delta \triangleq E\left[\left\|\hat{\mathbf{y}}_R^{(1)} - \mathbf{y}_R^{(1)}\right\|^2\right] = \sum_{i=1}^{N_R} d_i \tag{20}$$

The above results are summarized in the following proposition:

Proposition 1:

The source coding of vector $\mathbf{y}_R^{(1)}$ *with side information* $\mathbf{y}_D^{(1)}$ *at the destination can be performed at a rate* ρ *by applying a Conditional Karhunen-Loève Transform (CKLT) at the relay followed by separate Wyner-Ziv (WZ) coding of each CKLT output sequence at rate* $r_i \geq 0$ *such that:*

$$\sum_{i=1}^{N_R} r_i \leq \rho \tag{21}$$

The following relationship holds between the WZ coding rate, the distortion and the compression noise for each component of the CKLT-transformed source:

$$r_i = \log\left(1 + s_i / \eta_i\right) = \log\left(s_i / d_i\right) \tag{22}$$

In particular, the rate-distortion trade-off $\rho(\delta)$ *is achieved by the reverse-waterfilling algorithm:*

$$d_i = \begin{cases} \lambda & if \;\; 0 \leq \lambda < s_i \\ s_i & otherwise \end{cases} \text{ with } \lambda \text{ s.t. } \sum_{i=1}^{N_{eig}} d_i = \delta \tag{23}$$

Proof: See Appendix A.

Note that in the scalar case $N_R = 1$, $d_1 = \delta$ and Proposition 1 boils down to the rate-distortion coding results of [12]. The relationship $\eta_i = s_i d_i / \left(s_i - d_i\right)$ shows that the compression noise on a component of the transformed vector is approximately equal to the distortion when the latter is small. However, when $r_i \to 0$ (i.e. no bit is allocated to represent the ith eigenmode) or equivalently $d_i \to s_i$, then $\eta_i \to +\infty$. The last part of Proposition 1 is the generalization of a well-known result (see e.g. §13.3.3 in [21]): the rate-distortion function of parallel Gaussian source is obtained by allocating the distortion according to a reverse-waterfilling algorithm on the eigenvalues of the source covariance matrix. Here, equation (23) corresponds to the same algorithm applied to the eigenvalues of the

conditional covariance matrix of $\mathbf{y}_R^{(1)}$ given $\mathbf{y}_D^{(1)}$. The term reverse-waterfilling comes from the fact that it can be implemented by progressively increasing the distortion level on every component of the transformed vector until the total distortion δ is reached, under the constraint that the distortion on the ith component cannot exceed the ith eigenvalue of the conditional covariance matrix. As shown by equation (6), the latter reduces to the covariance when the product $\mathbf{R}_{R,D}^{(1)}\left(\mathbf{R}_D^{(1)}\right)^{-1}\mathbf{R}_{D,R}^{(1)}$ goes to zero. This happens for instance when the SNR at the destination is small, then whatever the value of the cross-correlation $\mathbf{R}_{R,D}^{(1)}$, the product $\mathbf{R}_{R,D}^{(1)}\left(\mathbf{R}_D^{(1)}\right)^{-1}\mathbf{R}_{D,R}^{(1)}$ tends towards zero. This also happens for instance if $\mathbf{R}_S^{(1)}=\left(P_S/N_S\right)\mathbf{I}_{N_S}$ and the channels $\mathbf{H}_0$ and $\mathbf{H}_1$ are orthogonal. In these two cases, the CKLT degenerates into a KLT and the side information at the destination cannot be exploited to reduce the rate required at the relay to encode its observation.

C. Impact of the compression on the CF achievable rate

In this section, we assume that the compression at the relay is performed as described in §II.B, and we exploit the results of Proposition 1 in order to clarify the relationship between the achievable rate of the CF strategy described in §II.A and the compression noise defined in §II.B. The expression of the CF achievable rate is given by the following proposition:

Proposition 2:

The partial CF coding strategy defined *in §II.A achieves a rate given by the solution of the following optimization problem:*

$$R_{CF}=\max_{\substack{t\in[0;1],\ \boldsymbol{\eta}>0\\ \mathbf{R}_S^{(1)}\succeq 0,\ \mathbf{R}_S^{(2)}\succeq 0,\ \mathbf{R}_R^{(2)}\succeq 0\\ \mathrm{tr}\left(\mathbf{R}_S^{(1)}\right)\le P_S,\ \mathrm{tr}\left(\mathbf{R}_S^{(2)}\right)\le P_S,\ \mathrm{tr}\left(\mathbf{R}_R^{(2)}\right)\le P_R}} tR_0+(1-t)R_2 \tag{24}$$

where

- R_0 is equal to the capacity of a virtual MIMO channel $\mathcal{C}\left(\mathbf{R}_S^{(1)},\tilde{\boldsymbol{\Sigma}}^{-1/2}\tilde{\mathbf{H}}\right)$
- $$\tilde{\mathbf{H}}\triangleq\begin{bmatrix}\mathbf{H}_0\\ \mathbf{H}_1\end{bmatrix}\ \text{and}\ \tilde{\boldsymbol{\Sigma}}\triangleq\begin{bmatrix}\boldsymbol{\Sigma}_D & \mathbf{0}\\ \mathbf{0} & \boldsymbol{\Sigma}_R+\mathbf{U}\mathrm{diag}\left(\boldsymbol{\eta}\right)\mathbf{U}^H\end{bmatrix} \tag{25}$$
- $\boldsymbol{\eta}$ is subject to the following inequality constraint:

$$t\sum_{i=1}^{N_R}\log\left(1+s_i/\eta_i\right)\le(1-t)\,R_1 \tag{26}$$

- (R_1,R_2) is constrained to lie within the capacity region of the MAC from (S,R) to D.

Proof:

From the definition of the coding strategy in §II.A:

$$R_0 = I\left(\mathbf{x}_S^{(1)};\mathbf{y}_D^{(1)},\hat{\mathbf{y}}_R^{(1)}\right) \tag{27}$$

where $\hat{\mathbf{y}}_R^{(1)}$ is given by equation (13):

$$\hat{\mathbf{y}}_R^{(1)} = \mathbf{U}\mathbf{A}\mathbf{U}^H\mathbf{y}_R^{(1)} + \mathbf{U}\mathbf{A}\boldsymbol{\psi} + \mathbf{U}\mathbf{K}\mathbf{y}_D^{(1)}$$

Removing $\mathbf{U}\mathbf{K}\mathbf{y}_D^{(1)}$ from $\hat{\mathbf{y}}_R^{(1)}$ does not affect the mutual information (27), and from the property (47) multiplying the remainder by $\mathbf{U}\mathbf{A}^{-1}\mathbf{U}^H$ does not affect (27) either. Therefore,

$$I\left(\mathbf{x}_S^{(1)};\mathbf{y}_D^{(1)},\hat{\mathbf{y}}_R^{(1)}\right) = I\left(\mathbf{x}_S^{(1)};\mathbf{y}_D^{(1)},\tilde{\mathbf{y}}_R^{(1)}\right) \tag{28}$$

where $\tilde{\mathbf{y}}_R^{(1)}$ is defined as:

$$\tilde{\mathbf{y}}_R^{(1)} \triangleq \mathbf{y}_R^{(1)} + \mathbf{U}\boldsymbol{\psi} \tag{29}$$

From (27)-(29), the rate R_0 can therefore be expressed as:

$$R_0 = \mathcal{C}\left(\mathbf{R}_S^{(1)},\tilde{\boldsymbol{\Sigma}}^{-1/2}\tilde{\mathbf{H}}\right) \tag{30}$$

Finally, equation (26) is a direct application of Proposition 1. This concludes the proof.

It can be noticed that in the single-antenna case Proposition 2 boils down to the CF achievable rates derived in [6]. From the above equations, it can also be observed that R_{CF} meets the MIMO capacity bound when $t\to 1$ and $\boldsymbol{\eta}\to\mathbf{0}$. Unfortunately, equation (26) shows that achieving these two conditions simultaneously would require $R_1\to+\infty$. The trade-off which maximizes R_{CF} is investigated in the next section.

III. Maximizing the Achievable Rate

The maximization of the CF achievable rate as formulated in Proposition 2 is a highly non-trivial non-convex optimization problem. In this section we start by dividing this problem into sub-problems which can be solved by

means of convex optimization techniques. First, in §III.A a closed-form expression for the optimum WZ coding rates is derived assuming all other parameters constant, then in §III.B the joint optimization of the WZ coding rates and the transmit covariance matrices at the source and relay is addressed. Finally in §III.C an iterative procedure is proposed for the whole problem. The optimality of this procedure cannot be guaranteed, but simulations in §IV.B will show that it can approach capacity in some scenarios of practical interest.

A. Optimization of the compression at the relay

Let assume that $\mathbf{R}_S^{(1)}$, R_1, R_2 and t are fixed and only R_0 needs to be maximized with respect to the compression noise $\boldsymbol{\eta}$. We show in Appendix B that R_0 can be decomposed as follows:

$$\begin{aligned} &R_0 = R_{0,d} + R_{0,r} \\ &R_{0,d} \triangleq \mathcal{C}\left(\mathbf{R}_S^{(1)}, \mathbf{H}_0\right) \text{ and } R_{0,r} \triangleq \sum_{i=1}^{N_R} \log\left(\frac{s_i + \eta_i}{\sigma^2 + \eta_i}\right) \end{aligned} \tag{31}$$

The term $R_{0,d}$ is the mutual information between the signal transmitted by the source and the destination observation during the first slot. The term $R_{0,r}$ is the additional information brought by the compressed relay observation reconstructed at the destination. Note that only $R_{0,r}$ depends on $\boldsymbol{\eta}$ and shall be optimized. A contribution to $R_{0,r}$ can be associated to each component of the transformed relay observation. Equation (31) shows that this contribution is maximized and tends to $\log(s_i / \sigma^2)$ when the compression noise variance η_i is negligible compared to the thermal noise variance σ^2. As expected, this contribution vanishes when $\eta_i \to +\infty$, reflecting the fact that a highly distorted signal component cannot convey information. Also note that the components for which equality holds in (12) do not contribute to $R_{0,r}$, although they affect the total distortion δ. The optimization of $R_{0,r}$ with respect to $\boldsymbol{\eta}$ yields the following proposition:

Proposition *3:*

The Wyner-Ziv coding rates r_i *and* compression *noise* η_i *which maximize the achievable rate of the partial CF relaying strategy are such that:*

$$r_i = \left[\mu + \log\left(\left(s_i - \sigma^2 \right) / \sigma^2 \right) \right]^+ \tag{32}$$

$$\eta_i = s_i / \left(2^{r_i} - 1 \right) \tag{33}$$

where μ *is a constant such that the following constraint is satisfied:*

$$t \sum_{i=1}^{N_R} r_i \le (1-t) R_1 \tag{34}$$

Proof: See Appendix B.

In equation (32), the ratio $(s_i - \sigma^2)/\sigma^2$ can be interpreted as a useful signal to thermal noise ratio per component, since equality occurs in (12) when the *i*th component contains only noise. From (31), it is clear that an eigenmode with a large $(s_i - \sigma^2)/\sigma^2$ ratio will be a large contributor to $R_{0,r}$ and ultimately to the CF achievable rate R_{CF}, provided that the compression noise variance η_i on this eigenmode is low-enough. The term $\log\left((s_i - \sigma^2)/\sigma^2\right)$ in (32) can thus be viewed as a rate penalty for the eigenmodes which have a lower potential contribution to R_{CF}. The penalty tends to $-\infty$ when $s_i \to \sigma^2$, and in this case the corresponding CKLT output will not be encoded. It is interesting to compare equation (32) with the reverse-waterfilling algorithm of proposition 1 that is obtained when minimizing the total distortion under a rate constraint. In reverse waterfilling, the algorithm tries to spread the total distortion δ uniformly, under the constraint $d_i \le s_i$. Such a strategy leads to $r_i = \log(s_i / d)$ with d a constant corresponding to the distortion on the eigenmodes which are finally encoded (i.e. $r_i > 0$). If the average number of bits available per vector $R_1(1-t)/t$ is large enough, even the eigenmodes such that $s_i = \sigma^2$ will be encoded, although they cannot convey any information, which shows that reverse-waterfilling is sub-optimum in our problem. The CF achievable rate loss when adopting the reverse-waterfilling algorithm instead of that of proposition 3 is illustrated by simulations in [17].

Having optimized the compression at the relay, we now address another sub-problem which is the optimization of source and relay transmit covariance during the second slot, before dealing with the whole problem (24).

B. Source and relay precoding during the second slot

Before introducing a joint optimization procedure, the expression of the rates R_1 and R_2 shall be clarified. The partial CF strategy described in §II.A considers the simultaneous transmission of the two independent messages ω_1 and ω_2 simultaneously to D during the second slot. Fixed $\mathbf{R}_S^{(2)}$ and $\mathbf{R}_R^{(2)}$, the MAC achievable rate region is defined by the following pentagon (see e.g. ch. 10 in [3]):

$$R_1 \le \mathcal{C}\left(\mathbf{R}_R^{(2)}, \mathbf{H}_2 / \sigma\right) \tag{35}$$

$$R_2 \le \mathcal{C}\left(\mathbf{R}_S^{(2)}, \mathbf{H}_0 / \sigma\right) \tag{36}$$

$$R_1 + R_2 \le \mathcal{C}\left(\begin{bmatrix} \mathbf{R}_S^{(2)} & \mathbf{0} \\ \mathbf{0} & \mathbf{R}_R^{(2)} \end{bmatrix}, \begin{bmatrix} \mathbf{H}_0 & \mathbf{H}_2 \end{bmatrix} / \sigma\right) \tag{37}$$

Equality in (36) is reached when D decodes ω_1 first, then removes the contribution of $\mathbf{x}_R^{(2)}$ from $\mathbf{y}_D^{(2)}$ so that ω_2 is decoded interference-free. Equality in (35) is reached by decoding ω_2 first, and the sum-rate side (37) is reached by time-sharing between the two decoding orders. Solving (24) under the constraints (26) and (35)-(37) seems a very difficult non-convex optimization problem. We start by simplifying it by imposing a decoding order at the destination:

Proposition 4:

In the single antenna case, the achievable rate of partial Compress-and-Forward is maximum when the relayed message is decoded first.

Proof: See Appendix C

In other words, R_{CF} varies along the sum-rate side of the MAC achievable rate pentagon and is optimum only at the corner of this pentagon corresponding to the decoding of ω_1 prior to ω_2. This is contrary to a statement in a footnote of [6]. Note that we were able to prove this proposition only in the single antenna case, but in the rest of the paper we conjecture that the proposition remains valid in the multiple antenna case.

Having fixed the decoding order at D, we now turn to the optimization of R_{CF} w.r.t. $\mathbf{R}_S^{(2)}$ and $\mathbf{R}_R^{(2)}$ (or equivalently w.r.t. R_1 and R_2) for a given t. Let first consider the "full CF" case where only the Relay is allowed to transmit during the 2nd slot (i.e. $R_2 = 0$, $\mathbf{R}_S^{(2)} = \mathbf{0}$). In this case maximizing R_1 w.r.t. $\mathbf{R}_R^{(2)}$ is a mandatory preliminary step in the maximization of R_{CF}, and this maximization consists in transmit power waterfilling on the eigenmodes of $\mathbf{H}_2$ as in [4]. Indeed, from (32) and (34), it is clear that given $\mathbf{R}_S^{(1)}$ (i.e. given $\mathbf{s}$), increasing R_1 allows to increase μ and therefore to increase each rate r_i or equivalently from (33) to reduce $\boldsymbol{\eta}$ component-wise. This increases R_0 which in the full CF case is the only contributor to R_{CF}. For partial CF, a larger R_{CF} is achieved in non-trivial cases by letting $R_2 > 0$. From Proposition 4, the decoding of ω_2 is interference-free. Therefore, R_2 can be maximized w.r.t. $\mathbf{R}_S^{(2)}$ by waterfilling on the eigenmodes of $\mathbf{H}_0$. Likewise, given $\mathbf{R}_S^{(2)}$, R_1 can be maximized w.r.t. $\mathbf{R}_R^{(2)}$ by waterfilling on the eigenmodes of $\left(\boldsymbol{\Sigma}_D + \mathbf{H}_0 \mathbf{R}_S^{(2)} \mathbf{H}_0^H\right)^{-1/2} \mathbf{H}_2$. In the following, we decide to optimize R_2 and R_1 successively in the order described above. The intuition behind this simplification is the following: CF is known to outperform DF only when the SNR is much larger on the R-D link than on the S-D link, therefore we can restrict the optimization of CF to this scenario. Thus, we assume that the signal to noise-plus-interference ratio when decoding ω_1 is high. In this case, waterfilling amounts to equal power allocation over all the eigenmodes of $\mathbf{H}_2$, and the impact of $\mathbf{R}_S^{(2)}$ on the optimization of R_1 becomes negligible.

C. Iterative procedure for joint optimization

Having derived in the previous section an optimization procedure for the Source and Relay precoders during the second slot, we now assume that R_1, R_2 and t are fixed and address the maximization of R_0 w.r.t. the Source precoder during the first slot and the compression at the relay, before addressing the whole problem (24).

The following two sub-problems can be identified:

- Fixed $\mathbf{U}$ and $\boldsymbol{\eta}$, the optimization of R_0 w.r.t. $\mathbf{R}_S^{(1)}$ in (30) is obtained by transmit power waterfilling on the eigenmodes of $\tilde{\boldsymbol{\Sigma}}^{-1/2} \tilde{\mathbf{H}}$ as in [4].

- Fixed $\mathbf{R}_S^{(1)}$, the CKLT is determined and the compression noise $\boldsymbol{\eta}$ which maximizes R_0 is given by Proposition 3.

This suggests the use of the iterative Gauss-Seidel algorithm (see §2.7 in [23]) to jointly optimize $\mathbf{R}_S^{(1)}$ and $\boldsymbol{\eta}$. Integrating this algorithm with the source and relay precoder optimization of §III.B, we now propose the following procedure for solving (24):

Iterative procedure for maximizing R_{CF} *:*

1. Maximize R_2 w.r.t. $\mathbf{R}_S^{(2)}$ by transmit power waterfilling on the eigenmodes of $\mathbf{H}_0$
2. Maximize R_1 w.r.t. $\mathbf{R}_R^{(2)}$ by transmit power waterfilling on the eigenmodes of $\left(\boldsymbol{\Sigma}_D + \mathbf{H}_0 \mathbf{R}_S^{(2)} \mathbf{H}_0^H\right)^{-1/2} \mathbf{H}_2$
3. Outer loop: Maximize R_{CF} w.r.t. $t \in [0;1]$

   Inner-loop: Maximize R_0 w.r.t. $\mathbf{R}_S^{(1)}$ and $\boldsymbol{\eta}$ by iterating between steps a. and b. ($\tilde{\boldsymbol{\Sigma}}$ is initialized to $\sigma^2 \mathbf{I}_{N_R+N_D}$):

   a. Maximize R_{CF} w.r.t. $\mathbf{R}_S^{(1)}$ by transmit power waterfilling on the eigenmodes of $\tilde{\boldsymbol{\Sigma}}^{-1/2}\tilde{\mathbf{H}}$.

   b. Maximize R_{CF} w.r.t. $\boldsymbol{\eta}$ by optimum Wyner-Ziv coding rates allocation (Proposition 3).

The outer loop is a one-dimensional maximization of $R_{CF} = tR_0 + (1-t)R_2$ with respect to $t \in [0;1]$. It can pratically be performed by uniformly quantizing this interval, and the quantization step determines the time accuracy of the final solution. Unfortunately the convergence of the Gauss-Seidel algorithm in the inner-loop cannot be guaranteed. Indeed, the conditions for convergence given in §2.7 of [23] cannot be verified, because $\mathbf{U}$ depends on $\mathbf{R}_S^{(1)}$ and therefore the assumption that $\tilde{\boldsymbol{\Sigma}}$ is fixed in the optimization step 3.a is an approximation. Thus the joint optimization procedure cannot be claimed optimal. The convergence of the inner-loop will be assessed in the next section for realistic SNR values, and the sub-optimality of the whole procedure will be evaluated by a comparison with the cut-set bound and the achievable rate of other relaying strategies.

IV. Practical Implementation Considerations and Simulation Results

In sections II and III an achievable rate was derived for the partial CF strategy with full CSI. In this section, we start by discussing implementation aspects such as practical source coding, CSI availability and the extension to MIMO-OFDM. We then analyze by achievable rate simulations the optimization procedure of section III, and compare it in scenarios of practical interest with other capacity bounds.

A. Practical implementation considerations

We review below several constraints which arise when considering a pratical implementation of CF relaying.

1) MIMO-OFDM transmission

The MIMO results presented in this paper can be easily extended to MIMO-OFDM by considering block-diagonal frequency-domain channel matrices. For instance the channel matrix on the S-D link can be defined as:

$$\mathbf{H}_0 \triangleq \mathrm{diag}\left(\mathbf{H}_{0,i}\right)_{i=1,\ldots,N_C} \tag{38}$$

where $\mathbf{H}_{0,i}$ is the $N_D \times N_S$ channel matrix on the *i*th subcarrier out of a total of N_C subcarriers. From parallel channel arguments (see e.g. 10.4 in [21]), the signal transmitted by S on different subcarriers can be assumed uncorrelated and therefore the conditional covariance will also be block-diagonal:

$$\mathbf{R}_{R|D}^{(1)} = \mathrm{diag}\left(\mathbf{R}_{i,R|D}^{(1)}\right)_{i=1,\ldots,N_C} \tag{39}$$

From (39), the CKLT can also be expressed as a block-diagonal matrix:

$$\mathbf{R}_{R|D}^{(1)} = \mathbf{U}\mathrm{diag}\left(\mathbf{s}\right)\mathbf{U}^H \quad \text{with} \quad \mathbf{U} = \mathrm{diag}\left(\mathbf{U}_1, \mathbf{U}_2, \ldots, \mathbf{U}_{N_C}\right) \tag{40}$$

Therefore a per-subcarrier CKLT $\mathbf{U}_i^H$ can be defined. It must be applied to the vector of length N_R formed by stacking the *i*th Discrete Fourier Transform output for each antenna at the relay. Achievable rates are now obtained by summing independent contributions over the N_C subcarriers. The optimum WZ coding rates can still be computed from Proposition 3, with the summation index *i* now ranging from 1 to $N_C N_R$, reflecting the fact that the total rate R_1 available at the relay to quantize its observation shall be shared between all the spatial eigenmodes over all the subcarriers.

2) Practical channel coding and modulation

A state-of-the-art broadband wireless system (e.g. [2]) typically encodes finite-length packets with a turbo-code or an LDPC, and maps the output onto finite alphabet symbols (e.g. QAM). In this case, the precoders and rate allocation algorithms presented in this paper are not directly applicable, since they are designed assuming Gaussian i.i.d. codewords of infinite length. This issue is well-known in OFDM systems and solutions such as bit and power loading have been proposed as realistic alternatives to waterfilling [24]. Another more recent approach includes the finite alphabet assumption in the information-theoretic optimization [25]. The extension to finite length and finite alphabet codewords is out of the scope of this paper but is an interesting topic for future work.

3) Practical WZ coding

The CF achievable rates in sections II and III are derived assuming an ideal WZ coding applied to each CKLT output. Practical WZ coding implementations are reviewed in [26]. The WZ encoder consists in a quantizer followed by a Slepian-Wolf (SW) encoder. In order to approach the rate-distortion trade-off, the quantizer shall be rate-distortion[3] approaching and the channel code used for SW coding shall be capacity-approaching. In [27] and [26], practical WZ coding operating at less than 0.5dB from the rate-distortion curve is obtained by combining LDPC-based SW Coding with trellis-coded quantization. Therefore, quasi-ideal rate-distortion coding is a realistic assumption from an implementation standpoint.

4) Source coding without side-information

The rate-distortion trade-off for gaussian vectors without side information is given in section 13.3.3 of [21]. The relay shall apply a KLT followed by rate-distortion (without side-information) encoding of each output. In this case, the rate allocation of Proposition 3 cannot be applied, but the total distortion can still be minimized by reverse-waterfilling. An even simpler source coding which does not even require CSI consists in applying per antenna quantization without any linear transform. In this case, an OFDM signal can be quantized directly in time-domain, which is not possible for techniques exploiting CSI. Overall these source-coding strategies result in a simpler implementation but may lead to a large reduction of the achievable rate.

[3] Here, for the quantizer we are refering to the rate-distortion trade-off without side information.

5) CSI availability

In point-to-point MIMO, CSI can be obtained by either explicit signaling or channel reciprocity. However, for cooperative relaying explicit signaling is always required because the knowledge of the three channel matrices is required to perform the maximization (24). The quantization of CSI to minimize the signaling load and its effect on the CF achievable rate are therefore interesting topics for future studies. Note that the per-antenna quantization as described in IV.A.4) does not demand explicit CSI signaling but leads to a reduced achievable rate.

B. Simulation results

In this section we analyze by simulations the achievable rate performance of partial CF, and perform comparisons with other relaying strategies for the TDD MIMO relay channel with full CSI. In the comparison, we will consider non-cooperative DF and partial DF strategies for which achievable rates are computed in [19], but not LR as explained in the introduction. In non-cooperative DF, only R transmits during the second slot and D does not process the signal received during the first slot. The capacity of the general relay channel remains unknown, and to the best of our knowledge the tightest known upper-bound on the capacity of the (fixed-dynamic) TDD MIMO relay channel with full CSI is the cut-set bound computed in [18] and solved by convex optimization in [19].

We consider a downlink TDD mobile relaying scenario where a Base Station (S) equipped with $N_S = 4$ antennas per sector transmits to a dual antenna Mobile Station (D) which is assisted by another dual antenna MS (R) in its neighborhood. The average SNRs on the S-D, S-R and R-D links are denoted respectively γ_0, γ_1 and γ_2. In the simulations we assume that $\gamma_0 = \gamma_1$ varies between 0dB and 20dB and γ_2 is fixed to 30dB (i.e. R and D are close to each other). The MIMO fading on each link is modeled by i.i.d. complex Gaussian components. On Figure 2, the average CF achievable rate obtained from the iterative procedure of §III.C is plotted (solid black curves) for both the partial and full CF strategies. The dashed curves represent the MIMO channel capacity of the S-D link, the achievable rate with the non-cooperative DF and with the partial DF strategies. On the figure, the three curves associated to full CF, partial DF and the S-D link capacity (i.e. no relaying) almost overlap and it can be observed that only the partial CF strategy yields a significant rate increase over the S-D link capacity, whereas non-cooperative DF relaying achieves a rate much lower than the S-D link capacity. Further simulation results (not

plotted here) show that at $\gamma_2 = 30\text{dB}$, partial DF starts to outperform partial CF only when γ_1 is at least 10dB higher than γ_0, i.e. when the capacity on the S-R link becomes much higher than on the S-D link. On the figure, the dotted curves represent the cut-set bound and the capacity of the Virtual MIMO channel from S to (R,D), obtained by waterfilling the source transmit power on the eigenmodes of $\tilde{\mathbf{H}}$. The cut-set bound is much lower than the V-MIMO bound, which shows that although γ_2 is high, the capacity of the R-D link cannot be assumed infinite. Partial CF achieves a rate very close to the cut-set bound and is therefore almost capacity-achieving, although in our simulations we stopped the inner-loop after only two iterations. Thus, partial CF seems well-suited to this downlink mobile cooperation scenario, and improving the optimization procedure could in this case only yield a marginal rate increase.

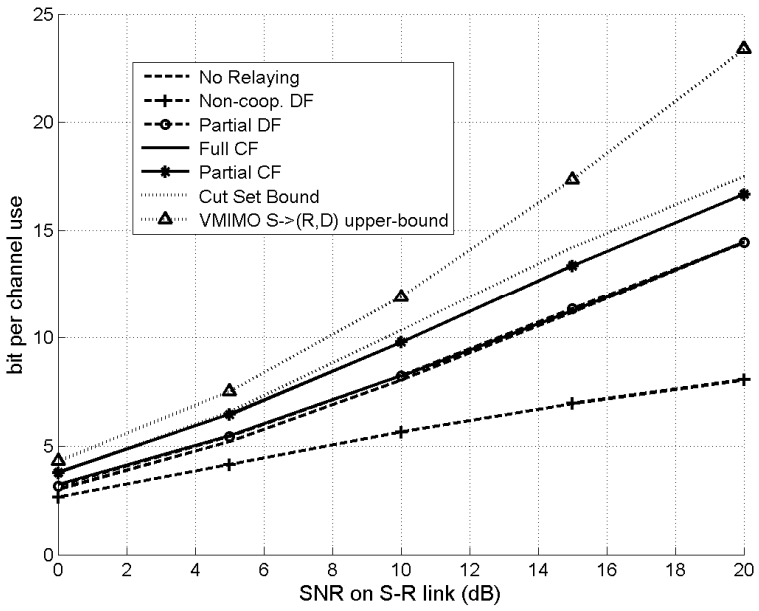

Figure 2: Average CF achievable rate and comparison with DF strategies and capacity upper-bounds

On Figure 3 we compare the achievable rate obtained by applying the complete optimization procedure of §III.C with a simpler optimization in which we fix $\mathbf{R}_S^{(1)} = \mathbf{R}_S^{(2)} = \left(P_S / N_S\right)\mathbf{I}_{N_S}$ and $\mathbf{R}_R^{(2)} = \left(P_R / N_R\right)\mathbf{I}_{N_R}$ and only optimize t and $\boldsymbol{\eta}$. As explained in [3], optimizing the source transmit covariance increases the S-D link capacity by a 3dB power gain (because $N_S = 2N_D$) plus a waterfilling gain that becomes negligible at high SNR. The SNR gain is only about 1dB for the cut-set bound and for partial CF. Intuitively, this smaller gain is justified by the fact that

the number of transmit antennas is equal to the number of receive antennas on the 4×4 virtual MIMO channel $\tilde{\mathbf{H}}$ and therefore the rate R_0, which is the main component of R_{CF} at low γ_0, does not benefit from any power gain.

To conclude, partial CF outperforms partial DF and is almost capacity-achieving in scenarios where the capacity of the R-D link is sufficiently high. This condition may occur either in cellular uplink where a fixed relay benefits from a strong link to the BS (see simulation results in [17]) or in the downlink mobile relaying scenario of this paper. In this second case, the optimization of the source transmit covariance matrix makes sense especially if the number of transmit antennas at the BS is larger than the total number of antennas at the relay and destination.

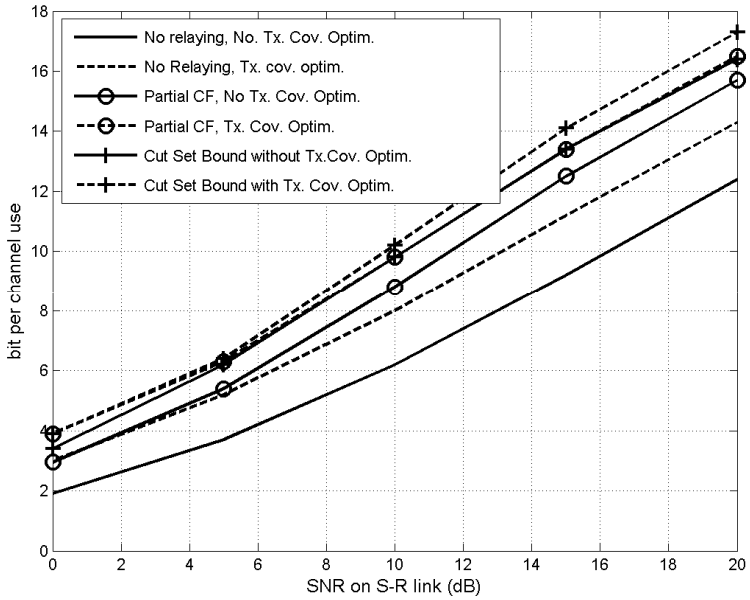

Figure 3: Capacity bounds with and without transmit covariance optimization.

V. CONCLUSIONS

In this paper, we apply distributed vector compression techniques to perform Compress-and-Forward relaying on the Gaussian TDD MIMO relay channel with full CSI. We propose an iterative procedure which jointly optimizes the transmit covariance matrices, the compression at the relay and the time resource allocation. In some scenarios of practical interest, the achievable rate approaches the capacity, outperforming other relaying strategies. We also address some implementation-related issues, such as CSI availability, the extension to MIMO-OFDM and practical source and channel coding. For these implementation issues, we identify several topics which deserve more investigation, for instance the impact of imperfect and quantized CSI. Other promising areas of research include the

MIMO signal processing design for other transmission strategies such as those designed for two-way, two-path [10], multi-user and multiple-relay channels (e.g. [8] and [10]) or for coordinated cellular networks (e.g. [13]).

APPENDIX

A. *Proof of Proposition 1*

The proof can be obtained as a special case of Theorem 3 and corollary 4 in [16]. The latter considers a Gaussian source vector $\mathbf{x}$ which is split into two (correlated) parts $\mathbf{x}_1$ and $\mathbf{x}_2$ and shall be reconstructed from a compressed version of $\mathbf{x}_1$ and a noisy observation of $\mathbf{x}_2$. Our problem is slightly different as we are only interested in reconstructing $\mathbf{x}_1$ and not the whole vector $\mathbf{x}$, thus we do not take into account the distortion on $\mathbf{x}_2$. As in [16], let first consider the rate-distortion coding of the Gaussian vector $\mathbf{y}_R^{(1)}$ with side information $\mathbf{y}_D^{(1)}$ at both the encoder and the decoder. It can be realized by the distribution $f\left(\hat{\mathbf{y}}_R^{(1)} \middle| \mathbf{y}_R^{(1)}, \mathbf{y}_D^{(1)}\right)$ which is generated on Figure 4, where $\mathbf{A}$ and $\boldsymbol{\psi}$ are defined by (15) and (14).

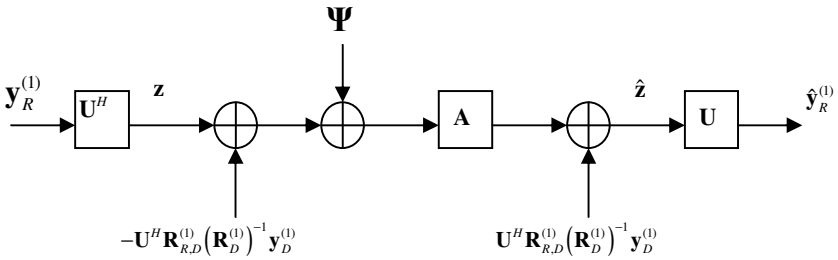

Figure 4: Rate-distortion coding of gaussian vector $\mathbf{y}_R^{(1)}$ with side information $\mathbf{y}_D^{(1)}$ at the encoder and decoder

Because the CKLT is a unitary transform, it preserves the quadratic distortion:

$$\delta \triangleq E\left[\left\|\mathbf{y}_R^{(1)} - \hat{\mathbf{y}}_R^{(1)}\right\|^2 \middle| \mathbf{y}_D^{(1)}\right] = E\left[\left|\hat{\mathbf{z}} - \mathbf{z}\right|^2 \middle| \mathbf{y}_D^{(1)}\right] \tag{41}$$

Furthermore, a fundamental property of the CKLT is that the transformed vector $\mathbf{z} \triangleq \mathbf{U}^H \mathbf{y}_R^{(1)}$ has conditionally uncorrelated components given the side information $\mathbf{y}_D^{(1)}$. Therefore, rate-distortion encoding can be performed separately on each component of the transformed vector using the same scheme as in section 3 of [12]: first, the conditional expectation $E\left[\mathbf{z} \middle| \mathbf{y}_D^{(1)}\right] = \mathbf{U}^H \mathbf{R}_{R,D}^{(1)} \left(\mathbf{R}_D^{(1)}\right)^{-1} \mathbf{y}_D^{(1)}$ is removed, then rate-distortion coding of independent Gaussian variables is performed (section 13.3.3 in [21]), and finally the conditional expectation is added back at the

destination to obtain the reconstructed signal $\hat{\mathbf{z}}$ which is transformed into $\hat{\mathbf{y}}_R^{(1)}$ by an inverse CKLT. The rate-distortion function with side information at both the encoder and decoder is found by minimizing the information rate distortion function with respect to the distribution $f\left(\hat{\mathbf{y}}_R^{(1)}\middle|\mathbf{y}_R^{(1)},\mathbf{y}_D^{(1)}\right)$ for a given sum-distortion δ:

$$r_{R|D}(\delta)=\min_{f\left(\hat{\mathbf{y}}_R^{(1)}\middle|\mathbf{y}_R^{(1)},\mathbf{y}_D^{(1)}\right)} I\left(\mathbf{y}_R^{(1)};\hat{\mathbf{y}}_R^{(1)}\middle|\mathbf{y}_D^{(1)}\right) \tag{42}$$

Note that the two schemes of Figure 1 and Figure 4 result in the same input-output relationship. From equations (13)-(16) , it is clear that the distribution is determined by the choice of a compression noise vector $\boldsymbol{\eta}$. Let now define the vector $\mathbf{d}$ as the squared distortion per component of the transformed vector $\mathbf{z}$:

$$d_i \triangleq E\left[\left|\hat{z}_i-z_i\right|^2\middle|\mathbf{y}_D^{(1)}\right] \tag{43}$$

The following lines show the relationship between the compression noise $\boldsymbol{\eta}$ and the component-wise distortion $\mathbf{d}$. From (13) and (5) we can write:

$$\hat{\mathbf{z}}-\mathbf{z}\triangleq\mathbf{U}^H\left(\hat{\mathbf{y}}_R^{(1)}-\mathbf{y}_R^{(1)}\right)=\left(\mathbf{A}-\mathbf{I}\right)\mathbf{U}^H\left(\mathbf{y}_R^{(1)}-E\left[\mathbf{y}_R^{(1)}\middle|\mathbf{y}_D^{(1)}\right]\right)+\mathbf{A}\boldsymbol{\psi} \tag{44}$$

Then by definition of conditional covariance and since the matrix $\mathbf{A}$ is diagonal, inserting (44) into (43) gives:

$$d_i=\left(a_i-1\right)^2 s_i+a_i^2\eta_i \tag{45}$$

Finally, replacing a_i in (45) by its definition given by (15) leads to equation (18): $d_i=s_i\eta_i/\left(s_i+\eta_i\right)$. Therefore, the distribution $f\left(\hat{\mathbf{y}}_R^{(1)}\middle|\mathbf{y}_R^{(1)},\mathbf{y}_D^{(1)}\right)$ is equivalently determined by the choice of either $\boldsymbol{\eta}$ or $\mathbf{d}$. It will be shown in section III.A that the distribution which minimizes the sum-distortion in (42) may not be the best for our CF relaying problem. We therefore compute the rate required to achieve a component-wise distortion $\mathbf{d}$:

$$
\begin{aligned}
r_{R|D}^{(1)}(\mathbf{d}) &= I\left(\mathbf{y}_R^{(1)};\hat{\mathbf{y}}_R^{(1)}\left|\mathbf{y}_D^{(1)}\right.\right) \\
&\overset{(a)}{=} I\left(\mathbf{U}^H\mathbf{y}_R^{(1)};\mathbf{U}^H\hat{\mathbf{y}}_R^{(1)}\left|\mathbf{y}_D^{(1)}\right.\right) \triangleq I\left(\mathbf{z};\hat{\mathbf{z}}\left|\mathbf{y}_D^{(1)}\right.\right) \\
&= h\left(\hat{\mathbf{z}}\left|\mathbf{y}_D^{(1)}\right.\right) - h\left(\hat{\mathbf{z}}\left|\mathbf{y}_D^{(1)},\mathbf{z}\right.\right) \\
&\overset{(b)}{=} h\left(\mathbf{A}\mathbf{z}+\mathbf{A}\boldsymbol{\psi}+\mathbf{K}\mathbf{y}_D^{(1)}\left|\mathbf{y}_D^{(1)}\right.\right) - h\left(\mathbf{A}\mathbf{z}+\mathbf{A}\boldsymbol{\psi}+\mathbf{K}\mathbf{y}_D^{(1)}\left|\mathbf{z},\mathbf{y}_D^{(1)}\right.\right) \\
&\overset{(c)}{=} h\left(\mathbf{z}+\boldsymbol{\psi}\left|\mathbf{y}_D^{(1)}\right.\right) - h\left(\boldsymbol{\psi}\right) \\
&\overset{(d)}{=} \sum_{i=1}^{N_R}\log\left(s_i+\frac{s_i d_i}{s_i-d_i}\right) - \sum_{i}^{N_R}\log\left(\frac{s_i d_i}{s_i-d_i}\right) \\
&= \sum_{i=1}^{N_R}\log\left(s_i/d_i\right) \triangleq \sum_{i=1}^{N_R} r_i
\end{aligned}
\tag{46}
$$

where

(a) follows from equation (13) in [20], which gives the entropy of the product of a proper complex Gaussian vector $\mathbf{x}$ by a non-singular matrix $\mathbf{M}$:

$$h\left(\mathbf{M}\mathbf{x}\right) = h\left(\mathbf{x}\right) + 2\log\left|\det\left(\mathbf{M}\right)\right| \tag{47}.$$

(b) is straightforward from equation (13)

(c) stems from (47) and the fact that the entropy of a known variable is zero.

(d) from the fact that the components of z are conditionally independent given $\mathbf{y}_D^{(1)}$ by definition of the CKLT.

As in [12], let denote by $r^*(\mathbf{d})$ the rate with side information at the decoder only. The equality between $r^*(\mathbf{d})$ and $r_{R|D}(\mathbf{d})$ follows from section 3 of [12] and from the equivalence between Fig. 1 and Fig. 2, which leads to:

$$I\left(\mathbf{y}_R;\mathbf{v}\left|\mathbf{y}_D\right.\right) = I\left(\mathbf{y}_R;\hat{\mathbf{y}}_R\left|\mathbf{y}_D\right.\right) \tag{48}$$

where $\mathbf{v}$ is defined on Figure 1. Finally, the rate-distortion function $r^*(\delta)$ is obtained by minimizing the rate under total squared distortion δ. This constrained problem is convex and the solution is given by the well-known (section 13.3.3 in [21]) reverse waterfilling algorithm.

B. Proof of Proposition 3

Applying the chain rule for mutual information to (27) gives:

$$R_0 = I\left(\mathbf{x}_S^{(1)}; \tilde{\mathbf{y}}_R^{(1)}, \mathbf{y}_D^{(1)}\right) = \underbrace{I\left(\mathbf{x}_S^{(1)}; \mathbf{y}_D^{(1)}\right)}_{\triangleq R_{0,d}} + \underbrace{I\left(\mathbf{x}_S^{(1)}; \tilde{\mathbf{y}}_R^{(1)} \mid \mathbf{y}_D^{(1)}\right)}_{\triangleq R_{0,r}} \tag{49}$$

The first term $R_{0,d}$ is equal to $\mathcal{C}\left(\mathbf{R}_S^{(1)}, \mathbf{H}_0\right)$. The second term $R_{0,r}$ can be computed as follows:

$$\begin{aligned} R_{0,r} &= h\left(\mathbf{y}_R^{(1)} + \mathbf{U}\boldsymbol{\psi} \mid \mathbf{y}_D^{(1)}\right) - h\left(\mathbf{y}_R^{(1)} + \mathbf{U}\boldsymbol{\psi} \mid \mathbf{x}_S^{(1)}, \mathbf{y}_D^{(1)}\right) \\ &\overset{(a)}{=} h\left(\mathbf{y}_R^{(1)} + \mathbf{U}\boldsymbol{\psi} \mid \mathbf{y}_D^{(1)}\right) - h\left(\mathbf{y}_R^{(1)} + \mathbf{U}\boldsymbol{\psi} \mid \mathbf{x}_S^{(1)}\right) \\ &= \log\left|\mathbf{R}_{R|D}^{(1)} + \mathbf{U}\mathrm{diag}\left(\boldsymbol{\eta}\right)\mathbf{U}^H\right| - \log\left|\mathbf{R}_{R|S}^{(1)} + \mathbf{U}\mathrm{diag}\left(\boldsymbol{\eta}\right)\mathbf{U}^H\right| \\ &\overset{(b)}{=} \log\left|\mathbf{U}\mathrm{diag}\left(\mathbf{s}+\boldsymbol{\eta}\right)\mathbf{U}^H\right| - \log\left|\mathbf{U}\left(\sigma^2\mathbf{I}_{N_R} + \mathrm{diag}\left(\boldsymbol{\eta}\right)\right)\mathbf{U}^H\right| \\ &= \sum_{i=1}^{N_R} \log\left(\left(s_i + \eta_i\right)/\left(\sigma^2 + \eta_i\right)\right) \end{aligned} \tag{50}$$

where (a) comes from the fact that $\mathbf{y}_D^{(1)}$ is a noisy version of $\mathbf{H}_0\mathbf{x}_S^{(1)}$ and (b) from the fact that white thermal noise was assumed in our signal model.

The maximization of $R_{0,r}$ w.r.t. $\boldsymbol{\eta}$ can now be performed. From (17) we have $\eta_i = s_i / \left(2^{r_i} - 1\right)$, which can be inserted into (50), resulting in the following equivalent problem:

$$\begin{aligned} &\max_{\mathbf{r}} \ \sum_{i=1}^{N_R} \log\left(\frac{s_i 2^{r_i}}{\sigma^2\left(2^{r_i}-1\right) + s_i}\right) \\ &\text{s.t.} \begin{cases} \sum_{i=1}^{N_R} r_i \le R_1 \\ r_i \ge 0 \ \text{ for } \ 1 \le i \le N_R \end{cases} \end{aligned} \tag{51}$$

It can be checked that this objective is concave in $\mathbf{r}$, and since the inequality constraints are affine the problem is convex is standard form. The KKT conditions yield after a few simple calculations the solution *(32)*.

C. Proof of Proposition 4

Let parameterize the sum-rate side of the achievable rate region (37), denoting by $\alpha \in [0;1]$ the fraction of the time during which ω_1 is decoded first, assuming that the rest of the time ω_2 is decoded first. In the single-antenna case the channel matrices are complex scalar denoted by H_0, H_1 and H_2 which are normalized in the following such that $\sigma^2 = 1$. The achievable rates read as:

$$R_1 = \alpha \log\left(1+\frac{|H_2|^2 P_R}{1+|H_0|^2 P_S}\right)+(1-\alpha)\log\left(1+|H_2|^2 P_R\right) \tag{52}$$

$$R_2 = \alpha \log\left(1+|H_0|^2 P_S\right)+(1-\alpha)\log\left(1+\frac{|H_0|^2 P_S}{1+|H_2|^2 P_R}\right) \tag{53}$$

$$R_0 = \log\left(1+|H_0|^2 P_S+|H_1|^2 P_S/(1+\eta)\right) \tag{54}$$

where

$$\eta = \left(1+|H_1|^2 P_S+|H_0|^2 P_S\right)\left(2^{R_1(1-t)/t}-1\right)^{-1}\left(1+|H_0|^2 P_S\right)^{-1} \tag{55}$$

From (4) and (52)-(55), it can be shown (after tedious calculations) that:

$$\frac{\partial R_{CF}}{\partial \alpha} = -2^{R_1(1-t)/t}\left(2^{R_1(1-t)/t}-1\right)^{-1}(1+\eta)^{-1}(1-t)\frac{\partial R_1}{\partial \alpha} \geq 0 \tag{56}$$

Since from (52) we have $\partial R_1 / \partial \alpha \leq 0$, therefore $\partial R_{CF} / \partial \alpha \geq 0$. In general, inequality (56) is strict and $\alpha = 1$ is optimum, Q.E.D.

REFERENCES

[1] Nosratinia, T. E. Hunter, A. Hedayat, "Cooperative Communication in Wireless Networks", *IEEE Communications Magazine*, Oct 2004

[2] IEEE 802.16m Task Group "Air Interface for Fixed and Mobile Broadband Wireless Access Systems - Advanced Air Interface" PAR approved by the IEEE-SA Standards Board on 6 Dec. 2006.

[3] D. Tse and P. Viswanath "Fundamentals of Wireless Communication" ISBN-13 978-0-521-84527-4, Cambridge University Press, 2005.

[4] E. Telatar, "Capacity of multi-antenna gaussian channels" *European Trans. Telecom.*, vol. 10, pp. 585--595, 1999

[5] T. M. Cover and A.A. El Gamal, "Capacity theorems for the relay channel", *IEEE Trans. on Information Theory*, vol. IT-25, No 5, pp 572-584, Sep 1979

[6] A. Høst-Madsen, J. Zhang "Capacity Bounds and Power Allocation for Wireless Relay channels" *IEEE Trans. on Information Theory*, Vol 51, N°6, pp. 2020-2040, June 2005

[7] M. Yuksel, E. Erkip "Multiple-Antenna Cooperative Wireless Systems: A Diversity–Multiplexing Tradeoff Perspective" *IEEE Trans. Inform. Theory*, Vol. 53, No 10, Oct 2007.

[8] G. Kramer, M. Gastpar, P. Gupta, "Cooperative Strategies and Capacity Theorems for Relay Networks", *IEEE Trans. on Information Theory*, Vol 51. No 9. Sep 2005

[9] A. El Gamal, M. Mohseni, S. Zahedi "Bounds on Capacity and Minimum Energy-Per-Bit for AWGN Relay Channels". *IEEE Trans. on Information Theory*, Vol 52, No 4, Apr 2006.

[10] B. Rankov, A. Wittneben, "Spectral efficient protocols for half-duplex fading relay channels", *IEEE Journal on Selected Areas in Communications (JSAC)*, Vol. 25, No. 2, Feb. 2007.

[11] A. Agustín and J. Vidal, "TDMA cooperation with spatial reuse of the relay slot and interfering power distribution information," *IEEE Int. Conf. Acoustics, Speech and Signal Processing (ICASSP)*, May 2006.

[12] A.D. Wyner "The rate-distortion function for source coding with side information at the decoder –II: General Sources" *Information and Control*, Vol 38, pp 60-80, 1978

[13] A. Sanderovich, S. Shamaï, Y. Steinberg, G. Kramer "Communication via Decentralized Processing" *IEEE Trans. Inform. Theory*, Vol. 54, No 7, Jul. 2008.

[14] A. Sanderovich, S. Shamai and Y. Steinberg "Distributed MIMO Receiver - Achievable Rates and Upper Bounds" *submitted to IEEE Trans. on Information Theory*, Sep. 2007, Avail. on Arxiv.org/abs/0710.0116.

[15] M. Gastpar, P.L. Dragotti and M. Vetterli "On compression using the distributed Karhunen Loeve Transform" *Proceedings of IEEE ICASSP*, May 2004

[16] M. Gastpar, P.L. Dragotti and M. Vetterli "The Distributed Karhunen-Loève Transform" *IEEE Trans. on Information Theory*, Vol 52, No 12, Dec. 2006.

[17] S. Simoens, O. Muñoz, J. Vidal "Achievable Rates of Compress-and-Forward Cooperative Relaying on Gaussian Vector Channels", *Proceedings of IEEE Int. Conf. Communications (ICC)*, Jun. 2007.

[18] S. Simoens, O. Muñoz, J. Vidal, A. Del Coso "Capacity Bounds for Gaussian MIMO relay channel with Channel State Information" *Proceedings of IEEE SPAWC*, Jul. 2008.

[19] S. Simoens, O. Muñoz, J. Vidal, A. Del Coso "On the Gaussian MIMO Relay Channel with full Channel State Information" *Submitted to IEEE Trans. on Signal Processing* in Sep. 2008.

[20] F.D. Neeser and J. Massey "Proper Complex Random Processes with Applications to Information Theory" *IEEE Trans. On Inform. Theory*, Vol. 39, No 4, Jul. 1993.

[21] T. Cover, J.A. Thomas "Elements of Information Theory" *Wiley-Interscience*, 1991

[22] T.K. Moon, W.C. Stirling "Mathematical Methods and Algorithms for Signal Processing" *Prentice Hall*, 2000

[23] D. P. Bertsekas "Nonlinear Programming". Athena Scientific. ISBN 1-886529-00-0. 2nd edition, 1999.

[24] P.S. Chow, J.M. Cioffi, J.A.C Bingham "A Practical Discrete Multitone Loading Algorithm for Data Transmission over Spectrallly Shaped Channels" *IEEE Trans. on Communications*, Vol 43, No 2/3/4 Feb/mar/Apr 1995.

[25] A. Lozano, A.M. Tulino, S. Verdu, "Mercury/waterfilling: optimum power allocation with arbitrary input constellations" *Int. Symposium on Information Theory (ISIT).* Sep 2005.

[26] Z. Xiong, A. D. Liveris and S. Cheng "Distributed Source Coding for Sensor Networks*" IEEE Signal Processing Magazine*, Vol 21, No 5, Sep. 2004.

[27] Z. Liu, V. Stankovic and Z. Xiong "Wyner-Ziv coding for the half-duplex relay channel" *Proceedings of IEEE ICASSP*, Mar. 2005

Sébastien Simoens graduated from Ecole Nationale Supérieure des Télécommunications (ENST) Paris in 1998 and received the PhD degree from the Signal Theory and Communications Department of the Technical University of Catalonia (UPC), Barcelona in July 2009. From 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 Nov. 2008, he has been with Thales Aerospace Division (Valence, France), working on signal processing for inertial navigation sensors.

Olga Muñoz-Medina received the M.S. degree in 1993 and the PhD degree in 1998, both in Electrical Engineering from the Universitat Politècnica de Catalunya (UPC), Spain. In 1994 she joined the Department of Signal Theory and Communications at the same University and became an Associate Professor in 2001. She teaches graduate and undergraduate courses related to Signal Processing. She has participated in industrial projects as Spread Spectrum Link for the Automatic Transfer Vehicle (ATV) with Alcatel Espacio S.A. for European Spatial Agency, and in European Comission funded projects as TSUNAMI, ROMANTIK, FIREWORKS and ROCKET. Her current research interests include physical layer aware MAC techniques, cooperative transmission for multihop systems and multi-base coordination.

Josep Vidal received the Telecommunication Engineering and the Ph. D. degrees from the Universitat Politècnica de Catalunya (UPC), Barcelona. From 1989 to 1990 he joined the LTS at the Ecole Polytechnique de Lausanne as associate researcher. He was awarded Premio Extraordinario de Doctorado in 1996. Since 1993 he has been teaching different topics in Telecommunication Engineering at UPC. His current research interests are in statistical signal processing, information and communication theory, where he has authored more than 100 journal and conference papers. Since 2000 has led UPC participation in the EC-funded projects, SATURN, ROMANTIK, FIREWORKS and ROCKET, as project coodinator in some of them. He has held research appointments with INP Toulouse and University of Hawaii.

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.