On the Optimization of Distributed Compression in Multi-Relay Cooperative Networks

Yinan Qi, Muhammad Ali Imran, Richard Demo Souza, and Rahim Tafazolli

Abstract

In this paper, we consider multi-relay cooperative networks for the Rayleigh fading channel, where each relay, upon receiving its own channel observation, independently compresses it and forwards the compressed information to the destination. Although the compression at each relay is distributed using Wyner-Ziv coding, there exists an opportunity for jointly optimizing compression at multiple relays to maximize the achievable rate. Considering Gaussian signalling, a primal optimization problem is formulated accordingly. We prove that the primal problem can be solved by resorting to its Lagrangian dual problem and an iterative optimization algorithm is proposed. The analysis is further extended to a hybrid scheme, where the employed forwarding scheme depends on the decoding status of each relay. The relays that are capable of successful decoding perform decode-and-forward and the rest conduct distributed compression. The hybrid scheme allows the cooperative network to adapt to the changes of the channel conditions and benefit from an enhanced level of flexibility. Numerical results from both spectrum and energy efficiency perspectives show that the joint optimization improves efficiency of compression and identify the scenarios where the proposed schemes outperform the conventional forwarding schemes. The findings provide important insights into the optimal deployment of relays in a realistic cellular network.

Index Terms

Multi-relay, decode-and-forward, compress-and-forward, Wyner-Ziv coding

I INTRODUCTION

Being considered as one of the potential enabling techniques for future communication networks, relaying gives rise to a plethora of interesting applications and new business opportunities. The relay model was first introduced by Van Der Meulen [1] and substantially developed by Cover and El Gamal [2]. Common relaying schemes including amplify-and-forward (AF), decode-and-forward (DF), and compress-and-forward (CF) have been investigated for single relay systems [3]-[8].

The study is then extended to multi-relay cooperative networks and analysis has been given in [9]-[14]. Cooperative diversity is analysed for AF in [9] and the sum rate and outage performance are studied for AF and DF, respectively, in [10] and [11]. An efficient pilot-based channel estimation scheme is proposed for multi-relay AF networks in [12]. Relay selection has also been widely investigated. In [13], selection criterion based on pairwise error probability is established for single carrier AF relay networks. The outage performance of DF based relay selection is analysed for cognitive relay networks and it is shown that the number of relays has a significant impact [14]. Recently, hybrid forwarding schemes have drawn lots of attention, where the relays adaptively switch between AF and DF depending on channel conditions and decoding status. It is shown that the enhanced flexibility improves not only the error performance but also achieves a significant throughput gain against conventional non-hybrid schemes [15]-[16]. However, most of these previous works on multi-relay cooperative networks are based on DF and AF, but CF is rarely addressed. The fundamental limit on the applicability of DF and AF is their high dependence on the quality of the source-relay links. With low quality source-relay links, for DF the relays might not be able to successfully decode the message and thus forward erroneous information to the destination, leading to error propagation. For AF, the relays amplify and forward noise and interference as well as the useful signals. Therefore AF is mainly useful in high-SNR environments.

This paper considers a multi-relay cooperative CF network based on Wyner-Ziv compression. Wyner-Ziv compression is proposed in [17] and [18] and further extended to multiple nodes by Gastpar in [19]. In Wyner-Ziv compression, the relays independently compress their received signals from the source and forward the compressed information to the destination. At the destination, the compressed information is decompressed with the help of side information to reconstruct the relays' receptions, which are then combined to decode the source message. The information-theoretic framework of such cooperative networks is given in [20], [21] and extended by [22] from successive decompression and decoding to a joint operation. The problem of distributed compression at multiple nodes and recovering at the destination node with side

information has been an open problem even after three decades. However, it has been pointed out that Wyner-Ziv coding can achieve any point in the Berger-Tung rate region [23]. In [19], Gastpar has established an achievable rate region for this problem, which can be regarded as a direct extension of the Burger-Tung coding and so far is the tightest lower bound obtained to the best of authors' knowledge. For this reason, we choose Wyner-Ziv coding in this paper. Apart from the analysis in [20]-[22], we derive the achievable rate of such networks in the Rayleigh fading channel under half-duplex mode and notice that the achievable rate depends not only on the channel conditions but, more importantly, also on the design of the distributed compression scheme. Hence, the essence of this work is to optimize the distributed compression scheme to maximize the achievable rate. Moreover, in this work, different from [21], the relay-destination links are not assumed ideal, but they interfere with each other forming a multiple access channel.

The contributions of this work can be summarized as follows:

- The system model of the multi-relay cooperative CF network is presented and its achievable rate is derived by modelling the compression noise as Gaussian noise added to the reconstructed relays' observations, subject to certain constraints imposed by the multiple access channel between the relays and the destination;

- It is shown that the achievable rate can be maximized by jointly optimizing the design of distributed compression at multiple relays and an achievable rate optimization problem is formulated upon this observation. Since this optimization problem is not in the standard concave form, we resort to its Lagrangian dual and prove that the duality gap between the primal and the dual problems is zero. Hence, a solution of the dual problem is also a solution of the primal problem. An iterative algorithm is then devised to search for the solution of the dual problem;

- The analysis is further extended to a hybrid scheme, where the relays can dynamically switch between DF and CF depending on their decoding status. This hybrid scheme is more flexible and its achievable rate is derived and optimized. The information-theoretic analysis of such scheme has been conducted in [20] and [24], where in the latter work a unified

relay framework employing nested blocks combined with backward decoding is investigated. Different from these works, the contribution of this paper is focused on the optimization of distributed compression for those relays employing CF.

The rest of the paper is organized as follows: In the next section, the system model is presented and the achievable rate is derived. The optimal design of distributed compression to maximize the achievable rate is given in Section III. The analysis is further extended to the hybrid CF/DF case in Section IV. Simulation results are presented from both spectrum and energy efficiency perspectives in Section V. The final section concludes the paper.

In this paper, we use capital letters, e.g., X, for random variables and lower case letters, e.g., x, to represent the realization of the variables. Vectors are denoted by bold letters, e.g., $\boldsymbol{X}$, and a superscript, e.g., $\boldsymbol{X}^n$, represents a vector with n elements $(X_1,\ldots,X_n)$. Inequalities of vectors are element-wise, i.e., $\boldsymbol{X} > \boldsymbol{Y}$ means $X_i > Y_i$, $\forall i$. Calligraphic letters are used to denote sets, e.g., $\mathcal{L} \equiv \{1,\ldots,L\}$, and the cardinality of a set $\mathcal{L}$ is $|\mathcal{L}|$. $\mathcal{S}$ is a subset of $\mathcal{L}$ and has a complementary set $\mathcal{S}^C$ where $\mathcal{S} \cup \mathcal{S}^C = \mathcal{L}$ and $\mathcal{S} \cap \mathcal{S}^C = \emptyset$. A set denoted as $\mathcal{S} \backslash i$ stands for a subset of $\mathcal{S}$, where the ith element is removed, and $x_{\mathcal{S}}$ denotes a set $\{x_1,\ldots,x_i,\ldots,x_{|\mathcal{S}|}\}$, $\forall i \in \mathcal{S}$. Mutual information and entropy are denoted as $I\,(\cdot)$ and $H\,(\cdot)$, respectively. All the logarithms are in base 2 and $\mathbf{E}\{\cdot\}$ denotes expectation.

II SYSTEM MODEL

Consider the multi-relay channel with L relays in Fig 1, where h_{si} and h_{id} denote the channel between the source and the ith relay and the channel between the ith relay and the destination, respectively, and h_{sd} stands for the source-destination channel. These channels are assumed to be quasi-static Rayleigh fading channels, i.e., they keep constant within one frame and change from frame to frame. The global channel state information (CSI) is assumed to be available at the destination. In a general multi-relay cooperative network, the relays can transmit and receive at the same time so that each relay will hear from the other relays. However, in this work, half-

duplex mode with duplex ratio α is assumed, where the relays either transmit or receive at one time, thus they do not hear from each other. The duration of one complete frame is T_f, which is divided into two phases. During the first phase, the source broadcasts to the relays and the destination. On the contrary, the source is silent and all the relays transmit simultaneously to the destination in phase 2. It should be noted that the source can also transmit during phase 2. In such a case, a successive cancellation method should be used at the destination to retrieve the whole source message. In this paper, we assume that the source is silent during phase 2 for simplicity.
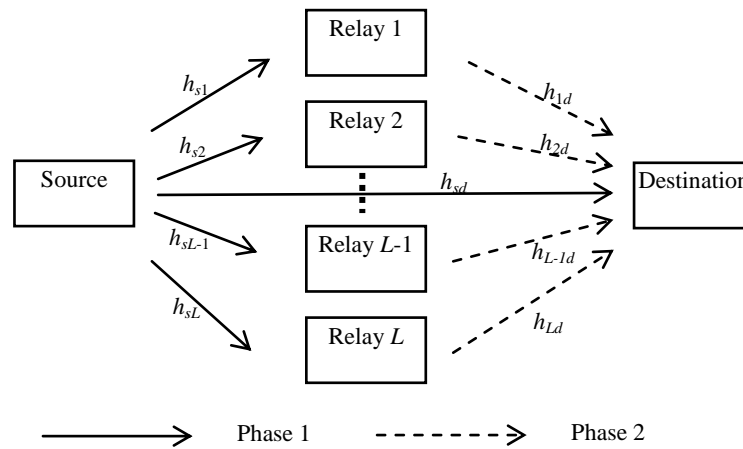

Figure. 1 Muti-Relay System

A message w is encoded and transmitted to the destination. We define a codebook $X_s^{\alpha n}$ with αn elements independent and identically distributed (*i.i.d.*) according to circularly symmetric complex Gaussian (CSCG) distribution with power P_s and zero mean. During phase 1, the source broadcasts $X_s^{\alpha n}$ and the destination and relays receive, respectively,

$$
\begin{aligned}
y_d^{(1)}\left[j\right] &= h_{sd}x_s\left[j\right]+n_d\left[j\right], \\
y_{ri}\left[j\right] &= h_{si}x_s\left[j\right]+n_i\left[j\right],
\end{aligned}
\quad i \in \mathcal{L}, \text{ and } 1 \le j \le \alpha n,
\tag{1}
$$

where n_d and n_i are additive noises at the destination and the *i*th relay, respectively, following CSCG distribution with zero mean and variance σ_d^2 and σ_i^2, respectively, and the superscript denotes phase. These noises are independent from each other and we assume that $\sigma_r^2 = \sigma_i^2$, $\forall i \in \mathcal{L}$.

At the end of phase 1, the *i*th relay performs Wyner-Ziv coding and compresses the received

signal into a bin index v_i [18]-[19]. A codeword $X_i^{(1-\alpha)n}$ is then defined for v_i at the ith relay with $(1-\alpha n)$ *i.i.d.* elements following CSCG distribution with power P_r and zero mean. During the relay-transmit phase, the ith relay transmits $X_i^{(1-\alpha)n}$ and the received signal at the destination is given as

$$y_d^{(2)}[j] = \sum_{i \in \mathcal{L}} h_{id} x_i[j] + n_d[j], \ \alpha n < j \le n. \tag{2}$$

At the destination, the compression indices are obtained first and then with the help of the side information, i.e., $y_d^{(1)}$, the relays' observations are reconstructed by exploiting the signal correlation and combined with $y_d^{(1)}$ to decode the message w.

The achievable rate of the multi-relay CF cooperative network is given by the following theorem.

Theorem 1: For the multi-relay cooperative network using Wyner-Ziv compression, the achievable rate is

$$R_{CF} = \alpha I\left(X_s; \hat{Y}_{r\mathcal{L}}, Y_d^{(1)}\right), \tag{3}$$

subject to constraints

$$\alpha I\left(\hat{Y}_{r\mathcal{S}}; Y_{r\mathcal{S}} \mid \hat{Y}_{r\mathcal{S}^C}, Y_d^{(1)}\right) \le (1-\alpha) I\left(X_{r\mathcal{S}}; Y_d^{(2)} \mid X_{r\mathcal{S}^C}\right), \forall \ \mathcal{S} \subseteq \mathcal{L},$$

where $\hat{Y}_{r\mathcal{L}}$ is a set of auxiliary random variables and can be interpreted as the estimated observations of the relays.

Proof: The theorem can be proved by extending [20, theorem 3] from full-duplex to half-duplex. The nonsingle letter bound on the rate R_{CF} is given as

$$nR_{CF} = \sum_{j=1}^{n} I\left(X_s[j]; \hat{Y}_{r\mathcal{L}}[j], Y_d[j]\right), \tag{4}$$

where

$$\sum_{j=1}^{n} I\left(\hat{Y}_{r\mathcal{S}}[j]; Y_{r\mathcal{S}}[j] \mid \hat{Y}_{r\mathcal{S}^C}[j], Y_d[j]\right) \le \sum_{j=1}^{n} I\left(X_{r\mathcal{S}}[j]; Y_d[j] \mid X_{r\mathcal{S}^C}[j]\right), \forall \ \mathcal{S} \subseteq \mathcal{L}.$$

$X_{r\mathcal{S}}$ and $Y_{r\mathcal{S}}$ denote the transmitted and received signals of all relays that belong to $\mathcal{S}$ and $\hat{Y}_{ri}$ is an auxiliary random variable and can be interpreted as the estimated relay observation Y_{ri}. During the first phase, all relays are silent so that $X_{ri}[j]=0$. On the other hand, in phase 2 the source is

silent and the relays only transmit so that $X_s[j]$, $Y_{ri}[j]$ and $\hat{Y}_{ri}[j]$ are 0. Given that the channel is memoryless, (4) can be rewritten as, ignoring the index j,

$$nR_{CF} = \alpha n I\left(X_s; \hat{Y}_{r\mathcal{L}}, Y_d^{(1)}\right),$$ (5)

subject to

$$\alpha n I\left(\hat{Y}_{r\mathcal{S}}; Y_{r\mathcal{S}} \mid \hat{Y}_{r\mathcal{S}^C}, Y_d^{(1)}\right) \le (1-\alpha) n I\left(X_{r\mathcal{S}}; Y_d^{(2)} \mid X_{r\mathcal{S}^C}\right), \forall \, \mathcal{S} \subseteq \mathcal{L}. \blacksquare$$

From the constraint equations of (5), it should be noted that not only the correlation between the relays' observations and the source information $Y_d^{(1)}$ but also the correlation between the relays' observations themselves should be exploited. This is reflected by the fact that the mutual information on the left side of the constraint equations is conditioned on both $Y_d^{(1)}$ and $\hat{Y}_{r\mathcal{S}^C}$, i.e., the reconstructed observations of other relays.

With CSCG distributed codebooks, using *Theorem* 1 yields the following results.

Proposition 1: With CSCG distributed codebooks, the multi-relay cooperative network using Wyner-Ziv compression achieves the rate up to

$$R_{CF} = \alpha \log\left(1 + \frac{\gamma_0 P_s}{\sigma_d^2} + \sum_{i \in \mathcal{L}} \frac{\gamma_i P_s}{\sigma_r^2 + \sigma_{wi}^2}\right),$$ (6)

subject to multiple constraints

$$f_{\mathcal{S}}\left(\sigma_{w\mathcal{L}}^2\right) \le \frac{1-\alpha}{\alpha} \log\left(1 + \frac{\sum_{i \in \mathcal{S}} \gamma_{id} P_r}{\sigma_d^2}\right), \forall \, \mathcal{S} \subseteq \mathcal{L}, \text{ where}$$

$$\begin{aligned}
f_{\mathcal{S}}\left(\sigma_{w\mathcal{L}}^2\right) &= \log\left(\left(\gamma_0 P_s + \sigma_d^2\right) \prod_{i \in \mathcal{L}} \left(\sigma_r^2 + \sigma_{wi}^2\right) + \sum_{i \in \mathcal{L}} \left(\gamma_i P_s \sigma_d^2 \prod_{k \in \mathcal{L}, k \ne i} \left(\sigma_r^2 + \sigma_{wk}^2\right)\right)\right) - \log\left(\prod_{i \in \mathcal{S}} \sigma_{wi}^2\right) \\
&\quad - \log\left(\left(\gamma_0 P_s + \sigma_d^2\right) \prod_{i \in \mathcal{S}^C} \left(\sigma_r^2 + \sigma_{wi}^2\right) + \sum_{i \in \mathcal{S}^C} \left(\gamma_i P_s \sigma_d^2 \prod_{k \in \mathcal{S}^C, k \ne i} \left(\sigma_r^2 + \sigma_{wk}^2\right)\right)\right)
\end{aligned}.$$

Here σ_{wi}^2 is the variance of the compression noise at the ith relay, $\sigma_{w\mathcal{L}}^2$ is a set consisting of σ_{wi}^2, $\forall \, i \in \mathcal{L}$, and $\gamma_0 = |h_{sd}|^2$, $\gamma_i = |h_{si}|^2$ and $\gamma_{id} = |h_{id}|^2$.

Proof: See Appendix A for the proof. ∎

III. DISTRIBUTED COMPRESSION OPTIMIZATION

As shown in (6), the achievable rate R_{CF} depends on the variances of the compression noise and thus the optimal distributed compression should optimize σ_{wi}^2 jointly to maximize R_{CF} subject to the identified constraint functions. The number of constraint functions, denoted as c, depends on the number of relays and for L relays, is given as

$$c = \sum_{i \in \mathcal{L}} \binom{L}{i} = \sum_{i \in \mathcal{L}} c_{L,i} \ . \tag{7}$$

It is worth noting that in addition to the compression noise, eq. (6) should also be maximized with respect to the duplex ratio α. However, it is very difficult to optimize α and $\sigma_{w\mathcal{L}}^2$ at the same time. Hence we maximize the achievable rate with a given α first and then optimize α numerically. The optimization problem can be formulated as, with fixed α,

$$\underset{\sigma_{w\mathcal{L}}^2}{\text{maximize}} \quad \Phi\left(\sigma_{w\mathcal{L}}^2\right) = \log\left(1 + \frac{\gamma_0 P_s}{\sigma_d^2} + \sum_{i \in \mathcal{L}} \frac{\gamma_i P_s}{\sigma_r^2 + \sigma_{wi}^2}\right), \quad \forall \ \mathcal{S}^l \tag{8}$$

$$\text{subject to} \quad f_{\mathcal{S}^l}\left(\sigma_{w\mathcal{L}}^2\right) \le C_{\mathcal{S}^l}, \ \sigma_{wi}^2 > 0, \ i \in \mathcal{L}, \ 1 \le l \le L$$

where

$$C_{\mathcal{S}^l} = \frac{1-\alpha}{\alpha} \log\left(1 + \frac{\sum_{i \in \mathcal{S}^l} \gamma_{id} P_r}{\sigma_d^2}\right), \quad \forall \mathcal{S}^l \subseteq \mathcal{L}, \text{ and } \left|\mathcal{S}^l\right| = l, \ 1 \le l \le L.$$

For each l, there are $c_{L,l}$ different $\mathcal{S}^l$, each corresponding to one constraint function. Since the maximization objective function is not in the standard concave form, we resort to its dual problem by formulating its Lagrangian dual as

$$L\left(\sigma_{w\mathcal{L}}^2, \boldsymbol{\lambda}^c\right) = \Phi\left(\sigma_{w\mathcal{L}}^2\right) + \boldsymbol{\lambda}^{cT}\left(\mathbf{C}^c - \mathbf{f}^c\left(\sigma_{w\mathcal{L}}^2\right)\right), \tag{9}$$

with

$$\boldsymbol{\lambda}^c = \left(\boldsymbol{\lambda}_{\mathcal{S}^1},...,\boldsymbol{\lambda}_{\mathcal{S}^L}\right)^T, \ \boldsymbol{\lambda}_{\mathcal{S}^l} = \left(\lambda_{\mathcal{S}_1^l},...,\lambda_{\mathcal{S}_{c_{L,l}}^l}\right), \ \mathbf{C}^c = \left(\mathbf{C}_{\mathcal{S}^1},...,\mathbf{C}_{\mathcal{S}^L}\right)^T, \ \mathbf{C}_{\mathcal{S}^l} = \left(C_{\mathcal{S}_1^l},...,C_{\mathcal{S}_{c_{L,l}}^l}\right),$$

$$\mathbf{f}^c\left(\sigma_{w\mathcal{L}}^2\right) = \left(\mathbf{f}_{\mathcal{S}^1}\left(\sigma_{w\mathcal{L}}^2\right),...,\mathbf{f}_{\mathcal{S}^1}\left(\sigma_{w\mathcal{L}}^2\right)\right)^T, \mathbf{f}_{\mathcal{S}^1}\left(\sigma_{w\mathcal{L}}^2\right) = \left(f_{\mathcal{S}_1^l}\left(\sigma_{w\mathcal{L}}^2\right),...,f_{\mathcal{S}_{c_{L,l}}^l}\left(\sigma_{w\mathcal{L}}^2\right)\right),$$

where $\mathcal{S}_j^l \subseteq \mathcal{L}$, and $\left|\mathcal{S}_j^l\right| = l$ for $1 \leq j \leq c_{L,l}$, $1 \leq l \leq L$. The dual function is then defined as a maximization function of (9)

$$\varphi(\boldsymbol{\lambda}) = \max_{\sigma_{w\mathcal{L}}^2} L\left(\sigma_{w\mathcal{L}}^2, \boldsymbol{\lambda}^c\right). \tag{10}$$

The dual problem takes the following form:

$$\begin{aligned} \underset{\boldsymbol{\lambda}^c}{\text{minimize}} \quad & \varphi\left(\boldsymbol{\lambda}^c\right) \\ \text{subject to} \quad & \boldsymbol{\lambda}^c \geq \mathbf{0}^c \end{aligned}. \tag{11}$$

The dual objective function $\varphi(\boldsymbol{\lambda}^c)$ is a convex function regardless of the concavity of the primal function $\Phi\left(\sigma_{w\mathcal{L}}^2\right)$ [25]. If we can prove that the duality gap between the primal problem (8) and the dual problem (11) is zero, we can solve the primal problem by resorting to the dual problem because they have the same solution.

Theorem 2: The duality gap of the primal problem (8) and dual problem (11) is zero.

Proof: The theorem is proved in Appendix B. ∎

With *Theorem* 2, the primal problem can be solved by searching for the solution of the dual problem. At first, we need to find the optimal $\sigma_{w\mathcal{L}}^2$ to maximize (9). Due to its high complexity, it is difficult to obtain a closed-form solution. A block-coordinate descent optimization algorithm is applied, where the objective function is optimized with respect to σ_{wi}^2 only while keeping other $\sigma_{w\mathcal{L}/i}^2$ unchanged [26]. This iterative optimization algorithm is conducted with respect to σ_{wi}^2 from $i=1$ to $i=L$ successively in one iteration and the same procedure is repeated until $\sigma_{w\mathcal{L}}^2$ converges.

We define a set $\mathcal{C}$ as $\left\{ \mathcal{S}_1^1, ..., \mathcal{S}_L^1, ..., \mathcal{S}_1^l, ..., \mathcal{S}_{c_{L,l}}^l, ..., \mathcal{S}_1^L \right\}$, where each element is a subset of $\mathcal{L}$. Then

a subset of $\mathcal{C}$ is defined as $\mathcal{C}_i$ whose elements are those subsets containing i. It can be easily

derived from (7) that

$$|\mathcal{C}_i| = (c+1)/2, \ |\mathcal{C}_i{}^C| = (c-1)/2.$$

It is proved in Appendix C that the maximization in (10) with respect to σ_{wi}^2 only is equivalent

to maximizing the following function:

$$\hat{\varphi}(\boldsymbol{\lambda}^c) = \max_{\underline{\sigma}_{wi}^2 > 0} \left\{ (1 - \lambda_\Sigma) \log\left(1 + \underline{\sigma}_{wi}^2 + A_i\right) + \sum_{\mathcal{S}_j^l \in \mathcal{C}_i^C} \left(\lambda_{\mathcal{S}_j^l} \log\left(1 + \underline{\sigma}_{wi}^2 + B_{\mathcal{S}_j^l i}\right) \right) + \lambda_{\mathcal{C}_i} \log\left(\underline{\sigma}_{wi}^2\right) - \log\left(1 + \underline{\sigma}_{wi}^2\right) \right\}$$

,(12)

where A_i and $B_{\mathcal{S}_j^l i}$ are given in Appendix C and

$$\underline{\sigma}_{wi}^2 = \sigma_{wi}^2 / \sigma_r^2, \ \lambda_\Sigma = \sum_{l \in \mathcal{L}} \sum_{j=1}^{c_{L,l}} \lambda_{\mathcal{S}_j^l} \ \lambda_{\mathcal{C}_i} = \sum_{\mathcal{S}_j^l \in \mathcal{C}_i} \lambda_{\mathcal{S}_j^l}.$$

Maximizing (12) can be solved by resorting to its derivatives in the following proposition.

Proposition 2: The optimal $\underline{\sigma}_{wi}^{2\ *}$ is chosen from a set $\mathcal{X}_{sub}$ containing all the positive roots of a

$(c+1)/2$ degree polynomial $q\left(\boldsymbol{\lambda}^c, \underline{\sigma}_{w,i}^2\right)$ given in Appendix D, expressed as

$$\underline{\sigma}_{wi}^{2\ *} = \arg\max_{\underline{\sigma}_{wi}^{2\ *} \in \mathcal{X}_{sub}} \theta\left(\boldsymbol{\lambda}^c, \underline{\sigma}_{wi}^2\right) \tag{13}$$

where $\theta\left(\boldsymbol{\lambda}^c, \underline{\sigma}_{wi}^2\right)$ is defined in Appendix D.

Proof: See Appendix D for proof. ∎

For $L=2$, the closed-form optimal $\underline{\sigma}_{wi}^{2\ *}$ can be derived and is given in Appendix D.

Once $\underline{\sigma}_{w\mathcal{L}}^{2\ *}$ is obtained, the dual minimization problem (11) can be solved by successively

optimizing elements of $\boldsymbol{\lambda}^c$. However, the range of $\boldsymbol{\lambda}^c$ is too large to be feasible. The following

proposition defines the feasible searching range of $\boldsymbol{\lambda}^c$ as $\mathbf{0}^c \leq \boldsymbol{\lambda}^c < \mathbf{1}^c$.

Proposition 3: If $\boldsymbol{\lambda}^c \geq \mathbf{1}^c$, $\hat{\varphi}(\boldsymbol{\lambda}^c) \leq 0$ and its maximum is achieved with $\underline{\sigma}_{wi}^2$ approaching $+\infty$, i.e., the maximum of $\hat{\varphi}(\boldsymbol{\lambda}^c)$ is achieved with finite $\underline{\sigma}_{wi}^2$ only when $\boldsymbol{\lambda}^c < \mathbf{1}^c$.

Proof: See Appendix E for proof. ∎

Due to the convexity of $\varphi(\boldsymbol{\lambda}^c)$, the subgradient method can be used for solving minimization problem (11) [25]. The searching direction of $\lambda_{\mathcal{S}_j^l}$ is given as

$$g_{\mathcal{S}_j^l}\left(\sigma_{w\mathcal{L}}^2{}^*\right) = \frac{\partial \varphi(\boldsymbol{\lambda}^c)}{\partial \lambda_{\mathcal{S}_j^l}} = C_{\mathcal{S}_j^l} - f_{\mathcal{S}_j^l}\left(\sigma_{w\mathcal{L}}^2{}^*\right). \tag{14}$$

The searching criterion is: if $g_{\mathcal{S}_j^l}\left(\sigma_{w\mathcal{L}}^2{}^*\right) \leq 0$, increase $\lambda_{\mathcal{S}_j^l}$; otherwise decrease $\lambda_{\mathcal{S}_j^l}$. The overall algorithm is given as

Step 1: Initialize $\boldsymbol{\lambda}^c{}_{\min} = \mathbf{0}^c$ and $\boldsymbol{\lambda}^c{}_{\max} = \mathbf{1}^c$;

Step 2: Let $\boldsymbol{\lambda}^c = (\boldsymbol{\lambda}^c{}_{\min} + \boldsymbol{\lambda}^c{}_{\max})/2$;

Step 3: Let $t=1$, initialize[1] $\underline{\sigma}_{wi}^{2\,(t)} = +\infty$ from $i=1$ to L;

Step 4: From $i=1$ to L, update $\underline{\sigma}_{wi}^{2\,(t+1)}$ based on (13);

Step 5: If $\sum_{i \in \mathcal{L}} \left| \underline{\sigma}_{wi}^{2\,(t+1)} - \underline{\sigma}_{wi}^{2\,(t)} \right| \leq \varepsilon_\sigma$, $\sigma_{w\mathcal{L}}^2{}^* = \sigma_{w\mathcal{L}}^{2\,(t+1)}$ and go to the next step; otherwise, $t=t+1$ and go to Step 4.

Step 6: For $1 \leq j \leq c$, if $g_j\left(\sigma_{w\mathcal{L}}^2{}^*\right) \leq 0$, $\lambda_{j,\min} = \lambda_j$; otherwise $\lambda_{j,\max} = \lambda_j$;

Step 7: If $\sum_{j=1}^c \left| \lambda_{j,\max} - \lambda_{j,\min} \right| \leq \varepsilon_\lambda$, $\boldsymbol{\lambda}^c$ converges to $(\boldsymbol{\lambda}^c{}_{\min} + \boldsymbol{\lambda}^c{}_{\max})/2$ and the algorithm is finished; otherwise, go back to Step 2.

IV. THE HYBRID CF/DF SCHEME

In the previous section, we study the solution where each relay compresses its own observation independently using Wyner-Ziv coding. However, some of the relays might be able to decode the

---

[1] Ideally, the initial value of $\underline{\sigma}_{wi}^2$ to start the iteration should be $+\infty$. Practically we choose a large enough value 10^{10} as the initial value. ε_σ, ε_λ and ε are very small values.

message successfully and help the final decoding at the destination by re-encoding and forwarding the message. By allowing the relays to employ DF more flexibility can be provided. At the destination, the compression indices from CF relays are decoded first and then cancelled out from the received signal. The observations of the relays are reconstructed and then coherently combined with the residue signal to jointly decode the message from the source. In this section, we will consider this more flexible hybrid scheme, where each relay chooses DF or distributed compression depending on its own decoding status, i.e., if the ith relay successfully decodes the message, it applies DF; otherwise, CF is employed.

If we define a subset of $\mathcal{L}$ as $\mathcal{D}$, and assume that the ith relay conducts DF as long as $i \in \mathcal{D}$. The achievable rate of such a case is given by the following theorem.

Theorem 3: For a multi-relay cooperative network with L relays, if a subset of relays, denoted as $\mathcal{D}$, employs DF and the rest of the relays perform distributed compression, the achievable rate is up to

$$R_{CDF} = \min\left\{ \min_{i \in \mathcal{D}} \left\{ \alpha I\left(X_s; Y_{ri}\right) \right\}, \alpha I\left(X_s; \hat{Y}_{r\mathcal{D}^C}, Y_d^{(1)}\right) + (1-\alpha) I\left(X_{r\mathcal{D}}; Y_d^{(2)} \mid X_{r\mathcal{D}^C}\right) \right\}, \quad (15)$$

subject to

$$\alpha I\left(\hat{Y}_{r\mathcal{S}}; Y_{r\mathcal{S}} \mid \hat{Y}_{r\mathcal{S}^C}, Y_d^{(1)}\right) \le (1-\alpha) I\left(X_{r\mathcal{S}}; Y_d^{(2)} \mid X_{r\mathcal{S}^C}\right), \; \forall \, \mathcal{S} \subseteq \mathcal{D}^C,$$

where $\mathcal{S} \cup \mathcal{S}^C = \mathcal{D}^C$.

Proof: Same as *Theorem* 1, this theorem is a straightforward extension of [20, *Theorem* 4]. The nonsingle letter bound is

$$nR_{CDF} = \sum_{j=1}^{n} \min\left\{ \min_{i \in \mathcal{D}} \left\{ I\left(X_s[j]; Y_{ri}[j]\right) \right\}, I\left(X_s[j], X_{r\mathcal{D}}[j]; \hat{Y}_{r\mathcal{D}^C}[j], Y_d[j] \mid X_{r\mathcal{D}^C}[j]\right) \right\}, \quad (16)$$

subject to

$$\sum_{j=1}^{n} I\left(\hat{Y}_{r\mathcal{S}}[j]; Y_{r\mathcal{S}}[j] \mid \hat{Y}_{r\mathcal{S}^C}[j], Y_d[j]\right) \le \sum_{j=1}^{n} I\left(X_{r\mathcal{S}}[j]; Y_d[j] \mid X_{r\mathcal{S}^C}[j]\right), \; \forall \, \mathcal{S} \subseteq \mathcal{D}^C.$$

Due to the half duplex operation, we have

$$nR_{CDF} = \min\left\{\min_{i\in\mathcal{D}}\left\{\alpha n I\left(X_s;Y_{ri}\right)\right\}, \alpha n I\left(X_s;\hat{Y}_{r\mathcal{D}^C},Y_d^{(1)}\right)+\left(1-\alpha\right)n I\left(X_{r\mathcal{D}};Y_d^{(2)}\mid X_{r\mathcal{D}^C}\right)\right\}, \quad (17)$$

subject to

$$\alpha n I\left(\hat{Y}_{r\mathcal{S}};Y_{r\mathcal{S}}\mid \hat{Y}_{r\mathcal{S}^C},Y_d^{(1)}\right)\le\left(1-\alpha\right)n I\left(X_{r\mathcal{S}};Y_d^{(2)}\mid X_{r\mathcal{S}^C}\right), \forall\ \mathcal{S}\subseteq\mathcal{D}^C.\ \blacksquare$$

When decoding the compression indices, the messages from the relays belonging to subset $\mathcal{D}$ are treated as noise. With CSCG distributed codebooks, it is proved in Appendix A that

$$R_{CDF\_\mathcal{D}} = \min\left\{\min_{i\in\mathcal{D}}\left\{\alpha\log\left(1+\frac{\gamma_i P_s}{\sigma_r^2}\right)\right\}, \alpha\log\left(1+\frac{\gamma_0 P_s}{\sigma_d^2}+\sum_{i\in\mathcal{D}^C}\frac{\gamma_i P_s}{\sigma_r^2+\sigma_{wi}^2}\right)+\left(1-\alpha\right)\log\left(1+\sum_{i\in\mathcal{D}}\frac{\gamma_{id} P_r}{\sigma_d^2}\right)\right\}$$

$$(18)$$

subject to constraints

$$f_{\mathcal{S}}\left(\sigma_{w\mathcal{D}^C}^2\right)\le\frac{1-\alpha}{\alpha}\log\left(1+\frac{\sum_{i\in\mathcal{S}}\gamma_{id} P_r}{\sum_{i\in\mathcal{D}}\gamma_{id} P_r+\sigma_d^2}\right), \forall\ \mathcal{S}\subseteq\mathcal{D}^C,$$

where

$$f_{\mathcal{S}}\left(\sigma_{w\mathcal{D}^C}^2\right) = \log\left(\left(\gamma_0 P_s+\sigma_d^2\right)\prod_{i\in\mathcal{D}^C}\left(\sigma_r^2+\sigma_{wi}^2\right)+\sum_{i\in\mathcal{D}^C}\left(\gamma_i P_s\sigma_d^2\prod_{k\in\mathcal{D}^C,k\ne i}\left(\sigma_r^2+\sigma_{wk}^2\right)\right)\right)-\log\left(\prod_{i\in\mathcal{S}}\sigma_{wi}^2\right)$$

$$-\log\left(\left(\gamma_0 P_s+\sigma_d^2\right)\prod_{i\in\mathcal{S}^C}\left(\sigma_r^2+\sigma_{wi}^2\right)+\sum_{i\in\mathcal{S}^C}\left(\gamma_i P_s\sigma_d^2\prod_{k\in\mathcal{S}^C,k\ne i}\left(\sigma_r^2+\sigma_{wk}^2\right)\right)\right)$$

.

The DF subset $\mathcal{D}$ is not static but dynamically depends on the decoding status of each individual relay. The dynamic behaviour of set $\mathcal{D}$ is defined in the way that the relays that could decode the source message are included in $\mathcal{D}$ and the relays themselves decide if they are part of $\mathcal{D}$ or not. The achievable rate of the dynamic hybrid CF/DF scheme should be maximized with respect to the DF subset as

$$R_{CDF} = \max_{\mathcal{D}}\left\{R_{CDF\_\mathcal{D}}\right\} \quad (19)$$

subject to the same constraint in (18). In (17), the CF part $\alpha I\left(X_s; \hat{Y}_{r_D c}, Y_d^{(1)}\right)$ can still be maximized with optimized distributed compression and the same optimization algorithm can be used for the hybrid CF/DF scheme.

VI NUMERICAL RESULTS

In this section, we evaluate the performance of the proposed non-hybrid CF and hybrid CF/DF schemes and compare their performance against the existing ones. In addition to the spectrum efficiency analysis, energy efficiency results based on practical power models are also presented to provide insights into optimal relay operation & deployment in realistic communication networks. Unless mentioned, the duplex ratio is optimized numerically.

A. Spectrum Efficiency Analysis

Fig 2 shows the outage probability for SNR setting $\mathbf{E}\{P_s\gamma_0/\sigma_d^2\}=\mathbf{E}\{P_s\gamma_i/\sigma_r^2\}=0$dB and $\mathbf{E}\{P_r\gamma_{id}/\sigma_d^2\}=30$dB with the duplex ratio $\alpha=0.9$. It is shown that at any target rate, the outage probability of different approaches follow the same order: hybrid CF/DF < CF 3 < CF 1 < direct transmission, which means that CF outperforms direct transmission and the performance can be further improved by the hybrid CF/DF scheme. However, the gain diminishes as the target rate increases. Fig. 3 plots the rates of the following forwarding schemes at peak values (outage probability = 0.95) for 3 relays:

- Jointly optimized CF: the received signals are independently compressed at the relays but the compression is jointly optimized using the proposed iterative algorithm;

- Quantization and forward (QF) [27]: the relays' receptions are quantized only before being forwarded and no signal correlation is exploited;

- Independent CF: the compression at each relay is optimized independently;

- Upper limit: the upper bound when the variances of the compression noise approach zero in (6).
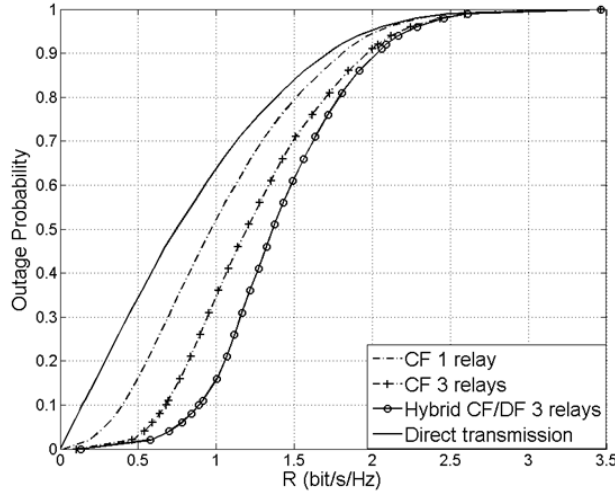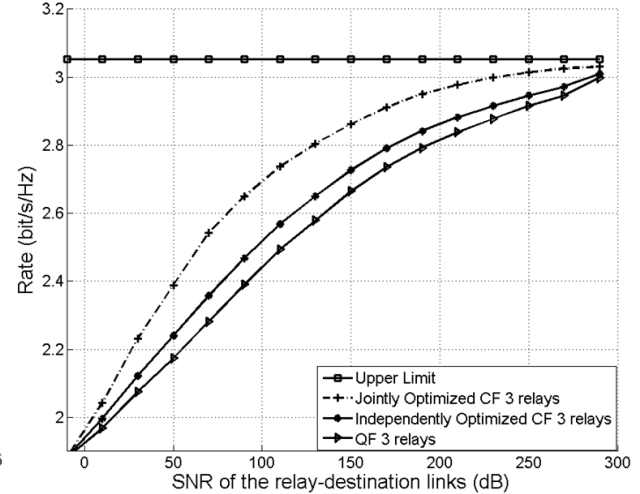
Figure 2 Outage probability
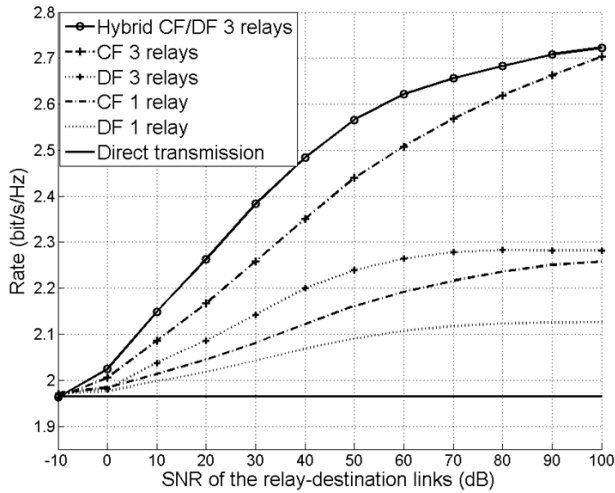

Figure 3 Rates of CF and QF
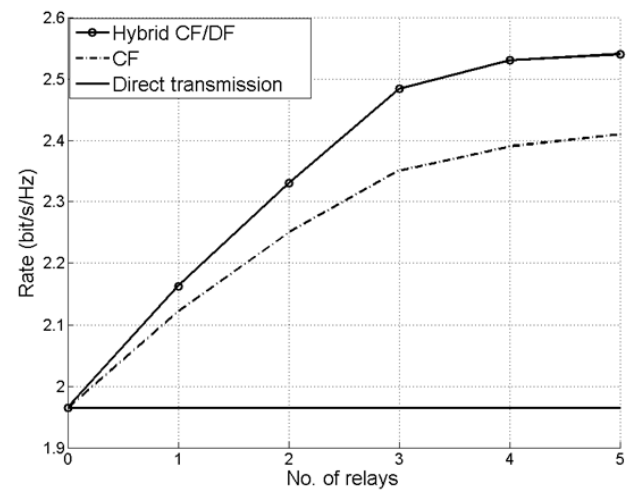

Figure 4 Rates of DF, CF and Hybrid CF/DF

Figure 5 Rates with different number of relays

We notice that all schemes are able to approach the upper limit when the relay-destination links become very strong. The independent compression shows around 10dB gain against QF and this gain is more significant when the proposed joint optimization algorithm is employed, e.g., at a given rate 2.6bit/s/Hz, more than 40dB additional gain is achieved.

Fig. 4 compares the rates of different forwarding schemes with SNR setting $\mathbf{E}\{P_s\gamma_0/\sigma_d^2\}$ =$\mathbf{E}\{P_s\gamma_i/\sigma_r^2\}$=0dB. The DF scheme is selective DF, where the relay only re-encodes and forwards the received message when the message is successfully decoded; otherwise, it remains silent. We notice that for selective DF, decoding at the relays tends to be erroneous at the low SNR region. Thus, in most of the cases, the relays keep silent during phase 2. If we let α

approach 1, the achievable rate of selective DF can approach direct transmission. For CF scheme, we can also let α approach 1 thus the compression noises approach infinity and the achievable rate of CF can approach direct transmission as well at the low SNR region. For hybrid CF/DF, at the low SNR region, the relays cannot decode the source message and they employ CF in most of the cases. Similarly, if we let α approach 1, the achievable rate of hybrid CF/DF will approach direct transmission. In short, at the low SNR region, all schemes including selective DF, pure CF and hybrid CF/DF approach the performance of direct transmission. However, when the relay-destination links are of relatively good quality, the rate of DF saturates and the gains of CF and hybrid CF/DF become more remarkable.

The gap between the hybrid scheme and the non-hybrid CF scheme becomes larger initially but eventually diminishes because CF is always the best choice with very strong relay-destination links. Fig. 5 plots the rates with different number of relays for SNR setting $\mathbf{E}\{P_s\gamma_0/\sigma_d^2\} = \mathbf{E}\{P_s\gamma_i/\sigma_r^2\} = 0$dB and $\mathbf{E}\{P_r\gamma_{id}/\sigma_d^2\} = 40$dB. From 1 to 3 relays, the rate is improved by 11% and 15% for CF and hybrid CF/DF, respectively, but from 3 to 5 relays, the improvement is only 4% for both schemes. It implies that with growing of the cooperative network, the gain of adding new relays becomes only marginal.

B. Energy Efficiency Analysis

Energy efficiency has become more and more important to support economic and sustainable future cellular networks. The energy efficiency analysis of relay systems has been investigated in [28]-[34]. In this paper, energy efficiency is evaluated in terms of consumed energy per transmitted bit. We consider a cellular network, assuming hexagon cells with radius r_c. The relays are assumed to be deployed at the cell edge. The base station (BS) to relay and user equipment (UE) distances are assumed to be 0.95 r_c and 0.99 r_c, respectively. Each cell consists of 3 sectors operating on orthogonal frequency bands. Hence, the interference is only from neighbouring cells. The mean antenna gains of the transmit and receive antennas are assumed to be 14dB and 0dB and the pathloss model is given as
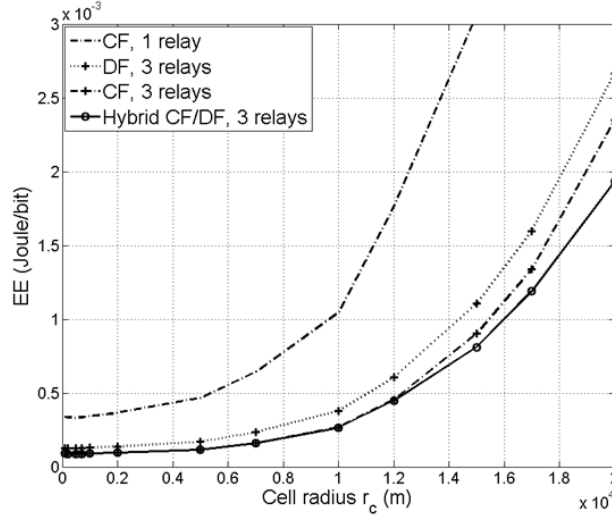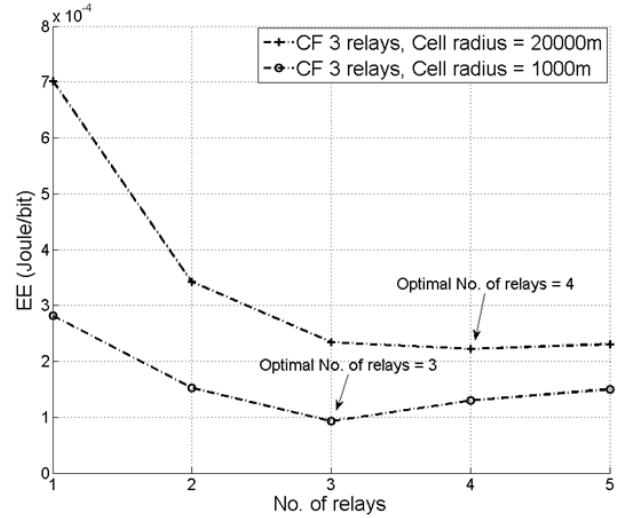
Figure 6 Energy efficiency (Joule/bit) Figure 7 Energy efficiency with different number of relays

$$PL(d) = 15.3 + 37.6*\log_{10}(d),$$

where d denotes distance [35]. The thermal noise power density is $N_0 = -171\text{dBm/Hz}$ and the bandwidth B is assumed to be 10MHz.

The total energy consumption is composed of not only the transmission power but also the circuitry energy consumption of all involved nodes. During phase 1, the BS broadcasts and the consumed energy can be expressed as $E^{(1)} = \alpha T_f(P_s + P_{s,tc} + LP_{r,rc} + P_{d,rc})$, where $P_{s,tc}$ is the circuitry power consumption of the BS while transmitting, $P_{r,rc}$ and $P_{d,rc}$ are the receiving circuitry power consumption of the relays and the UE, respectively. In the second phase, the relays transmit to the UE simultaneously and the BS is switched to sleep mode. The consumed energy is $E^{(2)} = (1 - \alpha)T_f(LP_r + LP_{r,tc} + P_{d,rc} + P_{s,sm})$, where $P_{r,tc}$ is the circuitry power consumption of a transmitting relay and $P_{s,sm}$ is the sleep mode power consumption of the BS. The overall energy efficiency in terms of Joule/bit can be expressed as $E = (E^{(1)} + E^{(2)})/BRT_f$. Based on the well defined power models of BS, relay and UE in [36]-[42], the power consumption values can be given as: $P_s = 19.95\text{W}$, $P_{s,tc} = 204\text{W}$, $P_r = 5.01\text{W}$, $P_{r,tc} = 23.25\text{W}$, $P_{r,tc} = 14.25\text{W}$, $P_{d,tc} = 1.80\text{W}$, and $P_{s,sm} = 75\text{W}$.

Fig 6 illustrates energy efficiency in terms of Joule/bit for different schemes. We first notice that the CF scheme has less energy consumption per bit than the DF scheme when the relays and the UE are at the cell edge and the hybrid scheme always outperforms non-hybrid ones due to its

capability of adapting to the channel changes. We also notice that the energy efficiency improvement of adding more relays becomes more remarkable with increased cell radius r_c. Fig. 7 depicts energy efficiency of different number of relays for urban deployment (cell size $r_c=1000$m) and rural deployment (cell size $r_c=20000$m), respectively. Since the achieved capacity gain of adding more relays diminishes but the extra energy consumption of additional relays is almost constant, at certain point introducing more relays to the cooperative network will only cause extra energy expenditure with very small rate improvement. Therefore energy efficiency of the cooperative network actually decreases. In other words, there should be an optimal number of relays to be deployed to achieve the best energy efficiency as shown in the figure.

VI CONCLUSIONS

In this paper, we investigate the performance of a multi-relay cooperative network employing distributed compression. The proposed scheme employs Wyner-Ziv compression and it is shown that the application of such a compression scheme gives rise to a compression noise variance optimization problem. This problem is solved by resorting to its dual problem using an iterative optimizing algorithm. The analysis is further extended to a hybrid scheme where the relays can choose from DF and CF adaptively. The proposed scheme is shown to be able to greatly improve the efficiency of compression by exploiting the signal correlation. In addition, significant gains can be obtained from both spectrum and energy efficiency perspectives by applying distributed compression, and further improvement can be achieved by allowing the relays to employ different forwarding schemes adaptively based on their own decoding status.

APPENDIX A

The mutual information in (3) can be expressed as

$$I\left(X_s;\hat{Y}_{r\mathcal{L}},Y_d^{(1)}\right)=H\left(\hat{Y}_{r\mathcal{L}},Y_d^{(1)}\right)-H\left(\hat{Y}_{r\mathcal{L}},Y_d^{(1)}\mid X_s\right), \tag{A1}$$

$$\begin{aligned}I\left(\hat{Y}_{r\mathcal{S}};Y_{r\mathcal{S}}\mid \hat{Y}_{r\mathcal{S}^C},Y_d^{(1)}\right)&=H\left(\hat{Y}_{r\mathcal{S}}\mid \hat{Y}_{r\mathcal{S}^C},Y_d^{(1)}\right)-H\left(\hat{Y}_{r\mathcal{S}}\mid Y_{r\mathcal{S}},\hat{Y}_{r\mathcal{S}^C},Y_d^{(1)}\right)\\&=H\left(\hat{Y}_{r\mathcal{L}},Y_d^{(1)}\right)-H\left(\hat{Y}_{r\mathcal{S}^C},Y_d^{(1)}\right)-H\left(\hat{Y}_{r\mathcal{S}}\mid Y_{r\mathcal{S}}\right)\end{aligned}, \tag{A2}$$

$$I\left(X_{r\mathcal{S}};Y_d^{(2)}\mid X_{r\mathcal{S}^C}\right)=H\left(Y_d^{(2)}\mid X_{r\mathcal{S}^C}\right)-H\left(Y_d^{(2)}\mid X_{r\mathcal{L}}\right), \tag{A3}$$

$\forall\ \mathcal{S}\subseteq\mathcal{L}$. According to [7] and [18], we can choose the auxiliary random variable $\hat{Y}_{ri}$ of the form $\hat{Y}_{ri}=Y_{ri}+W_i$, where W_i is independent of Y_{ri} and Y_d and follows CSCG distribution with variance σ_{wi}^2 and zero mean, referred to as "compression noise". Hence

$$H\left(\hat{Y}_{r\mathcal{L}},Y_d^{(1)}\right)=\log\left((\pi e)^{L+1}\left((\gamma_0 P_s+\sigma_d^2)\prod_{i\in\mathcal{L}}(\sigma_r^2+\sigma_{wi}^2)+\sum_{i\in\mathcal{L}}\left(\gamma_i P_s\sigma_d^2\prod_{k\in\mathcal{L},k\neq i}(\sigma_r^2+\sigma_{wk}^2)\right)\right)\right). \tag{A4}$$

$$H\left(\hat{Y}_{r\mathcal{S}^C},Y_d^{(1)}\right)=\log\left((\pi e)^{|\mathcal{S}^C|+1}\left((\gamma_0 P_s+\sigma_d^2)\prod_{i\in\mathcal{S}^C}(\sigma_r^2+\sigma_{wi}^2)+\sum_{i\in\mathcal{S}^C}\left(\gamma_i P_s\sigma_d^2\prod_{k\in\mathcal{S}^C,k\neq i}(\sigma_r^2+\sigma_{wk}^2)\right)\right)\right), \tag{A5}$$

$$H\left(\hat{Y}_{r\mathcal{L}},Y_d^{(1)}\mid X_s\right)=\log\left((\pi e)^{L+1}\left(\sigma_d^2\prod_{i\in\mathcal{L}}(\sigma_r^2+\sigma_{wi}^2)\right)\right), \tag{A6}$$

$$H\left(\hat{Y}_{r\mathcal{S}}\mid Y_{r\mathcal{S}}\right)=\log\left((\pi e)^{|\mathcal{S}|}\prod_{i\in\mathcal{S}}\sigma_{wi}^2\right), \tag{A7}$$

$$H\left(Y_d^{(2)}\mid X_{r\mathcal{S}^C}\right)=\log\left(\pi e\left(\sigma_d^2+\sum_{i\in\mathcal{S}}\gamma_{id}P_r\right)\right), \tag{A8}$$

$$H\left(Y_d^{(2)}\mid X_{r\mathcal{L}}\right)=\log\left(\pi e\sigma_d^2\right). \tag{A9}$$

Insert equations of (A4) and (A6) to (A1),

$$I\left(X_s;\hat{Y}_{r\mathcal{L}},Y_d^{(1)}\right)=\log\left(1+\frac{\gamma_0 P_s}{\sigma_d^2}+\sum_{i\in\mathcal{L}}\frac{\gamma_i P_s}{\sigma_r^2+\sigma_{wi}^2}\right), \tag{A10}$$

Insert (A4), (A5) and (A7) into (A2), $f_{\mathcal{S}}\left(\sigma_{w\mathcal{L}}^2\right)$ is obtained. Insert (A8) and (A9) to (A3),

$$I\left(X_{r\mathcal{S}};Y_d^{(2)}\mid X_{r\mathcal{S}^C}\right)=\log\left(1+\frac{\sum_{i\in\mathcal{S}}\gamma_{id}P_r}{\sigma_d^2}\right). \tag{A11}$$

Then *Proposition* 1 is proved.

For the hybrid CF/DF case, the mutual information expressions are

$$I\left(X_s;Y_{ri}\right)=H\left(Y_{ri}\right)-H\left(Y_{ri}\mid X_s\right)=\log\left(\frac{\gamma_i P_s+\sigma_r^2}{\sigma_r^2}\right), \tag{A12}$$

$$I\left(X_s;\hat{Y}_{r\mathcal{D}^C},Y_d^{(1)}\right)=H\left(\hat{Y}_{r\mathcal{D}^C},Y_d^{(1)}\right)-H\left(\hat{Y}_{r\mathcal{D}^C},Y_d^{(1)}\mid X_s\right), \tag{A13}$$

$$I\left(X_{r\mathcal{D}};Y_d^{(2)}\mid X_{r\mathcal{D}^C}\right)=H\left(Y_d^{(2)}\mid X_{r\mathcal{D}^C}\right)-H\left(Y_d^{(2)}\mid X_{r\mathcal{L}}\right), \tag{A14}$$

$$I\left(\hat{Y}_{r\mathcal{S}};Y_{r\mathcal{S}}\mid\hat{Y}_{r\mathcal{S}^C},Y_d^{(1)}\right)=H\left(\hat{Y}_{r\mathcal{D}^C},Y_d^{(1)}\right)-H\left(\hat{Y}_{r\mathcal{S}^C},Y_d^{(1)}\right)-H\left(\hat{Y}_{r\mathcal{S}}\mid Y_{r\mathcal{S}}\right), \tag{A15}$$

$$I\left(X_{r\mathcal{S}};Y_d^{(2)}\mid X_{r\mathcal{S}^C}\right)=H\left(Y_d^{(2)}\mid X_{r\mathcal{S}^C}\right)-H\left(Y_d^{(2)}\mid X_{r\mathcal{D}^C}\right), \tag{A16}$$

$\forall\ \mathcal{S}\subseteq\mathcal{D}^C$. The entropy expressions are given as

$$H\left(Y_d^{(2)}\mid X_{r\mathcal{S}^C}\right)=\log\left(\pi e\left(\sum_{i\in\mathcal{D}}\gamma_{id}P_r+\sum_{i\in\mathcal{S}}\gamma_{id}P_r+\sigma_d^2\right)\right), \tag{A17}$$

$$H\left(Y_d^{(2)}\mid X_{r\mathcal{D}^C}\right)=\log\left(\pi e\left(\sum_{i\in\mathcal{D}}\gamma_{id}P_r+\sigma_d^2\right)\right). \tag{A18}$$

Inserting (A4)-(A9) and (A17)-(A18) to (A12)-(A16) gives (18).

APPENDIX B

Let $\mathbf{C}_X^c$ and $\mathbf{C}_Y^c$ be vectors of constraints. Let $R_{CF,X}$ and $R_{CF,Y}$ be the optimal solutions to the primal problem with constraints $\mathbf{C}_X^c$ and $\mathbf{C}_Y^c$. We assume that codebooks $\mathcal{X}_{X,i}$ and $\mathcal{X}_{Y,i}$ with coding rate $R_{X,i}$ and $R_{Y,i}$, respectively, are used in the link between the ith relay-destination link and satisfy the constraints. Assuming another case where

$$\mathbf{C}_Z^c=v\mathbf{C}_X^c+\left(1-v\right)\mathbf{C}_Y^c, \tag{B1}$$

with $0\le v\le1$ and considering the idea of time-sharing, we also assume a case Z that at the ith relay, codebook $\mathcal{X}_{Z,i}$ is constructed by using the first vn symbols of the first codebook $\mathcal{X}_{X,i}$ and the last $(1-v)n$ symbols of the second code book $\mathcal{X}_{Y,i}$. The rate of this new codebook is

$$R_{Z,i}=vR_{X,i}+(1-v)R_{Y,i}, \tag{B2}$$

and the constraint (B1) is also satisfied for case Z. Clearly, with the new code book, the rate can be achieved up to

$$R_{CF,Z}=vR_{CF,X} +(1-v)R_{CF,Y}. \tag{B3}$$

It is pointed out in [43], lemma 10.4.1] that time-sharing of compression cannot decrease the compression noise. Since the compression noises are at the denominators of (8), it also means that (8) cannot be increased by time-sharing of compression. Let $R^*{}_{CF,Z}$ be the optimal solution to the primal problem with constraints $\mathbf{C}_Z^c$, we have

$$R^*{}_{CF,Z} \geq R_{CF,Z} = vR_{CF,X} + (1-v)R_{CF,Y}. \tag{B4}$$

In addition, if we increase the constraint $\mathbf{C}^c$, the distortion, i.e., compression noise should be decreased in nature. Hence eq. (8) can be increased. It leads to the conclusion that $\Phi(\cdot)$ is a non-decreasing concave function with regard to the constraint $\mathbf{C}^c$.

Let us define an image set $\mathcal{I}$:

$$\mathcal{I} = \left\{ \left(z_0, z_1, ..., z_c\right) \in \mathbb{R}^{c+1} : z_i \geq f_i\left(\sigma_{w\mathcal{L}}^2\right), 1 \leq i \leq c \text{ and } z_0 = \Phi\left(\sigma_{w\mathcal{L}}^2\right) \text{ for } \sigma_{w\mathcal{L}}^2 \geq 0 \right\}, \tag{B5}$$

and a line in this $(c+1)$-dimension space:

$$\mathcal{I}_L \equiv \left\{ z_i = C_i, 1 \leq i \leq c \right\}, \tag{B6}$$

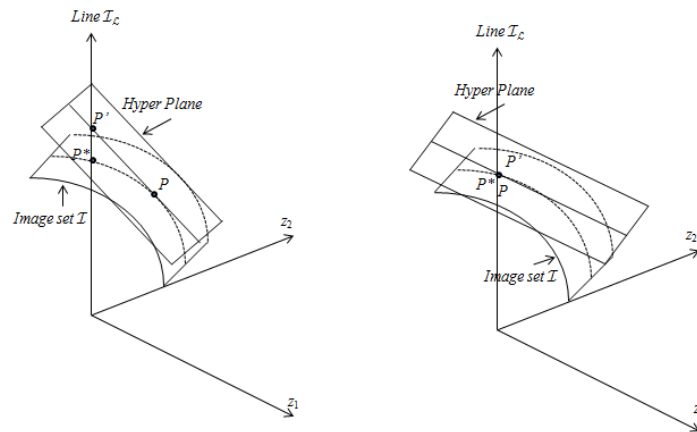as shown in figure B1, where c is assumed to be 2 to demonstrate the concept clearly.

Figure B1 Image set $\mathcal{I}$ ($c=2$)

The primal problem is to find the point in the image set $\mathcal{I}$ which maximizes z_0 subject to $z_i \leq C_i$, $1 \leq i \leq c$. Hence the intersection point of the image set $\mathcal{I}$ and the line $\mathcal{I}_L$ is the optimal solution to the primal problem because of the proven concavity and non-decreasing properties of function $\Phi(\cdot)$, denoted as

$$P^* = \left[\Phi\left(\sigma_{w\mathcal{L}}^{2\;*}\right), f_1\left(\sigma_{w\mathcal{L}}^{2\;*}\right), ..., f_c\left(\sigma_{w\mathcal{L}}^{2\;*}\right) \right], \tag{B7}$$

where $\sigma_{w\mathcal{L}}^{2\;*}$ maximizes the primal problem. To evaluate the dual function, we have, with fixed $\boldsymbol{\lambda}^c$,

$$\begin{aligned} &\text{maximize } \Phi\left(\sigma_{w\mathcal{L}}^2\right) + \boldsymbol{\lambda}^{cT}\left(\mathbf{C}^c - \mathbf{f}^c\left(\sigma_{w\mathcal{L}}^2\right)\right), \\ &\text{subject to } \sigma_{w\mathcal{L}}^2 \geq 0 \end{aligned} \tag{B8}$$

This is equivalent to

$$\begin{aligned} &\text{maximize } \quad z_0 + \sum_{i=0}^{c} \lambda_i \left(C_i - z_i\right), \\ &\text{subject to } \quad \left(z_0, ..., z_c\right) \in \mathcal{I} \end{aligned} \tag{B9}$$

The objective function $z_0 + \sum_{i=0}^{c} \lambda_i \left(C_i - z_i\right)$ is a hyperplane with slope $\boldsymbol{\lambda}^c$ and the problem corresponds to determining the lowest plane with slope $\boldsymbol{\lambda}^c$ which intersects with the image set $\mathcal{I}$. Again, due to the concavity and non-decreasing properties, the optimal value can be obtained from the hyperplane tangent to the image set $\mathcal{I}$ at point P. The intersection point of the hyperplane with line $\mathcal{I}_L$ is the optimal value of $\varphi(\boldsymbol{\lambda}^c)$, denoted as P'. A randomly selected hyperplane might not be able to make the duality gap zero as shown in the left of figure B1. However, among all the hyperplanes tangent to the image set $\mathcal{I}$, if we choose $\boldsymbol{\lambda}^{c*}$ such that hyperplane is tangent to the image set $\mathcal{I}$ at point $P=P'=P^*$, the duality gap is zero.

APPENDIX C

The Lagrangian function can be expressed as

$$\begin{aligned} L\left(\sigma_{w\mathcal{L}}^2, \boldsymbol{\lambda}^c\right) = &\left(1 - \sum_{l \in \mathcal{L}} \sum_{j=1}^{c_{L,l}} \lambda_{\mathcal{S}_j^l}\right) H\left(\hat{Y}_{r\mathcal{L}}, Y_d^{(1)}\right) - H\left(\hat{Y}_{r\mathcal{L}}, Y_d^{(1)} \mid X_s\right) \\ &+ \sum_{l \in \mathcal{L}} \sum_{j=1}^{c_{L,l}} \lambda_{\mathcal{S}_j^l} H\left(\hat{Y}_{r\mathcal{S}_j^{lC}}, Y_d^{(1)}\right) + \sum_{l \in \mathcal{L}} \sum_{j=1}^{c_{L,l}} \lambda_{\mathcal{S}_j^l} H\left(\hat{Y}_{r\mathcal{S}_j^l} \mid Y_{r\mathcal{S}_j^l}\right) + \sum_{l \in \mathcal{L}} \sum_{j=1}^{c_{L,l}} \lambda_{\mathcal{S}_j^l} C_{\mathcal{S}_j^l} \end{aligned} \tag{C1}$$

If we fix $\sigma^2_{w\mathcal{L}\backslash i}$, maximizing $L\left(\sigma^2_{w\mathcal{L}}, \boldsymbol{\lambda}^c\right)$ is equivalent to maximizing

$$\left(1-\sum_{l\in\mathcal{L}}\sum_{j=1}^{c_{L,l}}\lambda_{\mathcal{S}_j^l}\right)\log\left((\gamma_0 P_s+\sigma_d^2)\prod_{i\in\mathcal{L}}(\sigma_r^2+\sigma_{wi}^2)+\gamma_i P_s\sigma_d^2\prod_{k\in\mathcal{L},k\neq i}(\sigma_r^2+\sigma_{wk}^2)+\sum_{l\in\mathcal{L},l\neq i}\left(\gamma_l P_s\sigma_d^2\prod_{k\in\mathcal{L},k\neq l}(\sigma_r^2+\sigma_{wk}^2)\right)\right)$$

$$+\sum_{l\in\mathcal{L}}\sum_{j=1}^{c_{L,l}}\lambda_{\mathcal{S}_j^l}\log\left((\gamma_0 P_s+\sigma_d^2)\prod_{i\in\mathcal{S}_j^{lC}}(\sigma_r^2+\sigma_{wi}^2)+\sum_{i\in\mathcal{S}_j^{lC}}\left(\gamma_i P_s\sigma_d^2\prod_{k\in\mathcal{S}_j^{lC},k\neq i}(\sigma_r^2+\sigma_{wk}^2)\right)\right)+\sum_{l\in\mathcal{L}}\sum_{j=1}^{c_{L,l}}\lambda_{\mathcal{S}_j^l}\log\left(\prod_{i\in\mathcal{S}_j^l}\sigma_{wi}^2\right) \qquad \text{.(C2)}$$

$$-\log\left(\sigma_d^2(\sigma_r^2+\sigma_{wi}^2)\right)$$

$\mathcal{C}_i^C$ does not contain element i. Hence maximizing the third term of (C2) is equivalent to maximizing $\sum_{\mathcal{S}_j^l\in\mathcal{C}_i}\lambda_{\mathcal{S}_j^l}\log(\sigma_{wi}^2)$ with fixed $\sigma^2_{w\mathcal{L}/i}$. The first and second terms of (C2) can be re-organized as (C3) and (C4), respectively.

$$\left(1-\sum_{l\in\mathcal{L}}\sum_{j=1}^{c_{L,l}}\lambda_{\mathcal{S}_j^l}\right)\log\left(\sigma_r^2+\sigma_{wi}^2+\sigma_r^2 A_i\right)$$

$$+\left(1-\sum_{l\in\mathcal{L}}\sum_{j=1}^{c_{L,l}}\lambda_{\mathcal{S}_j^l}\right)\log\left((\gamma_0 P_s+\sigma_d^2)\prod_{k\in\mathcal{L},k\neq i}(\sigma_r^2+\sigma_{wk}^2)+\sum_{l\in\mathcal{L},l\neq i}\left(\gamma_l P_s\sigma_d^2\prod_{k\in\mathcal{L},k\neq i,k\neq l}(\sigma_r^2+\sigma_{wk}^2)\right)\right) \qquad \text{, (C3)}$$

$$\sum_{\mathcal{S}_j^l\in\mathcal{C}_i^C}\lambda_{\mathcal{S}_j^l}\log\left(\sigma_r^2+\sigma_{wi}^2+\sigma_r^2 B_{\mathcal{S}_j^l,i}\right)$$

$$+\sum_{\mathcal{S}_j^l\in\mathcal{C}_i^C}\lambda_{\mathcal{S}_j^l}\log\left((\gamma_0 P_s+\sigma_d^2)\prod_{\substack{k\in\mathcal{S}_j^{lC}\\k\neq i}}(\sigma_r^2+\sigma_{wk}^2)+\sum_{\substack{l\in\mathcal{S}_j^{lC}\\l\neq i}}\left(\gamma_l P_s\sigma_d^2\prod_{\substack{k\in\mathcal{S}_j^{lC}\\k\neq i,k\neq l}}(\sigma_r^2+\sigma_{wk}^2)\right)\right) \qquad \text{(C4)}$$

where

$$A_i=\frac{\gamma_i P_s\sigma_d^2\prod_{k\in\mathcal{L},k\neq i}(\sigma_r^2+\sigma_{wk}^2)}{\sigma_r^2\left((\gamma_0 P_s+\sigma_d^2)\prod_{k\in\mathcal{L},k\neq i}(\sigma_r^2+\sigma_{wk}^2)+\sum_{l\in\mathcal{L},l\neq i}\left(\gamma_l P_s\sigma_d^2\prod_{k\in\mathcal{L},k\neq i,k\neq l}(\sigma_r^2+\sigma_{wk}^2)\right)\right)},$$

$$B_{\mathcal{S}_j^l,i}=\frac{\gamma_i P_s\sigma_d^2\prod_{k\in\mathcal{S}_j^{lC},k\neq i}(\sigma_r^2+\sigma_{wk}^2)}{\sigma_r^2\left((\gamma_0 P_s+\sigma_d^2)\prod_{k\in\mathcal{S}_j^{lC},k\neq i}(\sigma_r^2+\sigma_{wk}^2)+\sum_{l\in\mathcal{S}_j^{lC},l\neq i}\left(\gamma_l P_s\sigma_d^2\prod_{k\in\mathcal{S}_j^{lC},k\neq i,k\neq l}(\sigma_r^2+\sigma_{wk}^2)\right)\right)},$$

<div align="right">(C5)</div>

$\forall i \in \mathcal{L}$, $\mathcal{S}_j^l \in \mathcal{C}_i^C$. The second terms of (C3) and (C4) do not contain σ_{wi}^2 and thus can be ignored when maximizing with respect to σ_{wi}^2. We normalize (C2) with σ_r^2 and insert (C3)-(C5) to (C2) to give (12).

APPENDIX D

Since the logarithm function is a monotonically increasing function, maximizing (12) can be further simplified as maximizing

$$\theta\left(\boldsymbol{\lambda}^c, \underline{\sigma}_{wi}^2\right) = \frac{\left(1+\underline{\sigma}_{wi}^2 + A_i\right)^{(1-\lambda_\Sigma)} \displaystyle\prod_{\mathcal{S}_j^l \in \mathcal{C}_i^C}\left(1+\underline{\sigma}_{wi}^2 + B_{\mathcal{S}_j^l i}\right)^{\lambda_{\mathcal{S}_j^l}} \left(\underline{\sigma}_{wi}^2\right)^{\lambda_{\mathcal{C}_i}}}{1+\underline{\sigma}_{wi}^2},$$

$$= \frac{y\left(\boldsymbol{\lambda}^c, \underline{\sigma}_{wi}^2\right)}{1+\underline{\sigma}_{wi}^2} \tag{D1}$$

subject to $\underline{\sigma}_{wi}^2 > 0$, where

$$y\left(\boldsymbol{\lambda}^c, \underline{\sigma}_{wi}^2\right) = \left(1+\underline{\sigma}_{wi}^2 + A_i\right)^{(1-\lambda_\Sigma)} \prod_{\mathcal{S}_j^l \in \mathcal{C}_i^C}\left(1+\underline{\sigma}_{wi}^2 + B_{\mathcal{S}_j^l i}\right)^{\lambda_{\mathcal{S}_j^l}} \left(\underline{\sigma}_{wi}^2\right)^{\lambda_{\mathcal{C}_i}}.$$

The derivative of (D1) is given as,

$$\frac{\partial \theta\left(\boldsymbol{\lambda}^c, \underline{\sigma}_{wi}^2\right)}{\partial \underline{\sigma}_{wi}^2} = \frac{\left(1+\underline{\sigma}_{wi}^2\right) y'\left(\boldsymbol{\lambda}^c, \underline{\sigma}_{wi}^2\right) - y\left(\boldsymbol{\lambda}^c, \underline{\sigma}_{wi}^2\right)}{\left(1+\underline{\sigma}_{wi}^2\right)^2}, \tag{D2}$$

where

$$y'\left(\boldsymbol{\lambda}^c, \underline{\sigma}_{wi}^2\right) = \left(1-\lambda_\Sigma\right)\left(1+\underline{\sigma}_{wi}^2 + A_i\right)^{-\lambda_\Sigma} \prod_{\mathcal{S}_j^l \in \mathcal{C}_i^C}\left(1+\underline{\sigma}_{w,i}^2 + B_{\mathcal{S}_j^l i}\right)^{\lambda_{\mathcal{S}_j^l}} \left(\underline{\sigma}_{wi}^2\right)^{\lambda_{\mathcal{C}_i}}$$

$$+\left(1+\underline{\sigma}_{wi}^2 + A_i\right)^{(1-\lambda_\Sigma)} \sum_{\mathcal{S}_j^l \in \mathcal{C}_i^C}\left(\lambda_{\mathcal{S}_j^l}\left(1+\underline{\sigma}_{wi}^2 + B_{\mathcal{S}_j^l i}\right)^{\lambda_{\mathcal{S}_j^l}-1} \prod_{\substack{\hat{\mathcal{S}}_j^l \in \mathcal{C}_i^C, \\ \hat{\mathcal{S}}_j^l \neq \mathcal{S}_j^l}}\left(1+\underline{\sigma}_{wi}^2 + B_{\hat{\mathcal{S}}_j^l i}\right)^{\lambda_{\hat{\mathcal{S}}_j^l}}\right)\left(\underline{\sigma}_{wi}^2\right)^{\lambda_{\mathcal{C}_i}}.$$

$$+\lambda_{\mathcal{C}_i}\left(1+\underline{\sigma}_{wi}^2 + A_i\right)^{(1-\lambda_\Sigma)} \prod_{\mathcal{S}_j^l \in \mathcal{C}_i^C}\left(1+\underline{\sigma}_{wi}^2 + B_{\mathcal{S}_j^l i}\right)^{\lambda_{\mathcal{S}_j^l}} \left(\underline{\sigma}_{wi}^2\right)^{\lambda_{\mathcal{C}_i}-1}$$

We rewrite (D2) as

$$\frac{\partial\theta\left(\boldsymbol{\lambda}^c,\underline{\sigma}_{wi}^2\right)}{\partial\underline{\sigma}_{wi}^2}=\frac{\hat{y}\left(\boldsymbol{\lambda}^c,\underline{\sigma}_{wi}^2\right)\left(\left(1+\underline{\sigma}_{wi}^2\right)\dfrac{y'\left(\boldsymbol{\lambda}^c,\underline{\sigma}_{wi}^2\right)}{\hat{y}\left(\boldsymbol{\lambda}^c,\underline{\sigma}_{wi}^2\right)}-\dfrac{y\left(\boldsymbol{\lambda}^c,\underline{\sigma}_{wi}^2\right)}{\hat{y}\left(\boldsymbol{\lambda}^c,\underline{\sigma}_{wi}^2\right)}\right)}{\left(1+\underline{\sigma}_{wi}^2\right)^2}, \tag{D3}$$

where

$$\hat{y}\left(\boldsymbol{\lambda}^c,\underline{\sigma}_{wi}^2\right)=\left(1+\underline{\sigma}_{wi}^2+A_i\right)^{-\lambda_\Sigma}\prod_{\mathcal{S}_j^l\in\mathcal{C}_i^C}\left(1+\underline{\sigma}_{wi}^2+B_{\mathcal{S}_j^l i}\right)^{\lambda_{\mathcal{S}_j^l}-1}\left(\underline{\sigma}_{wi}^2\right)^{\lambda_{\mathcal{C}_i}-1}>0.$$

Thus letting (D3) be 0 is equivalent to letting

$$q\left(\boldsymbol{\lambda}^c,\underline{\sigma}_{wi}^2\right)=\left(1+\underline{\sigma}_{wi}\right)\frac{y'\left(\boldsymbol{\lambda}^c,\underline{\sigma}_{wi}^2\right)}{\hat{y}\left(\boldsymbol{\lambda}^c,\underline{\sigma}_{wi}^2\right)}-\frac{y\left(\boldsymbol{\lambda}^c,\underline{\sigma}_{wi}^2\right)}{\hat{y}\left(\boldsymbol{\lambda}^c,\underline{\sigma}_{wi}^2\right)}=0. \tag{D4}$$

Combine (D1) to (D3), (D4) can be given as

$$
\begin{aligned}
&q\left(\boldsymbol{\lambda}^c,\underline{\sigma}_{wi}^2\right)\\
&=\left(1+\underline{\sigma}_{w,i}^2\right)\left[
\begin{aligned}
&\left(1-\lambda_\Sigma\right)\underline{\sigma}_{wi}^2\prod_{\mathcal{S}_j^l\in\mathcal{C}_i^C}\left(1+\underline{\sigma}_{wi}^2+B_{\mathcal{S}_j^l i}\right)+\lambda_{\mathcal{C}_i}\left(1+\underline{\sigma}_{wi}^2+A_i\right)\prod_{\mathcal{S}_j^l\in\mathcal{C}_i^C}\left(1+\underline{\sigma}_{wi}^2+B_{\mathcal{S}_j^l i}\right)\\
&+\underline{\sigma}_{wi}^2\left(1+\underline{\sigma}_{wi}^2+A_i\right)\sum_{\mathcal{S}_j^l\in\mathcal{C}_i^C}\lambda_{\mathcal{S}_j^l}\prod_{\mathcal{S}_k^l\in\mathcal{C}_i^C,k\neq j}\left(1+\underline{\sigma}_{wi}^2+B_{\mathcal{S}_j^l k}\right)
\end{aligned}
\right]\\
&\quad-\underline{\sigma}_{wi}^2\left(1+\underline{\sigma}_{wi}^2+A_i\right)\prod_{\mathcal{S}_j^l\in\mathcal{C}_i^C}\left(1+\underline{\sigma}_{wi}^2+B_{\mathcal{S}_j^l i}\right)
\end{aligned} \tag{D5}
$$

Clearly, $q\left(\boldsymbol{\lambda}^c,\underline{\sigma}_{wi}^2\right)$ is a polynomial function of $\underline{\sigma}_{wi}^2$ and the term with the highest degree is

$$\left(1-\lambda_\Sigma\right)\left(\underline{\sigma}_{wi}^2\right)^{\left|\mathcal{C}_i^C\right|+2}+\sum_{\mathcal{S}_j^l\in\mathcal{C}_i^C}\lambda_{\mathcal{S}_j^l}\left(\underline{\sigma}_{wi}^2\right)^{\left|\mathcal{C}_i^C\right|+2}+\lambda_{\mathcal{C}_i}\left(\underline{\sigma}_{wi}^2\right)^{\left|\mathcal{C}_i^C\right|+2}-\left(\underline{\sigma}_{wi}^2\right)^{\left|\mathcal{C}_i^C\right|+2}=0. \tag{D6}$$

Therefore, the actual highest degree of $q\left(\boldsymbol{\lambda}^c,\underline{\sigma}_{wi}^2\right)$ is $\left|\mathcal{C}_i^C\right|+1=(c+1)/2$ and $q\left(\boldsymbol{\lambda}^c,\underline{\sigma}_{wi}^2\right)$ has $(c+1)/2$

roots, denoted as a set $\mathcal{X}=\{\underline{\sigma}_{w1}^2,\ldots,\underline{\sigma}_{w(c+1)/2}^2\}$. However, not all members of root set $\mathcal{X}$ are viable

solutions. Since $\underline{\sigma}_{wi}^2>0$, only positive roots should be considered. We define $\mathcal{X}_{\text{sub}}$ as a subset of $\mathcal{X}$

including only positive elements.

   Considering a special case that $\underline{\sigma}_{wi}^2\to+\infty$, we have

$$\lim_{\sigma_{wi}^2 \to +\infty} q\left(\lambda^c, \underline{\sigma}_{wi}^2\right)$$

$$= \underline{\sigma}_{wi}^2 \left((1-\lambda_\Sigma)\left(\underline{\sigma}_{wi}^2\right)^{\left|\mathcal{C}_i^C\right|+1} + \lambda_{\mathcal{C}_i}\left(\underline{\sigma}_{wi}^2\right)^{\left|\mathcal{C}_i^C\right|+1} + \left(\sum_{\mathcal{S}_j^l \in \mathcal{C}_i^C} \lambda_{\mathcal{S}_j^l} \right)\left(\underline{\sigma}_{wi}^2\right)^{\left|\mathcal{C}_i^C\right|+1} \right) - \underline{\sigma}_{wi}^2 {}^{\left|\mathcal{C}_i^C\right|+2}, \quad \text{(D7)}$$

$$= 0$$

which means $+\infty$ is also a local maximum. Therefore, a new set is defined as $\underline{\mathcal{X}}_{\text{sub}}=\{\mathcal{X}_{\text{sub}},+\infty\}$.

Note that the members of $\underline{\mathcal{X}}_{\text{sub}}$ only guarantee that the first derivative of $\theta\left(\lambda^c, \underline{\sigma}_{wi}^2\right)$ is zero at those points thus can be regarded as local maximum or minimum. In order to make sure that only maximum is found, the optimal $\underline{\sigma}_{wi}^2 *$ should be chosen as

$$\underline{\sigma}_{wi}^{2 \ *} = \underset{\underline{\sigma}_{wi}^2 \in \underline{\mathcal{X}}_{\text{sub}}}{\arg\max} \, \theta\left(\lambda^c, \underline{\sigma}_{wi}^2\right). \quad \text{(D8)}$$

If $L=2$, $q\left(\lambda^c, \underline{\sigma}_{wi}^2\right)$ can be expressed as

$$q\left(\lambda^c, \underline{\sigma}_{wi}^2\right)$$

$$= \left(1+\underline{\sigma}_{wi}^2\right)\left((1-\lambda_\Sigma)\underline{\sigma}_{wi}^2\left(1+\underline{\sigma}_{wi}^2+B_{\mathcal{S}_j^l i}\right) + \lambda_{\mathcal{C}_j}\left(1+\underline{\sigma}_{wi}^2+A_i\right)\left(1+\underline{\sigma}_{wi}^2+B_{\mathcal{S}_j^l i}\right) + \lambda_{\mathcal{S}_j^l i}\underline{\sigma}_{wi}^2\left(1+\underline{\sigma}_{wi}^2+A_i\right) \right)$$

$$- \underline{\sigma}_{wi}^2\left(1+\underline{\sigma}_{wi}^2+A_i\right)\left(1+\underline{\sigma}_{wi}^2+B_{\mathcal{S}_j^l i}\right)$$

$$= a_{1,i}\left(\underline{\sigma}_{wi}^2\right)^2 + a_{2,i}\underline{\sigma}_{wi}^2 + a_{3,i}$$

$$= a_{1,i}\left(\left(\underline{\sigma}_{wi}^2 + \frac{a_{2,i}}{2a_{1,i}}\right)^2 + \frac{4a_{1,i}a_{3,i}-a_{2,i}^2}{4a_{1,i}^2} \right)$$

$$= a_{1,i}\left(\left(\underline{\sigma}_{wi}^2 + \underline{a}_{2,i}\right)^2 + \underline{a}_{3,i} \right)$$

,(D9)

where either $i=1$ and $j=2$ or $i=2$ and $j=1$, and

$$a_{1,i} = \left(1-\lambda_{\Sigma}\right)\left(1+B_{S_{ji}^l}\right) + \lambda_{C_i}\left(2+A_i+B_{S_{ji}^l}\right) + \lambda_{S_{ji}^l}\left(1+A_i\right) - \left(1+A_i+B_{S_{ji}^l}\right),$$

$$a_{2,i} = \left(1-\lambda_{\Sigma}\right)\left(1+B_{S_{ji}^l}\right) + \lambda_{C_i}\left(2+A_i+B_{S_{ji}^l}\right) + \lambda_{S_{ji}^l}\left(1+A_i\right) + \left(\lambda_{C_i}-1\right)\left(1+A_i\right)\left(1+B_{S_{ji}^l}\right),$$

$$a_{3,i} = \lambda_{C_i}\left(1+A_i\right)\left(1+B_{S_{ji}^l}\right),$$

$$\underline{a}_{2,i} = \frac{a_{2,i}}{2a_{1,i}},$$

$$\underline{a}_{3,i} = \frac{4a_{1,i}a_{3,i}-a_{2,i}^2}{4a_{1,i}^2}.$$

Thus $q\left(\lambda^c, \underline{\sigma}_{wi}^2\right)$ is a two degree polynomial. Similar to (D3), let

$$a_{1,i}\left(\left(\sigma_{wi}^2+\underline{a}_{2,i}\right)^2+\underline{a}_{3,i}\right)=0. \tag{D10}$$

The viable solutions are discussed in the following table.

Table-D1

$a_{1,i}$	$\underline{a}_{2,i}$	$\underline{a}_{3,i}$	Solutions to $q\left(\lambda^c, \underline{\sigma}_{wi}^2\right)$
>0	X	>0	$q\left(\lambda^c, \underline{\sigma}_{wi}^2\right)>0$ and $\theta\left(\lambda^c, \underline{\sigma}_{wi}^2\right)$ is monotonically increasing with the maximum value achieved at $+\infty$.
>0	>0	≤0	Same as above.
>0	<0	≤0	(D10) has two roots: $\pm\sqrt{-\underline{a}_{3,i}}-\underline{a}_{2,i}$. Function $q\left(\lambda^c, \underline{\sigma}_{wi}^2\right)$ intersects with q-axial at point $(0, a_{3,i})$, where $a_{3,i}\geq0$. In such a case, $-\sqrt{-\underline{a}_{3,i}}-\underline{a}_{2,i}$ is only a local maximum. If $\underline{\sigma}_{wi}^2 > +\sqrt{-\underline{a}_{3,i}}-\underline{a}_{2,i}$, $q\left(\lambda^c, \underline{\sigma}_{wi}^2\right)>0$ and $\theta\left(\lambda^c, \underline{\sigma}_{wi}^2\right)$ is monotonically increasing to achieve another maximum at $+\infty$. Hence we need to compare two local maximums to find the global maximum.
<0	X	X	(D10) has two roots: $\pm\sqrt{-\underline{a}_{3,i}}-\underline{a}_{2,i}$. Function $q\left(\lambda^c, \underline{\sigma}_{wi}^2\right)$

			intersects with q-axial at point $(0, a_{3,i})$, where $a_{3,i} \geq 0$. Only the larger solution $+\sqrt{-\underline{a}_{3,i}} - \underline{a}_{2,i} > 0$ and serves as a global maximum.
=0	≥ 0	X	$\theta(\lambda^c, \underline{\sigma}_{wi}^2)$ is monotonically increasing to achieve another maximum at $+\infty$.
=0	<0	X	(D10) has one root: $-a_{3,i} / a_{2,i}$, where the global maximum is achieved.

X means either positive or negative.

APPENDIX E

In order to maximize the Lagrangian dual (11), we can maximize the following equation

$$I\left(X_s; \hat{Y}_{r\mathcal{L}}, Y_d^{(1)}\right) - \sum_{l \in \mathcal{L}} \sum_{j=1}^{c_{L,l}} \lambda_{\mathcal{S}_j^l} I\left(\hat{Y}_{r\mathcal{S}_j^l}; Y_{r\mathcal{S}_j^l} \mid \hat{Y}_{r\mathcal{S}_j^l C}, Y_d^{(1)}\right),$$ (E1)

where $\lambda_{\mathcal{S}_j^l} \geq 0$, $\forall \mathcal{S}_j^l \subseteq \mathcal{L}$, $l \in \mathcal{L}$, $1 \leq j \leq |\mathcal{S}_j^l|$. For a particular $\lambda_{\mathcal{S}_j^l}$ we have

$$I\left(X_s; \hat{Y}_{r\mathcal{L}}, Y_d^{(1)}\right) - \sum_{l \in \mathcal{L}} \sum_{j=1}^{c_{L,l}} \lambda_{\lambda_{\mathcal{S}_j^l}} I\left(\hat{Y}_{r\mathcal{S}_j^l}; Y_{r\mathcal{S}_j^l} \mid \hat{Y}_{r\mathcal{S}_j^l C}, Y_d^{(1)}\right) \leq I\left(X_s; \hat{Y}_{r\mathcal{L}}, Y_d^{(1)}\right) - \lambda_{\mathcal{S}_j^l} I\left(\hat{Y}_{r\mathcal{S}_j^l}; Y_{r\mathcal{S}_j^l} \mid \hat{Y}_{r\mathcal{S}_j^l C}, Y_d^{(1)}\right). \text{ (E2)}$$

For $\forall i \in \mathcal{S}_j^l$, if we maximize the right term with respect to σ_{wi}^2 only, following the similar derivation as in Appendix C, it is given as

$$\Gamma\left(\underline{\sigma}_{wi}^2\right) = \left(1 - \lambda_{\mathcal{S}_j^l}\right) \log\left(1 + \underline{\sigma}_{wi}^2 + A_i\right) + \lambda_{\mathcal{S}_j^l} \log\left(\underline{\sigma}_{wi}^2\right) - \log\left(1 + \underline{\sigma}_{wi}^2\right)$$

$$= \log\left(\frac{\left(1 + \underline{\sigma}_{wi}^2 + A_i\right)^{1 - \lambda_{\mathcal{S}_j^l}} \underline{\sigma}_{wi}^{2 \lambda_{\mathcal{S}_j^l}}}{1 + \underline{\sigma}_{wi}^2}\right) = \log u\left(\underline{\sigma}_{wi}^2\right)$$ (E3)

The first derivative of $u(\cdot)$ is given as

$$\frac{\partial u\left(\underline{\sigma}_{wi}^2\right)}{\partial \underline{\sigma}_{wi}^2} = \frac{\lambda_{\mathcal{S}_j^l}\left(1 + \underline{\sigma}_{wi}^2\right)\left(1 + A_i\right) - A_i \underline{\sigma}_{wi}^2}{\left(1 + \underline{\sigma}_{wi}^2\right)^2 \left(1 + \underline{\sigma}_{wi}^2 + A_i\right)^{-\lambda_{\mathcal{S}_j^l}} \underline{\sigma}_{wi}^{2 \lambda_{\mathcal{S}_j^l} - 1}}.$$ (E4)

If $\lambda_{\mathcal{S}_j^l} \geq 1$, the numerator of (E4) is

$$\lambda_{\mathcal{S}_j^l}\left(1 + \underline{\sigma}_{wi}^2\right)\left(1 + A_i\right) - A_i \underline{\sigma}_{wi}^2 \geq \left(1 + \underline{\sigma}_{wi}^2\right)\left(1 + A_i\right) - A_i \underline{\sigma}_{wi}^2 = 1 + A_i + \underline{\sigma}_{wi}^2.$$ (E5)

Since $\underline{\sigma}_{wi}^2 \geq 0$ and $A_i > 0$, it follows that $\dfrac{\partial u\left(\underline{\sigma}_{wi}^2\right)}{\partial \underline{\sigma}_{wi}^2} > 0$, which means that $u(\cdot)$ is a monotonically increasing function of $\underline{\sigma}_{wi}^2$ and its maximum is achieved when $\underline{\sigma}_{wi}^2 \to +\infty$ and given as

$$\max_{\underline{\sigma}_{wi}^2} u\left(\underline{\sigma}_{wi}^2\right) = \lim_{\underline{\sigma}_{wi}^2 \to +\infty} \frac{\left(1+\underline{\sigma}_{wi}^2+A_i\right)^{1-\lambda_{\mathcal{S}_j^l}} \underline{\sigma}_{wi}^{2\lambda_{\mathcal{S}_j^l}}}{1+\underline{\sigma}_{wi}^2} = 1. \tag{E6}$$

Hence

$$I\left(X_s;\hat{Y}_{r\mathcal{L}},Y_d^{(1)}\right) - \sum_{l \in \mathcal{L}}\sum_{j=1}^{c_{L,l}} \lambda_{\mathcal{S}_j^l} I\left(\hat{Y}_{r\mathcal{S}_j^l};Y_{r\mathcal{S}_j^l} \mid \hat{Y}_{r\mathcal{S}_j^{lC}},Y_d^{(1)}\right) \leq \log 1 = 0. \tag{E7}$$

Considering

$$\lim_{\underline{\sigma}_{wi}^2 \to +\infty} \hat{\varphi}\left(\boldsymbol{\lambda}^c\right) = \lim_{\underline{\sigma}_{wi}^2 \to +\infty} \log \frac{\left(\underline{\sigma}_{wi}^2\right)^{1-\lambda_\Sigma}\left(\underline{\sigma}_{wi}^2\right)^{\lambda_{c_i^C}}\left(\underline{\sigma}_{wi}^2\right)^{\lambda_{c_i}}}{\underline{\sigma}_{wi}^2} = 0, \tag{E8}$$

we can conclude that if $\lambda_{\mathcal{S}_j^l} \geq 1$, $\arg\max\limits_{\underline{\sigma}_{wi}^2} \hat{\varphi}\left(\boldsymbol{\lambda}^c\right) = +\infty$.

REFERENCES

[1] E.C. Van Der Meulen, "Three-terminal communication channel," Adv. Appl. Prob., vol. 3, pp. 120-154, 1971.

[2] T. Cover, and A. E. Gamal, "Capacity theorems for the relay channel," *IEEE Trans. Inform. Theory*, vol. 25, pp. 572-584, Sep. 1979.

[3] R. U. Nabar, H. Bolcskei and F.W. Kneubuhler, "Fading relay channels: performance limits and space-time signal design*," IEEE J. Sel. Area. in Comm.*, vol. 22, no. 6, pp. 1099-1109, Aug 2004.

[4] L. Zhao and Z. Liao, "Power allocation for amplify-and-forward cooperative transmission over Rayleigh-fading channels," *Journal of Comm.*, vol. 3, no. 3, pp.33-42, July 2008.

[5] M. Janani, A. Hedayat, T. E. Hunter, and A. Nosratinia, "Coded cooperation in wireless communications: space-time transmission and iterative decoding," *IEEE Trans. Signal Processing*, vol. 52, pp 362-371, Feb. 2004.

[6] J. N. Laneman, D. N. C. Tse, G. W. Wornell, "Cooperative diversity in wireless networks: Efficient protocols and outage behavior", *IEEE Trans. Inform. Theory*, vol. 50, no. 12, pp. 3062–3080, Dec. 2004.

[7] M. Katz and S. Shamai, "Relay protocols for two colocated users," *IEEE Trans. Inform. Theory*, vol. 52, No. 6, pp. 2329-2344, Jun. 2006.

[8] A. Host-Madsen and J. Zhang, "Capacity bounds and power allocation for wireless relay channels," *IEEE Trans. Inform. Theory*, vol. 51, pp. 2020 – 2040, Jun 2005.

[9] X. Mestre, D. Gregoratti, "Diversity analysis of randomized linear dispersion codes in a half-duplex amplify-and-forward multiple-relay system," *IEEE Trans. Inform. Theory*, vol. 59, no.5, pp. 2936 – 2959, May 2013.

[10] S. A. Jafar, K. S. Gomadam and C. Huang, "Duality and rate optimization for multiple access and broadcast channels with amplify-and-forward relays," *IEEE Trans. Inform. Theory*, vol. 53, no. 10, pp. 3350-3370, Oct. 2007.

[11] Y. Hu, J. Gross, "On the outage probability and effective capacity of multiple decode-and-forward relay system," in *Proc. of Wireless Days (WD) 2012 IFIP*, pp. 1-8, Dublin, Nov. 2012.

[12] S. Choi, J. Baek, J. Han, and J. Seo, "Channel estimation using extended orthogonal codes for AF multiple-relay networks over frequency-selective fading channels," *IEEE Trans. Wireless Commun.*, vol. 63, no. 1, pp. 417-423, Jan. 2014.

[13] H. Eghbali, S. Muhaidat, S. A. Hejazi, and Y. Ding, "Relay selection strategies for single-carrier frequency-domain equalization multi-relay cooperative networks," *IEEE Trans. Wireless Commun.*, vol. 12, no. 5, pp. 2034-2045, May 2013.

[14] X. Zhang, Z. Yan, Y. Gao, W. Wang, "On the study of outage performance for cognitive relay networks (CRN) with the Nth best-relay selection in Rayleigh-fading channels," *IEEE Trans. Wireless Commun. Letters*, vol. 2, no. 1, pp. 110-113, Feb. 2013.

[15] Q. Huo, T. Liu, S. Sun, L. Song, and B. Jiao, "Selective combining for hybrid cooperative networks," IET Commun., vol. 8, pp. 471-482, 2014.

[16] I. K. Sileh, W. Xiang, A. Maxwell, "Joint subcarrier paring and resource allocation for adaptive hybrid relay protocol in OFDM systems," in *Proc. of IEEE PIMRC* 2013, pp. 2011-2015, London, UK, Sept. 2013.

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

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

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

[20] G. Kramer, M. Gastpar, and P. Gupta, "Cooperative strategies and capacity theorems for relay networks," *IEEE Trans. Inform. Theory*, vol. 51, No. 9, pp. 3037-3063, Sep. 2005.

[21] A. Sanderovich, S. Shamai, Y. Steinberg, and G. Kramer, "Communication via decentralized processing," *IEEE Trans. Inform. Theory*, vol. 54, No. 7, pp. 3008-3023, Jul. 2008.

[22] X. Wu and L Xie, "On the optimal compressions in the compress-and-forward relay schemes," *IEEE Trans. Inform. Theory*, vol. 59, no. 5, pp.2613-2628, May 2013.

[23] J. Chen and T. Berger, "Distributed source coding via successive refinement," 42nd Annual Allerton Conference on Communication, Control, and Computing 2006, online at www.ece.mcmaster.ca/~junchen/Allerton2004.pdf.

[24] X. Wu and L. Xie, "A unified relay framework with both D-F and C-F relay nodes," *IEEE Trans. Wireless Comm.*, vol. 60, no. 1, pp. 586-604, Jan. 2014.

[25] C. A. Floudas, *NonLinear and mixed-integer optimization*, Oxford Univ. Press, 1995.

[26] M.S. Bazaraa, H. Sherali, and C. M. Shetty, *Nonlinear programming: theory and algorithms*, 3rd. John Wiley, 2006.

[27] S. Yao. T. T. Kim, M. Skoglund, and H. V. Poor, "Half-duplex relaying over slow fading channels based on quantize-and-forward," *IEEE Trans. on Inf. Theory*, vol.59, no.2, pp.860 - 872, Feb. 2013.

[28] I. Ku, C. Wang, and J. Thompson, "Spectral-Energy efficiency tradeoff in relay-aided cellular networks," *IEEE Trans. on Wireless Comm.*, vol. 12, no. 10, pp. 4970-4982, Oct. 2013.

[29] Y. Qi, R. Hoshyar, M. A. Imran, R. Tafazolli, "H^2-ARQ_relaying: spectrum and energy efficiency perspectives" *IEEE J. Sel. Area. in Comm.*, vol. 29, No. 8, Sep. 2011.

[30] S. Cui, A. J. Goldsmith, and A. Bahai, "Energy-efficiency of MIMO and cooperative MIMO techniques in sensor networks," *IEEE J. Sel. Area. in Comm.*, vol. 22, no. 6, pp. 1089-1098, Aug. 2004.

[31] I. Stanojev, O. Simeone, Y. Bar-Ness, and D. Kim, "Energy efficiency of non-collaborative and collaborative hybrid-ARQ protocols," *IEEE Trans. on Wireless Comm.*, vol. 8, no. 1, pp. 326-335, Jan. 2009.

[32] I. Godor, et al, "D3.1: Most promising tracks of green network technologies," Tech. Rep. of FP7 project EARTH, Dec. 2010, online at https://www.ict-earth.eu/publications/deliverables/deliverables.html.

[33] F. Heliot et al, "D3.2: Green network technologies," Tech. Rep. of FP7 project EARTH, Dec. 2011, online at https://www.ict-earth.eu/publications/deliverables/deliverables.html.

[34] I. Godor et al, "D3.3: Final report on green network technologies," Tech. Rep. of FP7 project EARTH, Jun. 2012, online at https://www.ict-earth.eu/publications/deliverables/deliverables.html.

[35] H. Holtkamp et al, "D2.2: Definition and parameterization of reference systems and scenarios," Tech. Rep. of FP7 project EARTH, Jun. 2010, available online at https://www.ict-earth.eu/publications/deliverables/deliverables.html.

[36] Y. Qi, M. A. Imran, D. Sebella, B. Debaillie, R. Fantini and Y. Fernandes, "On the deployment opportunities for increasing energy efficiency in LTE-Advanced with relay nodes," in *Proc. of WWRF* 2012.

[37] 3GPP TR 36.814 v1.5.1 (2009-12), Further Advancements for E-UTRA, Physical Layer Aspects.

[38] M. A. Imran et al, "D2.3: Energy efficiency analysis of the reference systems, areas of improvements and target breakdown," Tech. Rep., Nov. 2010, online at https://www.ict-earth.eu/publications/deliverables/deliverables.html.

[39] G. Auer et al, "How much energy is needed to run a wireless network?" *IEEE Wireless Comm. Magazine*, vol. 18, pp. 40-49, Oct. 2011.

[40] B. Debaillie et al, "Flexible power modeling of LTE base stations," in *Proc. of IEEE WCNC 2012*, Paris, Apr. 2011.

[41] J. Huang et al, "A close examination of performance and power characteristics of 4G LTE networks," in *Proc. of MObiSys'12*, UK, Jun. 2012.

[42] Gao, Q. and Blow, K. J. and Holding, D. J. and Marshall, I. W. and Peng, X. H., *Radio range adjustment for energy efficient wireless sensor networks.* Ad Hoc Networks, 4(1), pp. 75-82, online at http://eprints.lancs.ac.uk/27224/.

[43] T. Cover and J. Thomas, *Elements of Information Theory*, Wiley Series in Telecommunications, 2^{ed} ed., 2006.