A Practical Feasibility Study of a Novel Strategy for the Gaussian Half-Duplex Relay Channel

Robin Rajan Thomas, Martina Cardone, Raymond Knopp, Daniela Tuninetti, and Bodhaswar T. Maharaj

***Abstract*— This paper presents a practical feasibility study of a novel two-phase three-part-message strategy for half-duplex relaying, which features superposition coding and interference-aware cancellation decoding. Aiming to analyze the performance of the proposed scheme in the non-asymptotic regime, this paper evaluates the spectral efficiency with finite block-length and discrete constellation signaling and compares it with the theoretical performance of Gaussian codes with asymptotically large block-lengths. The performance evaluation is carried out on an LTE simulation test bench. During each transmission phase, the modulation and coding scheme is adapted to the channel link qualities to enhance the overall spectral efficiency. A single-antenna source and relay, and a multi-antenna destination are assumed. The static Gaussian and two frequency selective channel models are considered for the proposed scheme. A spectral efficiency comparison with a baseline scheme (non-cooperative two-hop transmission, i.e., the source-destination link is absent) and with the point-to-point transmission strategy (no relay) is presented. The results confirm that physical-layer cooperation and multi-antennas are critical for performance enhancement in heterogeneous networks. Moreover, they show that physical layer cooperation advantages are within practical reach with existing LTE coded-modulation and interference-mitigation techniques, which are prevalent in modern user-equipment.**

***Index Terms*— Decode-and-forward, half-duplex relay, physical-layer cooperation, spectral efficiency, successive interference cancellation.**

This work was supported in part by the Eurecom's Industrial Partners: BMW Group Research & Technology, IABG, Monaco Telecom, Orange, SAP, ST Microelectronics, and Symantec. It was also partially funded by the EU Celtic+ Framework Program Project SHAR-ING. Other partially funded sources include the National Research Foun-dation (NRF) in South Africa and the National Science Foundation. The work of R. R. Thomas was supported in part by NRF under Grant 89803, and in part by the National Research Foundation (NRF) in South Africa. The work of M. Cardone was supported by NSF under Grant 1314937 and Grant 1514531. The work of D. Tuninetti was supported by NSF under Award 1218635. This paper was presented at the 2015 IEEE International Conference on Communications.

R. R. Thomas and R. Knopp are with the Mobile Communications Department (Communications Systems Department), Eurecom, 06410 Biot, France (e-mail: thomas@eurecom.fr; knopp@eurecom.fr).

M. Cardone is with the Electrical Engineering Department, Henri Samueli School, UCLA, Los Angeles, CA 90095 USA (e-mail: martina.cardone@ucla.edu).

D. Tuninetti is with the Electrical and Computer Engineering Department, The University of Illinois at Chicago, Chicago, IL 60607 USA (e-mail: danielat@uic.edu).

B. T. Maharaj is with the Electrical, Electronic and Computer Engineering Department, University of Pretoria, 0002 Pretoria, South Africa (e-mail: sunil.maharaj@up.ac.za).

I. Introduction

The benefits of cooperative communications, as a means to enable single-antenna terminals to cooperatively operate with efficiency and diversity gains usually reserved to multi-antenna systems, have been extensively studied [1]. The different cooperative communication tech-niques and relay strategies available in the literature are largely based on the seminal information theoretic work by Cover and El Gamal [2]. These advances have led to technical studies of practical relay architectures by 3GPP for inclusion in the LTE Release 9 standard [3], [4], which then led to the incorporation of physical layer relaying node operations in the LTE Release 10 (LTE-Advanced) specifications [5].

In this paper we provide a practical LTE-based implementation [6] of the novel two-phase three-part-message relay strategy proposed in [7] for the Gaussian Half-Duplex Relay Channel (HD-RC) that is known to be to within a constant gap of the cut-set upper bound on the capacity of the network. The two-phase three-part-message scheme proposed in [7] employs superposition of Gaussian codebooks at the source, Successive Interference Cancellation (SIC) both at the relay and at the destination, and Decode-and-Forward (DF) at the relay. The channel model has a direct link between the source and the destination, through which the source continuously sends information to the destination at a rate close to the capacity of that link. At the same time, the source leverages the relay to convey extra information to the destination at a rate that, in the high-SNR regime, can be interpreted as the minimum capacity of the source-relay (for the relay-listen phase) and relay-destination (for the relay-send phase) links minus the capacity of the source-destination link. The relative duration of the relay-listen and relay-send phases is determined so that the amount of information decoded in the former can be reliably conveyed in the latter. In this work, we first extend the model of [7] to the case of multi-antenna destination and then propose a practical implementation compliant with the LTE standard. A single- and a two-antenna destination as well as the Additive White Gaussian Noise (AWGN) and two LTE frequency selective channel models are considered. We bridge theory and practice by showing that low-implementation complexity and high-throughput HD relay schemes, which exploit physical layer cooperation, are within practical reach for near-future high-spectral efficiency Heterogeneous Network (HetNet) deployments. Moreover, we once again show that enabling physical-layer cooperation among nodes and using SIMO technology is of critical importance in existing and future wireless HetNets.

A. Related Work

Wireless relay networks have long been conceptualized to cooperatively form virtual antenna arrays to improve spatial diversity and serve as a medium for multi-hop communications. The implementation of the types of relay strategies in 3G and 4G cellular standards have been classified according to the Layers in the stack at which user-plane data is forwarded at the relay, i.e., either Layer 1, 2 or 3 relaying. For example, Layer 1 (physical layer) relaying is limited to the amplification of the desired signal, which is forwarded to the mobile terminal with power control supported by control signaling features; Layer 2 consists of a DF relay, which decodes and re-encodes the user data to be forwarded on Layer 2, in the absence of direct-link transmission; Layer 3 relays function in a similar fashion to Layer 2 relays, but can operate as a self-backhauling base station, which forwards the IP data packets on Layer 3 [8]–[10]. It is well known that AF relay networks are prone to noise enhancements at the relay side. Low-powered relay nodes can aid densification of existing networks in high traffic areas and are foreseen to improve the overall network performance in many aspects, such as throughput and coverage. Their employment in HetNets is crucial and this aspect has already been standardised in LTE-Advanced [5]. In the context of our proposed HD relay strategy, various practical studies have examined the implementation and performance of relay networks with respect to LTE frameworks as well as particular channel codes.

In [11], work was carried out to analyze the coverage performance of a DF relay algorithm in a typical LTE-Advanced system, while in [12] a spectral efficiency analytical comparison was made between AF and DF modes of operation at the relay station. These studies were limited to an analytical case aimed to prove the benefits in terms of coverage and throughput of existing relay types in an LTE framework, but lacked the end-to-end performance analysis of a full physical layer LTE scenario, which can be used as a reference for future practical deployments of relay networks. In contrast, in this work we seek to evaluate the spectral efficiency performance of a two-phase three-part-message relay strategy on a physical layer simulation test bench by using practical channel codes as specified in the LTE standard.

Various relay schemes using channel codes have been extensively studied under a generalized setting. Recently, in [13], the authors showed that the capacity of the line network with one HD relay (i.e., without a direct link between the source and the destination) is achieved by a discrete input, which stresses the importance of random switching (between transmission and reception phases) at the relay. Differently, in this work we use a transmission scheme that, although it does not achieve the capacity, can be implemented within the LTE standard and is provably optimal to within a constant gap, in addition to having an easy closed-form expression. In [14], a distributed turbo coding scheme was designed for a DF relay channel; this study showed diversity and coding gains, especially in low-SNR regimes. Similarly, the work proposed in [15], used Dynamic DF (DDF) and turbo codes to increase the coding gain at the relay as well as achieve full diversity at the destination in an HD relay channel. In [16], the authors investigated the performance of fountain codes in a basic cooperative relay scheme over a block fading channel. Comparisons were made between two protocols, i.e., distributed space-time coded and time-division, revealing a performance gain of the latter in terms of outage. In [17], a Reed-Solomon convolutional coded system was designed to validate a coded opportunistic AF and DF relay scheme for a quasi-static channel; the results showed that the coded opportunistic DF scheme outperforms the AF scheme thanks to the ability of decoding/re-encoding at the relay, hence eliminating any noise enhancement to which the AF scheme is susceptible. In [18], a physical layer network coding scheme for a two-part message HD satellite communication relay setting was presented, which was limited to an uncoded BPSK scenario with hard-decision decoding. Contrary to the works in [14]–[17] where no rate splitting is performed over the two time slots (phases), our relay strategy splits the message into three parts, superposes them for transmission and performs SIC with perfect channel estimation. Moreover, although these studies considered the use of practical codes to show coding and diversity gains of relay communications, the use of multiple receive antennas at the destination (which is supported by LTE), adaptive link transmission techniques (the work in [15] examines a similar adaption in terms of channel quality at the relay in a different context), and the study over frequency selective channels, are not taken into account. In contrast, our study aims to take a practical step toward implementing relays with an overall improvement in capacity.

Few studies have tackled the performance of relay strategies with respect to current LTE standards. In an LTE-Advanced HD relay study [19], the authors proposed the integration of a Quantize-Map-Forward (QMF) - a network generalization of the classical Compress-and-Forward (CF) relay strategy - for a two-relay diamond network.[1] These results paved the way for the implementation of diamond relay networks in the LTE standard. However, this study did not perform rate splitting and was limited to an AWGN channel with a single-antenna destination. Differently, in this work we consider a theoretically optimal (up to a constant gap) scheme which splits the message into three parts and also exploits direct transmission (from the source to the destination) and we assess its performance on an AWGN channel as well as on two LTE frequency channel models with a 2-antenna destination. Another example is the work performed in [20], where the Bit Error Rate (BER) performances of various memoryless and full-memory relay schemes were compared in a convolutional and LTE Turbo coded system. Contrary to [20], where the direct source-destination link is absent (i.e., no physical cooperation between the source and the relay), in this paper we exploit physical cooperation between the source and the relay as a means to enhance the system performance. A two-layered demodulation strategy was proposed in [21] to improve the DF BLock Error Rate (BLER) performance of an

[1] An N-relay diamond network is a relay network topology where the source can communicate with the destination only through N non-interfering relays.

asymmetric relay channel link using a soft-combining of the Log-Likelihood Ratios (LLRs) during the broadcast and relaying phases. Our two-phase three-part-message scheme considers the source transmitting in both phases, which differs from [21], where the source stays silent in the second phase. The authors in [22] used a software-defined radio framework, based on the open source GNU radio platform, to implement a scalar two-user Gaussian broadcast channel; their main merit was to show rate improvements from the use of SIC with respect to Time Division Multiplexing (TDM). The aforementioned scheme artificially introduces interference among user signals on the transmitter side, which are then decoded using SIC; the corresponding performance shows the benefits of superposition coding in terms of spectral efficiency. Similarly, the benefits of single- and multi-relay physical layer cooperation have been demonstrated with the GNU radio testbed [23]. The Wireless Open-Access Radio (WARP) testbed has also been used to validate the benefits of physical layer cooperation [24]. In [24], the authors compared the performance of the AF, DF and QMF relay schemes in an indoor environment using the IEEE 802.11 (WiFi) protocol. This particular empirical study exploits a single-part message HD relay network as a distributed 2×1 Multiple-Input-Single-Output (MISO) system to enhance the throughput. The work in [25] proposes a hybrid decoding approach, which represents a generalization of [24], for cooperative physical layer relaying involving the opportunistic use of QMF and DF relaying, which adapts according to the underlying network configuration. In this paper, we split the message into three parts and exploit superposition encoding and SIC decoding to enhance spectral efficiency in a single-user HD LTE relay scenario, which was not considered in [22]–[25].

B. Contributions

In this paper we provide a practical performance-based analysis of the novel theoretically optimal (up to a constant gap) two-phase three-part-message relay strategy proposed in [7]. The goal is to explore the feasibility of such a strategy using de-facto LTE coded-modulation and interference-mitigation techniques, which to the best of the authors' knowledge has not yet been investigated. The spectral efficiency performance is compared for two channel strength ratios of the source-destination link with respect to the relay-destination link. Furthermore, comparisons are made with (i) a baseline two-hop scheme operating as a DF relay network in absence of the source-destination link and (ii) a source-destination (point-to-point) direct transmission scheme. Our implementation uses channel coding libraries from the OpenAirInterface (OAI) package [26], an open-source software and hardware wireless communication experimentation platform developed in accordance with the evolving 3GPP LTE standard. Our main contributions can be summarized as follows:

1) We extend the two-phase three-part-message transmission scheme proposed in [7] to the case of multi-antenna destination. We show that this scheme achieves the cut-set upper bound on the capacity [2] to within 3.51 bits/dimension (bits/dim) independently of the values of the channel parameters and of the number of antennas at the destination. This scheme, thanks to the basic SIC relaying building blocks, is appealing for practical implementation as SIC receivers have been extensively studied for other channel models, such as MAC channels, and already implemented in practice in simple forms (in relation to LTE) [27].
2) We assess the performance of the theoretically optimal (up to a constant gap) two-phase three-part-message scheme by using practical channel codes as specified in the LTE standard and by considering an overall practically relevant system BLER value of 10^{-2}. For the static AWGN SIMO channel, i.e., when the destination is equipped with 2 antennas, we show a theoretical (with equal bandwidth allocation) and practical spectral efficiency gap of 0.28 bits/dim when the source-destination and the relay-destination links are of the same strength and of 0.67 bits/dim when the relay-destination link is 5 dB stronger than the source-destination link. These values indicate that high-throughput HD relay schemes are within practical reach for de facto 4G relays and receivers of today, which already implement turbo decoders and thus incur no additional complexity. Similarly, in the SISO case we show that the maximum spectral efficiency gap between theory and the presented implementation (with equal bandwidth allocation) is of 0.31 bits/dim when the strength of the source-destination and relay-destination links is the same and of 0.69 bits/dim when the relay-destination link is 5 dB stronger than the source-destination link.
   In addition, for the SIMO case, we observe an improved spectral efficiency and higher transmission rates compared to the SISO case. The array gain is hence exploited and reveals the benefits of the proposed HD relay strategy when employing two receive antennas at the destination.
3) We compare the rate performance of our scheme with respect to a baseline two-hop communication strategy, where the link between the source and the destination is absent, i.e., there is no physical layer cooperation between the source and the relay to convey information to the destination. In the SIMO case, the maximum difference between the proposed cooperative strategy and the baseline two-hop communication scheme is 3.39 bits/dim when the channel strength of the source-destination link is the same as the relay-destination link, and 2.88 bits/dim when the relay-destination link is 5 dB stronger than the source-destination link. Similarly, for the SISO scenario, spectral efficiency improvements of 3.15 bits/dim when the channel strength of the source-destination link is the same as the one of the relay-destination link, and of 2.97 bits/dim when the relay-destination link is 5 dB stronger than the source-destination link, are observed. These values show that the baseline two-hop communication scheme is not beneficial in terms of rate gain, especially when the channel strength of the source-destination link is the

same as the relay-destination link. We also provide a basic point-to-point comparison where a single-part message is transmitted from the source to the destination. This analysis shows that enabling physical-layer cooperation among nodes is of critical importance in today's and future wireless networks. The proposed scheme can, in fact, enable higher spectral efficiency performance deployment of Layer 1 relays in a manner that would be feasible from a business perspective, something that the standardized methods (what is referred to as the baseline two-hop strategy) are not capable of today, mainly because operators have not found a business case due to the limited spectral efficiency benefits [28]. It has already been shown that from a coverage standpoint the deployment of mid to high powered relay nodes can yield cost savings of at least 30% for the operators [29].

4) We also consider two different practical quasi-static fading channel models, namely the Extended Pedestrian A (EPA) and the Extended Typical Urban (ETU) models in order to assess the performance of the HD relay strategy in a realistic multipath environment. We found that the ETU model (root mean square delay spread of 991 ns) displays an overall average higher spectral efficiency over the EPA model (root mean square delay spread of 43 ns) [30], thanks to the higher frequency diversity of the former model (smaller coherence bandwidth of the ETU model with respect to the EPA model); the proposed scheme is hence well suited to LTE, which benefits from the resiliency of OFDM in such multipath environments. When the source-destination and the relay-destination links are of the same strength, the rate improvement of the proposed strategy over the baseline two-hop communication scheme is of 2.20 bits/dim for the EPA model and of 1.50 bits/dim for the ETU model. When the relay-destination link is 5 dB stronger than the source-destination link, the maximum improvement in rate is 1.71 bits/dim and 1.29 bits/dim for the EPA and ETU models, respectively. The rate gains are less than those in the AWGN case, but provide reasonable improvements over the existing baseline two-hop communication scheme in a more realistic channel scenario.

Our result shows that a practical implementation of HD relay techniques is possible with the modulation and coding formats already specified by the 3GPP LTE standard, and that the gap between theory and our proposed implementation is small. We believe that optimizing the resource allocation parameters (code rates, bandwidth allocation) in relation to our relay strategy and bench-marking its performance against second-order moderate block-length capacity results, in the spirit of what was initiated for the point-to-point channel in [31], may actually show the near optimality of the proposed two-phase relaying strategy.

C. Paper Organization

The rest of the paper is organized as follows. Section II introduces the system model and overviews the scheme proposed in [7], by adapting it to the case when the destination is equipped with multiple antennas. Section III presents the simulation test bench and discusses some additional design considerations that are crucial for a more realistic analysis. Section IV and Section V evaluates the BLER performance of the proposed scheme for the static AWGN and the frequency-selective fading channels, respectively. In particular, Section IV and Section V consider the single- and two-antenna destination cases and compare the achieved spectral efficiency with a baseline two-hop communication scheme that does not allow for direct link source-destination transmission. Section IV also compares the achieved spectral efficiency with the theoretical case. Finally, Section VI concludes the paper. Proofs may be found in the Appendix.

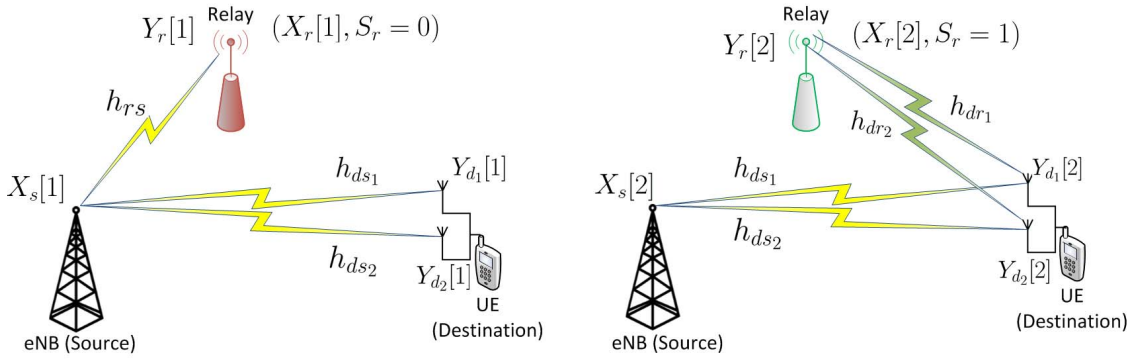
D. Notation

In the rest of the paper we use the following notation convention. With $[n_1 : n_2]$ we indicate the set of integers from n_1 to $n_2 \geq n_1$. Light-face letters indicate scalars and bold-face letters indicate vectors. Lower-case letters denote deterministic constants and upper-case letters denote random variables. We only have the following exceptions: (i) Y^j denotes a vector of length j with components $\left(Y_1, ..., Y_j\right)$; (ii) covariance matrices are indicated with a capital bold-face letter; (iii) messages are denoted with a lower-case letter. With $\mathbf{a}^H$ we indicate the Hermitian transpose of $\mathbf{a}$, with $\mathbf{a}^T$ the transpose of $\mathbf{a}$ and with a^* the complex conjugate of a. $|a|$ is the absolute value of a, $\|\mathbf{a}\|$ is the norm of the vector $\mathbf{a}$ and $|\mathbf{A}|$ is the determinant of the matrix $\mathbf{A}$; $\mathbf{I}_j$ is the identity matrix of dimension j; we use $[x]^+ := \max\{0, x\}$ for $x \in \mathbb{R}$. $X \sim \mathcal{N}\left(\mu, \sigma^2\right)$ indicates that X is a proper-complex Gaussian random variable with mean μ and variance σ^2. $\mathbb{E}[\cdot]$ indicates the expected value. Logarithms are in base 2. The spectral efficiency is measured in bits/dim and represents the number of bits per complex dimension, where a complex dimension corresponds to a channel use. Thus, bits/dim, bits/channel use and bits/s/Hz can be used interchangeably.

II. System Model and Transmission Strategy

An HD-RC consists of three nodes: the source, the relay, and the destination. The source has a uniformly distributed message $w \in \left[1 : 2^{NR}\right]$ for the destination where N denotes the codeword length and R the transmission rate. At time $i, i \in [1 : N]$, the source maps its message w into a channel input symbol $X_{s,i}(w)$ and the relay, if in transmission mode of operation, maps its past channel observations into a channel input symbol $X_{r,i}(Y_r^{i-1})$. At time N, the destination outputs an estimate of the message w based on all its channel observations Y_d^N as $\hat{w}(Y_d^N)$. A rate R is said to be ϵ-achievable if, for some block length N, there exists a code such that $\mathbb{P}[\hat{w} \neq w] \leq \epsilon$ for any $\epsilon > 0$. The capacity C is the largest nonnegative rate that is ϵ-achievable for $\epsilon \in (0, 1)$.

The SIMO Gaussian HD-RC is shown in Fig. 1, where the three nodes are the eNodeB (source), the relay, and the UE (destination), which is equipped with $n_d = 2$ antennas.



(a) Phase I- $S_r = 0$

(b) Phase II- $S_r = 1$

Fig. 1. Two-phase relay system model.

For the static case, the input/output relationship is

$$Y_r = h_{rs} X_s (1 - S_r) + Z_r \in \mathbb{C}, \tag{1a}$$

$$\mathbf{Y}_d = \mathbf{h}_r X_r S_r + \mathbf{h}_s X_s + \mathbf{Z}_d \in \mathbb{C}^2, \tag{1b}$$

where we let $\mathbf{Y}_d = \left[Y_{d_1}\ Y_{d_2}\right]^T$, $\mathbf{h}_r = \left[h_{dr1}\ h_{dr2}\right]^T$, $\mathbf{h}_s = \left[h_{ds1}\ h_{ds2}\right]^T$ and $\mathbf{Z}_d = \left[Z_{d_1}\ Z_{d_2}\right]^T$. The channel parameters $(h_{ds_i}, h_{dr_i}, h_{rs})$, $i \in [1:2]$ are fixed for the whole transmission duration (i.e., time needed to transmit one codeword) and assumed known to all nodes (i.e., full Channel State Information (CSI) - for the fading case, the CSI assumptions will be discussed in Section V), the inputs are subject to unitary power constraints, S_r is the switch random binary variable which indicates the state of the relay, i.e., when $S_r = 0$ the relay is receiving while when $S_r = 1$ the relay is transmitting, and the noises form independent proper-complex white Gaussian noise processes with zero-mean and unit-variance. The model is without loss of generality because non-unitary power constraints or noise variances can be incorporated into the channel gains. It is also worth noting that the scheme designed in this section as well as its derived performance guarantee hold for any value of $n_d \geq 1$. However, we will here focus on $n_d = 2$ as this is the case considered in the practical implementation. For the SIMO Gaussian HD-RC, the following is a generalization of [7, Proposition 5].

Proposition 1: For the static/non-fading SIMO Gaussian HD-RC the following rate is achievable

$$R = \log\left(1 + \|\mathbf{h}_s\|^2\right) + \frac{\underline{a}\ [\underline{b}]^+}{\underline{a} + [\underline{b}]^+}, \tag{2a}$$

$$\underline{a} := \log\left(1 + \frac{\|\mathbf{h}_r\|^2 + \|\mathbf{h}_r\|^2 \|\mathbf{h}_s\|^2 \left(1 - |\upsilon|^2\right)}{1 + \|\mathbf{h}_s\|^2}\right)$$

$$\upsilon := \frac{\mathbf{h}_s^H \mathbf{h}_r}{\|\mathbf{h}_s\| \|\mathbf{h}_r\|}, \tag{2b}$$

$$\underline{b} := \log\left(1 + \frac{|h_{rs}|^2}{1 + \|\mathbf{h}_s\|^2}\right) - \log\left(1 + \frac{\|\mathbf{h}_s\|^2}{1 + \|\mathbf{h}_s\|^2}\right). \tag{2c}$$

Moreover, R in (2a) *is to within* 3.51 *bits/dim from the cut-set upper bound to the capacity, irrespectively of the number of antennas at the destination and of the channel gains.*

Proof: We give next a sketch of the proof of Proposition 1. The complete proof can be derived by obvious modifications from the proof of [7, Proposition 5].

Codebooks: We study a scheme with the following four unit average power Gaussian codebooks to transmit a message that is rate-split into three parts:

$$\begin{aligned}
\mathcal{C}_{a1} &= \{X_{a1}^{N_1}(w_0) : w_0 \in [1 : M_0]\}, \\
\mathcal{C}_{a2} &= \{X_{a2}^{N_2}(w_0) : w_0 \in [1 : M_0]\}, \\
\mathcal{C}_b &= \{X_b^{N_1}(w_1) : w_1 \in [1 : M_1]\}, \\
\mathcal{C}_c &= \{X_c^{N_2}(w_2) : w_2 \in [1 : M_2]\},
\end{aligned}$$

by which we aim to achieve a rate of

$$\begin{aligned}
R &= R_a + R_b + R_c, \\
&= \frac{\log(M_0)}{N_1 + N_2} + \frac{\log(M_1)}{N_1 + N_2} + \frac{\log(M_2)}{N_1 + N_2}, \\
&= \frac{\log\left(M_0 M_1 M_2\right)}{N_1 + N_2}\ \text{[bits/dim]}.
\end{aligned}$$

The transmission is divided into two phases: the first phase (i.e., relay receiving) lasts N_1 channel uses, and the second phase (i.e., relay transmitting) lasts N_2 channel uses. The overall transmission duration for a codeword is $N = N_1 + N_2$ channel uses. In the following, in order to simplify the notation, we omit the length N_1 and N_2 in the superscript of the codewords.

Phase I: During this phase the relay is listening, i.e., $S_r = 0$ (see Fig. 1(a)). The source selects uniformly at random $w_0 \in [1 : M_0]$ (sent cooperatively with the relay to the destination) and $w_1 \in [1 : M_1]$ (sent directly to the destination). The transmitted signals are

$$X_s[1] = \sqrt{1 - \delta} X_b(w_1) + \sqrt{\delta} X_{a1}(w_0), \tag{3a}$$

$$X_r[1] = 0, \tag{3b}$$

$$\delta = \frac{1}{1 + \|\mathbf{h}_s\|^2}, \tag{3c}$$

where $\delta \in [0, 1]$ is the scaling parameter that allows for superposition coding; it is chosen in such a way that the message that is treated as noise at the destination is received at most at the power of the noise as in [7].

The relay applies successive decoding of $X_b(w_1)$ followed by $X_{a1}(w_0)$ from

$$Y_r[1] = h_{rs}\sqrt{1-\delta}X_b(w_1) + h_{rs}\sqrt{\delta}X_{a1}(w_0) + Z_r[1],$$

which is possible with arbitrary high reliability for sufficiently large N if

$$R_b \leq \gamma \log\left(1+|h_{rs}|^2\right) - \gamma \log\left(1+\frac{|h_{rs}|^2}{1+\|\mathbf{h}_s\|^2}\right),$$

$$R_a \leq \gamma \log\left(1+\frac{|h_{rs}|^2}{1+\|\mathbf{h}_s\|^2}\right), \quad (4)$$

where $\gamma = \frac{N_1}{N_1+N_2}$. The destination, by using Maximal-Ratio Combining (MRC), decodes $X_b(w_1)$ by treating $X_{a1}(w_0)$ as noise from

$$\mathbf{Y}_d[1] = \mathbf{h}_s\sqrt{1-\delta}X_b(w_1) + \mathbf{h}_s\sqrt{\delta}X_{a1}(w_0) + \mathbf{Z}_d[1],$$

which is possible with arbitrary high reliability for sufficiently large N if

$$R_b \leq \gamma \log\left(1+\|\mathbf{h}_s\|^2\right) - \gamma \log\left(1+\frac{\|\mathbf{h}_s\|^2}{1+\|\mathbf{h}_s\|^2}\right). \quad (5)$$

In order to obtain R_b in (5) we computed $\log\left(1+(1-\delta)\,\mathbf{h}_s^H\,\Sigma_1^{-1}\mathbf{h}_s\right)$ where $\Sigma_1 \in \mathbb{C}^{2\times 2}$ is the covariance matrix of the equivalent noise

$$\tilde{\mathbf{Z}}_1 = \mathbf{h}_s\sqrt{\delta}X_{a1}\,(w_0) + \mathbf{Z}_d[1],$$

and where the final expression follows as an application of the matrix inversion lemma.[2] Notice that the same rate is obtained if the receiver computes the scalar $\frac{\mathbf{h}_s^H}{\|\mathbf{h}_s\|}\mathbf{Y}_d[1] = \|\mathbf{h}_s\|\sqrt{1-\delta}X_b(w_1)+\|\mathbf{h}_s\|\sqrt{\delta}X_{a1}(w_0)+Z$, $Z\sim\mathcal{N}(0,1)$ and decodes X_b. Finally, by assuming $\|\mathbf{h}_s\|^2 < |h_{rs}|^2$ (we will see later that this assumption is without loss of generality), Phase I is successful if (4) and (5) are satisfied.

Phase II: During this phase the relay is transmitting, i.e., $S_r = 1$ (see Fig. 1(b)). The source selects uniformly at random $w_2 \in [1 : M_2]$ (sent directly to the destination) and the relay forwards its estimation of w_0 from Phase I, indicated as $\widehat{w}_0$. The transmitted signals are

$$X_s[2] = X_c(w_2),$$

$$X_r[2] = X_{a2}\,(\widehat{w}_0)\,.$$

The destination uses MRC and applies successive decoding based on

$$\mathbf{Y}_d[2] = \mathbf{h}_s\,X_c(w_2) + \mathbf{h}_r\,X_{a2}\,(\widehat{w}_0) + \mathbf{Z}_d[2].$$

In particular, it first decodes w_0 by using the received signal from both phases and by assuming that $\widehat{w}_0 = w_0$; this is true given the rate constraints found for Phase I. It then decodes $X_c(w_2)$, after having subtracted the contribution of its estimated w_0. Successful decoding is possible with arbitrary high reliability for sufficiently large N if

$$R_a \leq (1-\gamma)\log\left(1+\frac{\|\mathbf{h}_r\|^2+\|\mathbf{h}_r\|^2\|\mathbf{h}_s\|^2\left(1-|\upsilon|^2\right)}{1+\|\mathbf{h}_s\|^2}\right) + \gamma\log\left(1+\frac{\|\mathbf{h}_s\|^2}{1+\|\mathbf{h}_s\|^2}\right), \quad (6)$$

$$R_c \leq (1-\gamma)\log\left(1+\|\mathbf{h}_s\|^2\right), \quad (7)$$

where υ is defined in (2b). In order to obtain R_a in (6) we computed $\log\left(1+\mathbf{h}_r^H\,\Sigma_2^{-1}\mathbf{h}_r\right)$, where $\Sigma_2 \in \mathbb{C}^{2\times 2}$ is the covariance matrix of

$$\tilde{\mathbf{Z}}_2 = \mathbf{h}_s X_c\,(w_2) + \mathbf{Z}_d[2],$$

and where the final expression follows as an application of the matrix inversion lemma. By imposing that the rate R_a is the same in both phases, that is, that (4) and (6) are equal, we get that γ should be chosen equal to γ^*

$$\gamma^* = \frac{\underline{a}}{\underline{a}+\underline{b}}, \quad (8)$$

where $\underline{a}$ and $\underline{b}$ are defined in (2b) and (2c), respectively. Note that because of the assumption $\|\mathbf{h}_s\|^2 < |h_{rs}|^2$, we have $\underline{b} > 0$, i.e., $\underline{b} = [\underline{b}]^+$.

The rate sent directly from the source to the destination, that is, the sum of (5) and (7), is

$$R_b + R_c = \log\left(1+\|\mathbf{h}_s\|^2\right) - \underbrace{\gamma^*\log\left(1+\frac{\|\mathbf{h}_s\|^2}{1+\|\mathbf{h}_s\|^2}\right)}_{\in[0,1]}. \quad (9)$$

Therefore the total rate after the two phases of transmission is $R = R_b + R_c + R_a$ as in (2a), which implies

$$\mathsf{C} \geq R_b + R_c + R_a \geq \log(1+\|\mathbf{h}_s\|^2) + \frac{\underline{a}\;\underline{b}}{\underline{a}+\underline{b}}. \quad (10)$$

The rate expression for R in (2a), with $[b]^+$ rather than b, holds since for $\|\mathbf{h}_s\|^2 \geq |h_{rs}|^2$ it reduces to a direct transmission from the source to the destination. Note that the single-antenna result in [7, Proposition 5] is obtained as a special case of Proposition 1 by setting $\upsilon = 1$.

Gap Analysis: In Appendix we show that the scheme is optimal to within 3.51 bits/dim, irrespectively of the number of antennas at the destination and of the channel gains. ■

Next we propose a practical LTE-based implementation to achieve the rate in Proposition 1.

III. Simulation Test-Bench Design

The scheme described in Section II uses four Gaussian codebooks $\mathcal{C}_{a1}$, $\mathcal{C}_{a2}$, $\mathcal{C}_b$ and $\mathcal{C}_c$ to transmit the three sub-messages w_0, w_1 and w_2. In a practical implementation, the codes would not be Gaussian, but will contain symbols from a finite constellation. Since SIC is employed in the decoding operations both at the relay and at the destination, we need to understand the performance of these practical codes both in Gaussian and non-Gaussian noise. In particular, at the relay we need to understand the performance of $\mathcal{C}_b$ in non-Gaussian

[2] Matrix inversion lemma $(A + XBX^T)^{-1} = A^{-1} - A^{-1}X(B^{-1} + X^TA^{-1}X)^{-1}X^TA^{-1}$.

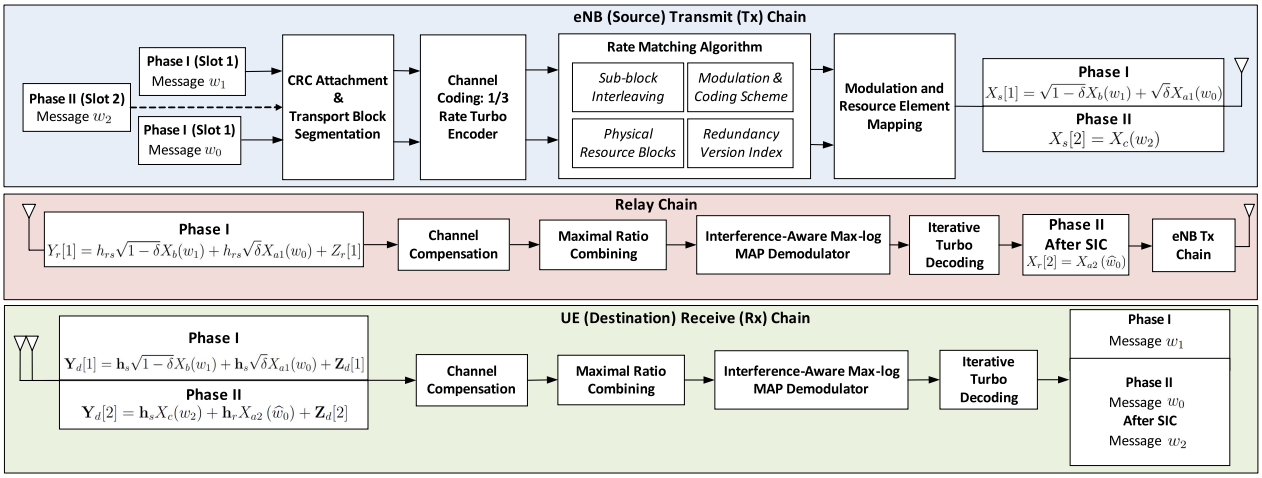


Fig. 2. Simulation block diagram for the eNB, relay and UE.

noise (first decoding step of Phase I) and of $\mathcal{C}_{a1}$ in Gaussian noise (second decoding step of Phase I); at the destination we have to understand the performance of $\mathcal{C}_{a1}$, $\mathcal{C}_{a2}$ and $\mathcal{C}_b$ in non-Gaussian noise ($\mathcal{C}_b$ in the first decoding step in Phase I and $\mathcal{C}_{a1}$, $\mathcal{C}_{a2}$ in the first decoding step in Phase II) and the performance of $\mathcal{C}_c$ in Gaussian noise (second decoding operation in Phase II when no error propagation). In the decoding stages where a message is treated as noise, we implement a demodulator that specially accounts for the fact that the overall noise is non-Gaussian. We consider different choices for the codebooks ($\mathcal{C}_{a1}$, $\mathcal{C}_{a2}$, $\mathcal{C}_b$, $\mathcal{C}_c$); for each choice, we ensure a BLER at the destination below a given threshold for all the decoding stages (here we seek to maintain the overall system BLER at the destination below 10^{-2} and hence the BLER corresponding to each of the three decoding operations is set to be at most $1/3 \cdot 10^{-2}$). Although the analysis is focused on the case when the destination is equipped with two antennas, the single-antenna case is also presented for completeness. The key question we seek to answer is the following: how does the spectral efficiency of practical codes compare to the theoretical performance in Proposition 1?

We develop a simulation testbed using the OAI (a platform for wireless communication experimentation) software libraries in order to evaluate the performance of the aforementioned scheme with practical codes (see Fig. 2). The software platform is based on 3GPP's evolving standard of LTE which consists of the essential features of a practical radio communication system, which closely align with the standards in commercially deployed networks. Fig. 2 shows the key functional units of the simulation design. The simulations were carried out on the Downlink Shared Channel (DL-SCH), which is the primary channel for transmitting user-data (or control information) from the eNodeB (eNB) to the UE [30]. The data messages are transported in units known as Transport Blocks (TBs) to convey the sub-messages w_0, w_1 and w_2. The TB Size (TBS) depends on the choice of the Modulation and Coding Scheme (MCS), which describes the modulation order and the coding rate of a particular transmission.

Processing: The TBs undergo a series of processing stages prior to modulation before the codeword can be mapped into the Resource Elements (REs) in the Physical DL-SCH (PDSCH). Error detection at the receiver is enabled by appending 24 Cyclic Redundancy Check (CRC) bits to the TB. The code block (comprising the TB and the CRC bits) has minimum and maximum sizes of 40 and 6144 bits, respectively, as required by the Turbo encoder. Filler bits are added if the code block is too small in size and the code block is segmented into blocks of smaller size if the maximum size is exceeded. The subsequent bit sequence is then fed into the $1/3$ rate Turbo encoder.

Channel Coding: The channel coding scheme comprises a 1/3 rate Turbo encoder, which follows the structure of a parallel concatenated convolutional code with two 8-state constituent encoders, and one Turbo code internal interleaver [32]. A single set of systematic bits and two sets of parity bits are produced at the output of the encoder as detailed in [32].

Rate Matching: The rate matching component ensures, through puncturing or repetition of the bits, that the output bits of the Turbo encoder match the available physical resources using the MCS, the Redundancy Version (RV) index and the Physical Resource Blocks (PRBs). For the numerical evaluations, an equal bandwidth allocation is chosen between the two phases of the relay strategy, i.e., the number of PRBs allocated in the first (relay listening) and second (relay transmitting) phases is the same. In other words, with reference to (8), we set $\gamma = 0.5$, which may not be the optimal choice. The sub-message w_0, transmitted by the source and the relay over the two phases, corresponds to two different RV indices with equal resource allocations. The selection is made possible through puncturing or repetition of the bits at the output of the encoder. The Circular Buffer (CB) generates puncturing patterns depending on the allocated resources, and the sub-block interleaver (which forms part of the CB) facilitates the puncturing of the three outputs of the encoder [30]. Furthermore, the code block is concatenated if segmentation was required prior to channel coding.

Modulation and REs Mapping: During this stage, complex-valued symbols are generated according to the chosen modulation scheme, i.e., QPSK, 16-QAM or 64-QAM, which are supported in LTE. In this study, we seek to investigate the performance of QPSK as well as higher-order modulations

such as 16-QAM and 64-QAM. In particular, we employ the same modulation order at the source and at the relay. Such a scheme performs well when the channel quality between the relay and the destination is not much better than that of the source-destination link. However, when the relay-destination link is significantly stronger than the source-destination link, a better performance/higher spectral efficiency could be attained with higher modulation schemes at the relay.

Channel Compensation and MRC: The channel compensation block is responsible for computing the Matched Filtered (MF) outputs and the effective channel magnitudes of the received signal. These parameters are required for the soft-decoding of the desired message using the interference-aware demodulator. The MRC block utilizes the MF outputs to constructively add the two received signals to maximize the post-processing SNR (notice that the MRC block is not needed in the case when the destination is equipped with a single-antenna).

Interference-Aware Demodulator: The demodulator comprises of a discrete constellation interference-aware receiver designed to be a low-complexity version of the max-log MAP detector. The main idea is to decouple the real and imaginary components through a simplified bit-metric using the MF output and thus reduce the search space by one complex dimension [33]–[35]. As a result, it is possible to decode the required codeword in the presence of an interfering codeword of the same (or different) modulation scheme. Thereafter, it is possible to strip out the decoded signal from the received signal and then decode the remaining signal in an interference-free channel in the case of no error propagation. The sub-message w_0 is decoded at the destination at the end of Phase II, i.e., X_{a1} (received in Phase I) and X_{a2} (received in Phase II) are combined to obtain the sub-message w_0.

A. Further Design Considerations for More Realistic Analysis

From a purely practical standpoint, two main issues related to low-layer signaling need to be addressed, in order for the proposed HD relay scheme to be viable for future (e.g., 5G) standardization within the LTE framework. The first pertains to timing advance of the relay transmission. In a conventional OFDM system, the receiver synchronizes its reception to the start of the OFDM symbols and this synchronization must be accurate enough so that the entire channel response falls within the duration of the cyclic-prefix. Otherwise, inter-symbol interference will significantly degrade the performance of the receiver. In the context of the physical-layer relaying scheme, the received signal in the second phase of the protocol is jointly transmitted from the eNB and the relay station. These two signal components must arrive more or less synchronized in order for the receiver to be able to receive the OFDM waveform without much difficulty. This clearly requires some form of timing adjustment since the path from the source to the destination going through the relay is longer than the direct path. The relay will adjust its timing to the received signal from the eNB (Phase I) and must advance the signal for each UE that it is serving according to the difference in path delay between the direct path and the path going through it to the destination. This will clearly require additional signaling. The second main issue is related to signaling in support of H-ARQ. In the first phase of the protocol, separate ACK/NAK signals need to be conveyed to the source from the relay (for two messages) and from the destination (for one message). Similarly, in the second phase, the destination must send two ACK/NAK signals to the source and one to the relay. This will clearly require significant modifications in low-layer signaling, but a feasible solution is well within reach. These design considerations represent an interesting research direction, which is currently under investigation.

IV. Performance Evaluation in Static AWGN Channel

In this section we numerically evaluate the performance of the proposed transmission strategy over the static/no-fading AWGN channel model. The transmission bandwidth of the simulated system is 5 MHz, corresponding to 25 PRBs. In the simulations, we considered square QAM constellations with *unit* energy. For example, the signal $X_s[1]$ transmitted by the source in Phase I (see (3a)) is a combination of $X_b(w_1)$ and $X_{a1}(w_0)$, which are both drawn from a QAM constellation with unit energy. Thereafter, they are scaled by the parameter δ, or $1-\delta$, and hence $X_s[1]$ has an average unit energy. Similarly, in Phase II the transmitted signals $X_s[2]$ and $X_r[2]$ are drawn from unit energy QAM constellations and hence, by definition, they have unit energy.

With reference to the system model in (1), we let $|h_{rs}|^2 = C$, $|h_{ds_1}|^2 = |h_{ds_2}|^2 = S$ (i.e., the two source-destination links are of the same strength), $|h_{dr_1}|^2 = |h_{dr_2}|^2 = I$ (i.e., the two relay-destination links are of the same strength), and we set the phase of the channel gains to some random value that is kept constant during the whole transmission duration (i.e., $N_1 + N_2$ channel uses). Perfect receive CSI is assumed at all nodes. For each of the decoding operations during Phases I and II, the BLER performances at the relay and destination are validated for different values of C (at the relay) and evaluated with respect to S (channel quality for the source-destination link). Furthermore, we use $\delta = \frac{1}{1+2S}$ as the superposition parameter in the SIMO scenario. An equal bandwidth allocation between the two phases of the strategy is also assumed, i.e., $\gamma = 0.5$.

Decoding at the Relay: At the end of Phase I, the relay first decodes w_1, then it strips it out from its received signal and finally decodes w_0, in an interference-free link. Error propagation, which results from feeding back incorrectly decoded symbols, is not considered here. The reason is that we consider a coded system with CRC bits (that is implemented in the 4G receivers of today) for which the residual error (misdetection) at the output occurs with a very low probability (e.g., 10^{-9}). The source-relay channel is assumed to be strong enough to guarantee a system BLER that is less than 10^{-3} at the relay (e.g., see the third and fourth columns of Table I). In other words, in order to have successful decoding operations (of both w_0 and w_1) at the relay at the end of Phase I, we assume a value of C such that BLER $\leq$ BLER$[X_b]$ + BLER$[X_{a1}] \leq 5 \cdot 10^{-4} + 5 \cdot 10^{-4} = 10^{-3}$. Under this assumption, it is worth noting that the relay successfully

TABLE I

MCS MAPPING FOR EACH DECODING OPERATION WITH $I/S = 0$ dB FOR A SIMO SCHEME

Phase I - X_b					**Phase II - (X_{a1}, X_{a2})**		**Phase II - X_c**		**Theoret-ical**	**Practical**	**Theoret-ical BS**	**Practical BS**
MCS	S [dB]	C_{X_b} [dB]	$C_{X_{a1}}$ [dB]	TBS [bits]	MCS	TBS [bits]	MCS	TBS [bits]	Rate [bits/dim]	Rate [bits/dim]	Rate [bits/dim]	Rate [bits/dim]
9	1.89	5.48	9.54	4008	11	4968	13	5736	2.44	2.26	1.01	1.12
14	5.84	9.07	12.10	6456	14	6456	17	7736	3.41	3.36	1.56	1.46
20	11.02	14.60	18.2	9912	20	9912	22	11448	5.10	5.09	2.36	1.93

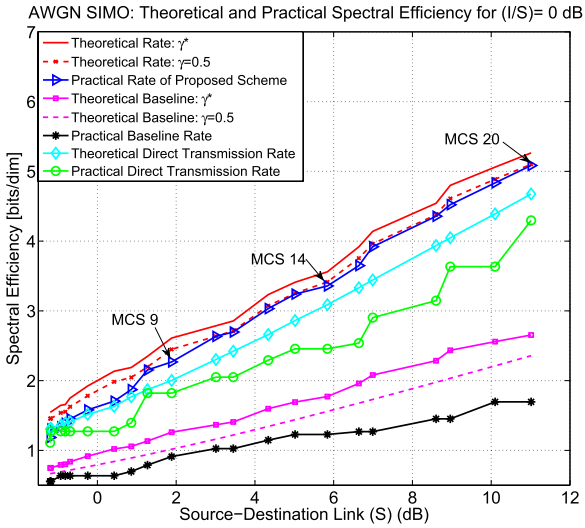


Fig. 3. SIMO spectral efficiency performance when (I/S) = 0 dB.

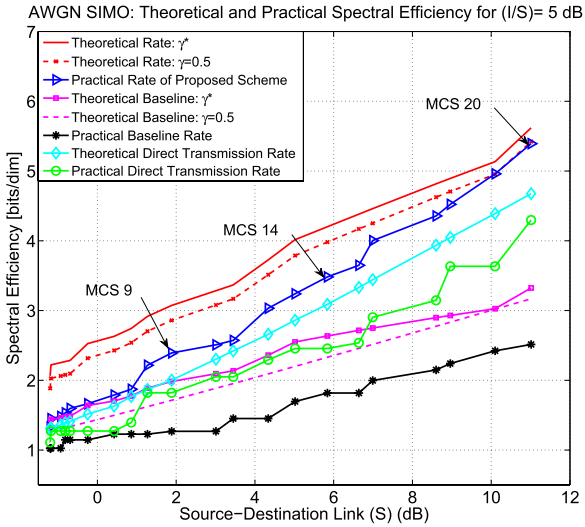


Fig. 4. SIMO spectral efficiency performance when (I/S) = 5 dB.

decodes the message w_0 with BLER$[X_{a1}] \leq 5 \cdot 10^{-4}$, which is a more stringent constraint than the one imposed at the destination of BLER$[X_{a1}] \leq 1/3 \cdot 10^{-2}$. Finally, at the beginning of Phase II, the sub-message w_0 is re-encoded and forwarded by the relay to the destination.

Decoding at the Destination: The overall SIMO (two antennas at the destination) spectral efficiency of the HD relay strategy at the end of Phase II is shown in Fig. 3 and Fig. 4 (see the blue curves). Each of the markers on the blue curves - practical rates achieved by the scheme described

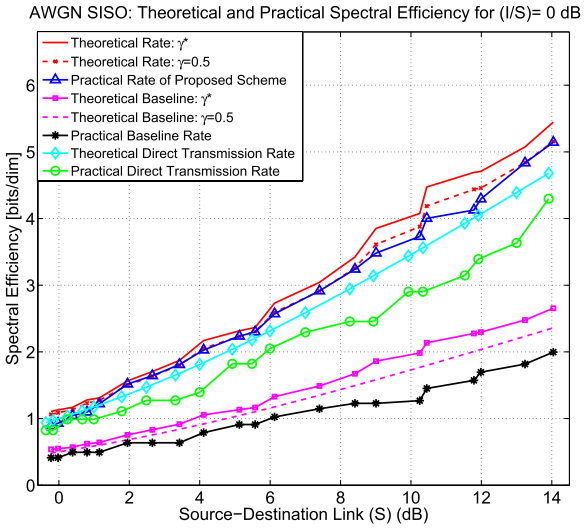


Fig. 5. SISO spectral efficiency performance when (I/S) = 0 dB.

in Section II - represents an MCS, which in total ranges from 0 to 20. Each MCS characterizes the type of signal transmission (i.e., QPSK, 16-QAM or 64-QAM) with a certain code rate. The rates are plotted for two different ratios of the relay-destination link and source-destination link, namely $I/S = 0$ dB (see Fig. 3) and $I/S = 5$ dB (see Fig. 4). This is shown to demonstrate that utilising a relay-destination link that is stronger than the source-destination link indeed boosts the rate performance with respect to direct transmission (see green curves) or conventional relay schemes (see black curves)-the baseline strategy, which will be explained later in this section. Similarly, Fig. 5 and Fig. 6 show the practical rate performance of the SISO case, i.e., when the destination is equipped with a single antenna.

Table I displays the link adaptation of the proposed strategy for the SIMO case when $I/S = 0$ dB with three different MCS values at the end of each phase; this illustrates the overall selection strategy in order to compute the practical spectral efficiencies.[3] We assumed an overall system decoding probability of 10^{-2} at the destination, i.e., BLER $\leq$ BLER$[X_b]$+BLER$[X_{a1}, X_{a2}]$+BLER$[X_c] \leq 1/3 \cdot 10^{-2} + 1/3 \cdot 10^{-2} + 1/3 \cdot 10^{-2} = 10^{-2}$. For each value of the MCS of X_b (first column of Table I), we assumed a value of S (second column of Table I), which ensures that BLER$[X_b] \leq 1/3 \cdot 10^{-2}$. Thereafter, for each value of the ratio I/S, we selected the MCS of (X_{a1}, X_{a2}) which, for each value of S

[3] Although here we report only Table I, which considers the SIMO case and $I/S = 0$ dB, a similar table was made for $I/S = 5$ dB and for the SISO case.

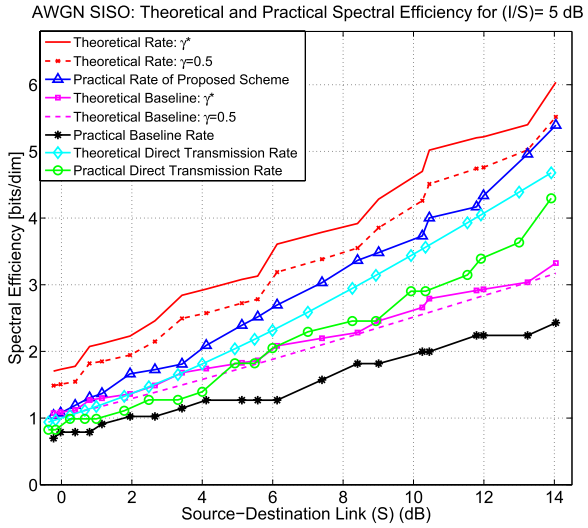


Fig. 6. SISO spectral efficiency performance when (I/S) = 5 dB.

(second column of Table I), ensures BLER$[X_{a1}, X_{a2}] \leq 1/3 \cdot 10^{-2}$. These MCS values are reported in the sixth column of Table I. Similarly, we proceeded in selecting the MCS of X_c (eighth column of Table I), so that BLER$[X_c] \leq 1/3 \cdot 10^{-2}$. The TBSs of each MCS (at the source and relay) are defined according to the 3GPP LTE standard [36] and are also reported in Table I.

Comparison With Theoretical Performance: One of our major goals in this work is to compare the theoretical and practical spectral efficiency performances of the proposed strategy. To this end, we proceeded as follows. Both for the SISO and for the SIMO cases, we drew two curves, corresponding to the theoretical achieved average (averaged over 10^4 different realizations of the phases of the channel parameters) spectral efficiency. In particular,

- Theoretical rate with γ^* (solid red curve): this curve is drawn by plugging into (2a) the values of (S, I, C) used to obtain the practical spectral efficiency. In other words, this rate is computed with the optimal γ^* defined in (8).
- Theoretical rate with $\gamma = 0.5$ (dashed red line): this curve is drawn by considering the values of (S, I, C) used to obtain the practical spectral efficiency and by computing

$$R = \mathbb{E}\left[\min\{\text{eq.(4)}, \text{eq.(6)}\}\right] + \text{eq.(7)} + \text{eq.(5)},$$

  with $\gamma = 0.5$. Hence, this rate is computed with the value of γ used in the simulations.

Clearly, as we also observe from Fig. 3, Fig. 4, Fig. 5 and Fig. 6, the theoretical rate with the optimal γ is no less than the theoretical rate with $\gamma = 0.5$. This is because in the latter case we are fixing $\gamma = 0.5$, which might not be optimal.

The spectral efficiency of our practical scheme - blue curve in Fig. 3, Fig. 4, Fig. 5 and Fig. 6 - (see for example the eleventh column of Table I for the SIMO case with $I/S = 0$ dB) was determined by using the ratio of the TBS (useful message length) with respect to the number of soft-bits (G-codeword size) together with the modulation order, which does not include the overhead bits such as the cyclic prefix, pilots and control channel information (PDCCH symbols). In particular,

$$R = \frac{\text{TBS}(X_b) + \text{TBS}\,(X_{a1}, X_{a2}) + \text{TBS}(X_c)}{\left(\frac{\text{G}_1}{\text{Q}_{\text{mod1}}} + \frac{\text{G}_2}{\text{Q}_{\text{mod2}}}\right)}, \qquad (11)$$

where G_1 is the number of soft-bits used to decode (X_b, X_{a1}) and G_2 to decode (X_{a2}, X_c), and Q_{mod1} and Q_{mod2} are the corresponding modulation orders.[4]

Simulation Analysis: For the SIMO case, from Fig. 3 and Fig. 4 we notice that the maximum difference between the theoretical rate (with $\gamma = 0.5$) and the achieved rate by the proposed scheme is 0.28 bits/dim at MCS = 6 when $I/S =$ 0 dB and 0.67 bits/dim at MCS = 7 when $I/S = 5$ dB. Similarly, for the SISO case, the maximum difference between the theoretical rate (with $\gamma = 0.5$) and the achieved rate by the proposed scheme is 0.31 bits/dim at MCS = 17 when $I/S = 0$ dB (see Fig. 5) and 0.69 bits/dim at MCS = 7 when $I/S = 5$ dB (see Fig. 6). These values indicate that high-throughput HD relay schemes are within practical reach for de facto 4G relays and receivers of today. These rate gaps between theory and practice can be mostly attributed to two key factors: (i) the TBs used are of finite length, differently from the theoretical assumption of an infinite block length and (ii) the coding and modulation schemes were chosen to match existing standards rather than being optimized for this setup. Furthermore, the fact that the difference when $I/S = 5$ dB is larger than when $I/S = 0$ dB is due to the fact that when the ratio I/S increases it becomes more critical to choose higher MCS values for the relay (with respect to the source) in order to fully exploit the strength of the relay-destination link. Adapting the modulation order and the number of PRBs across different rounds may reduce the gap between the theoretical and practical performances. Moreover, we also remark that the difference between theoretical and practical rates might be decreased by performing an optimization of the parameters δ (superposition factor) and γ (fraction of time the relay listens to the channel) in the interval $[0, 1]$, instead of considering them as fixed values (which has been deemed out of the scope of this study).

A. Baseline Relay Scheme

For comparisons with existing relay structures, we also considered a baseline two-hop communication scheme (BS), which mimics the relay structure of today's LTE networks, where the UE does not have a direct connection with the eNB, i.e., the source-destination link is absent and the eNB can communicate with the UE only through the relay (no physical layer cooperation). For both the SIMO and the SISO scenarios, the BS practical rate (black curve) is compared to the theoretical one (magenta dashed curve) in Fig. 3, Fig. 4, Fig. 5 and Fig. 6. The theoretical BS achievable rate with $\gamma = 0.5$ (twelfth column of Table I) is given by:

$$R_{\text{Th-BS}} = \min\{\gamma \log(1 + C), (1 - \gamma) \log(1 + \alpha I)\}, \qquad (12)$$

with $\alpha = 2$ in the SIMO case and $\alpha = 1$ in the SISO case. The optimal theoretical γ^* is obtained by equating the two

[4]TBS is the number of information bits, G is the number of coded bits and $\text{Q} \equiv \log M$ where M is the modulation order.

terms within the min in (12), i.e.,

$$\gamma^* = \frac{\log(1+\alpha I)}{\log(1+\alpha I)+\log(1+C)}. \quad (13)$$

Similar to how we did for the proposed three-part-message scheme, also for the baseline two-hop communication strategy, we plotted both the curve when $R_{\text{Th}-\text{BS}}$ in (12) is evaluated in γ^* in (13) (solid magenta curve) and the curve when $R_{\text{Th}-\text{BS}}$ in (12) is evaluated in $\gamma = 0.5$ (dashed magenta curve). Similar comments as those drawn for the proposed scheme also hold in this case. In the SIMO case, as we observe from Fig. 3 and Fig. 4, the maximum difference in spectral efficiency between the practical strategy and the practical BS rates (thirteenth column of Table I) is of 3.39 bits/dim for $I/S = 0$ dB, and of 2.88 bits/dim for $I/S = 5$ dB, which is a significant improvement in spectral efficiency of the cooperative relay strategy over the currently employed DF basic relay scheme. In the SISO case, we also notice that the maximum difference between the practical rate and the practical BS rate is of 3.15 bits/dim and of 2.97 bits/dim, respectively. The fact that the difference is larger when $I/S = 0$ dB than when $I/S = 5$ dB is due to the fact that when the ratio I/S is small, the presence of the source-destination link plays a significant role in enhancing the rate performance. Moreover, we notice that the curve representing the spectral efficiency of the scheme described in Section II is almost twice the one of the baseline strategy. This means that the scheme here proposed, not only provides an array gain with respect to the baseline strategy, but also a gDoF gain (pre-log factor). This can be explained as follows. The values of (S, I, C) here considered are more or less comparable; this implies, that in the high-SNR regime $R_{\text{Th}-\text{BS}}$ in (12) becomes $R_{\text{Th}-\text{BS}} \approx \frac{1}{2}\log(1+\mathsf{SNR})$, while R in (2) becomes $R \approx \log(1+\mathsf{SNR})$, i.e., $R \approx 2R_{\text{Th}-\text{BS}}$.

B. Direct Transmission Scheme (No Relay)

The theoretical (cyan curve) and practical (green curve) rates for direct communication between the source and the destination transmitting a single-part-message (basic point-to-point scenario), are also provided in Figs. 3, 4, 5 and 6. According to Figs. 3 and 4, the maximum difference in spectral efficiency between the three-part-message practical scheme and the single-part-message practical point-to-point scheme is of 1.21 bits/dim for $I/S = 0$ dB, and of 1.33 bits/dim for $I/S = 5$ dB. In the SISO case, we also observe a maximum difference of 1.20 bits/dim ($I/S = 0$ dB) and 1.33 bits/dim ($I/S = 5$ dB). These values show that the proposed three-part-message scheme has a higher spectral efficiency when compared to the direct transmission scheme, while maintaining the same slope. In particular, in order to observe a gDoF gain (in addition to the increased power efficiency) of our three-part-message scheme with respect to the point-to-point transmission, we could consider $I = S^\alpha$ and $C = S^\beta$ with $\min\{\alpha, \beta\} > 1$, which would be interesting from an information theoretic perspective (similar to [7]), in order to understand the asymptotics of the proposed scheme. However, when considering practical codes, we would eventually operate outside the practical SNR regime of the LTE codebooks. For example, if $\alpha = 2$, S would be limited to around 10 dB at the point where the links to and from the relay would be at the highest spectral efficiency. In this case, MCS formats beyond 6 bits/dim (64-QAM) would be required on the source-relay and relay-destination links, which are not part of the performance evaluation tools at our disposal. Our numerical evaluations show that the three-part-message scheme brings rate advantages (with respect to the direct transmission scheme) in the finite-SNR regime, thus not requiring the use of asymptotic performance indicators. The theoretical (cyan curve) and practical (green curve) point-to-point rates further strengthen the notion that the conventional relay scheme (as indicated by our baseline scenario) does not provide quantitative benefits, in terms of spectral efficiency, over our proposed scheme.

V. Performance Evaluation in Frequency-Selective Fading Channel

The AWGN channel modeling of the proposed relay strategy, analyzed in Section IV, represents an idealistic scenario. In an effort to model a practical scenario, we evaluate the spectral efficiency of the strategy using two well-known low mobility frequency-selective channel models defined by 3GPP, i.e., the EPA and ETU models [37]. In particular, we focus on the scenario where the destination is equipped with two antennas (SIMO). We assume channel state information at all nodes. For a more practically relevant scenario, the channel state information should be estimated/learned and this is currently part of ongoing work. The power delay profiles (amplitude and path delay) of the EPA and ETU channel models are derived from [30] and [37]. The EPA and ETU models consist of seven and nine discrete multipath components, each with a coherence bandwidth of 2.43 MHz and 0.2 MHz, respectively. The amplitude distribution for each tap in the EPA and ETU models is described by a Rayleigh fading process. The complex channel coefficients for both the source-destination (h_{ds_1} and h_{ds_2}) and relay-destination (h_{dr_1} and h_{dr_2}) links are generated according to the generalized channel transfer function (in the frequency domain) given by $h_{i,n} = \sum_{l=1}^{L} \alpha_l \exp(-j2\pi\tau_l n f_{sub})$, where i represents the symbol index, $n \in [1:300]$ represents the subcarrier index for a bandwidth of 5 MHz, α_l represents the complex path amplitude, l is the path index, L represents the total number of paths, τ_l is the path delay and f_{sub} represents the periodic subcarrier spacing of 15 kHz (as specified in LTE) [30]. Furthermore, we assume a zero Doppler shift for both channel models in line with the low-mobility assumption of the destination (UE). The link between the source and the relay is modeled to be a static AWGN channel with the aim of deploying a relay in a location with a high link quality from the source. An analysis of the proposed scheme over a fading channel would involve an evaluation of the achievable rate under a given outage probability, which would hold for the infinite block length and block-fading channel case. This would be performed by extending (2) to a vector channel with channel coefficients governed by the statistics of the EPA and ETU models and by evaluating the achievable rates at which we obtain the desired outage probability. The aforementioned analysis presents an

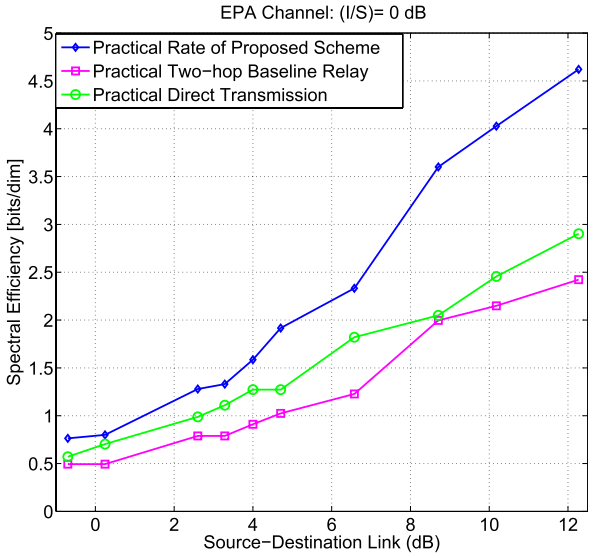


Fig. 7. EPA practical rate comparisons for (I/S) = 0 dB.

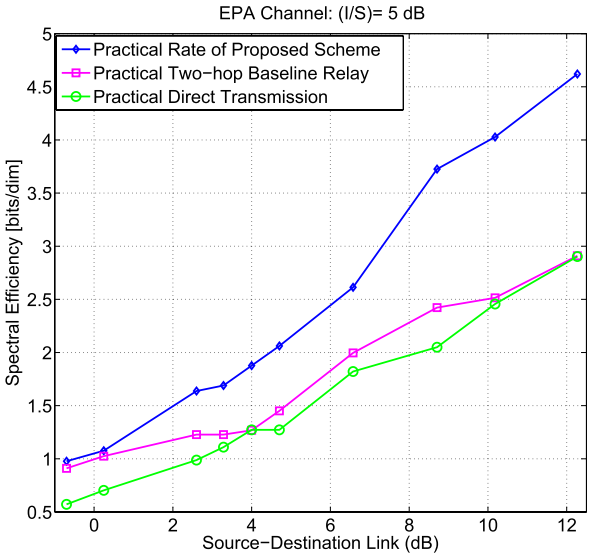


Fig. 8. EPA practical rate comparisons for (I/S) = 5 dB.

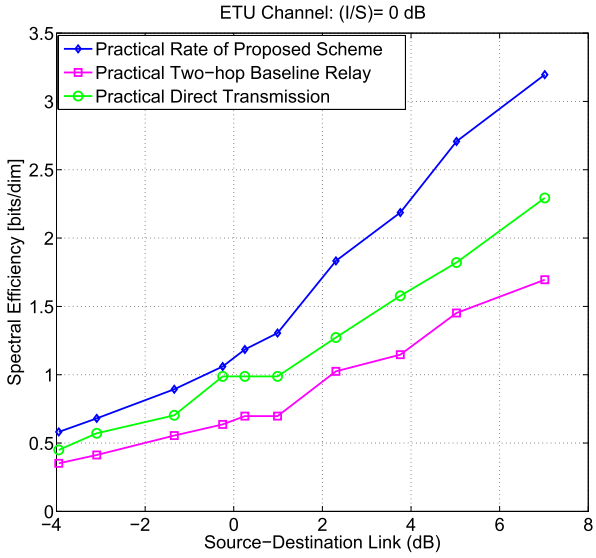


Fig. 9. ETU practical rate comparisons for (I/S) = 0 dB.

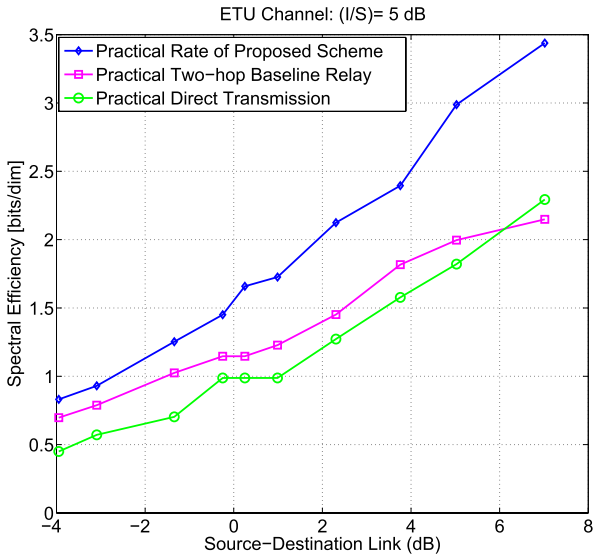


Fig. 10. ETU practical rate comparisons for (I/S) = 5 dB.

interesting challenge for future research directions. The major goal of the comparison carried out in the rest of this section is to show the rate advantage of our scheme in realistic channel models with respect to the existing baseline and point-to-point transmission schemes.

The analysis is as for the AWGN case in Section IV, except for the following. Due to the fact that the system was not designed and optimized for these frequency selective channels, it was a challenge to guarantee an overall system BLER of 10^{-2} at the destination at reasonable SNR values for higher MCSs, with an interfering codeword from the same discrete constellation. Hence, it was decided to relax the BLER constraint such that BLER $\leq$ BLER$[X_b]$+BLER$[X_{a1}, X_{a2}]$+ BLER$[X_c] \leq 10^{-1}$, only considering MCS values corresponding to QPSK transmissions.

The results in Fig. 7 and Fig. 8 show the performance of the three-part-message relay strategy (see blue curve) using the EPA channel model for $I/S = 0$ dB and $I/S = 5$ dB, respectively. Similarly, the performance of the ETU channel model, presented in Fig. 9 and Fig. 10, has also been investigated. In the case of the EPA channel, we observe that the maximum difference between the practical strategy (blue curve) and the BS rates (magenta line) is of 2.20 bits/dim for $I/S = 0$ dB and of 1.71 bits/dim for $I/S = 5$ dB. In the case of the ETU channel model, the maximum difference in spectral efficiency between the practical strategy and the BS rates is 1.50 bits/dim for $I/S = 0$ dB and 1.29 bits/dim for $I/S = 5$ dB. The proposed relay scheme also performs well in comparison to the point-to-point transmission strategy, justifying the feasibility of our proposed relay scheme. A maximum difference of 1.71 bits/dim can be observed between the practical and the point-to-point scheme (green curve) for the EPA case, when $I/S = 0$ dB and $I/S = 5$ dB. Similarly, for the ETU case, a maximum difference of 0.90 bits/dim when $I/S = 0$ dB and 1.16 bits/dim when $I/S = 5$ dB can be noted. However, it is interesting to observe that the benefit - in terms of spectral efficiency - of

utilizing a relay (baseline two-hop communication scheme) over the classical point-to-point communication scheme can be realized when considering realistic fading channel models in which the strength of the relay-destination link is 5 dB stronger than the source-destination link as shown in Fig. 8 and Fig. 10. The difference in spectral efficiency for the practical LTE channel model is noticeably less than the one in the AWGN case, highlighting the degrading effects of multipath on the proposed relay strategy. Nonetheless, even for these two practically relevant LTE channel models, the proposed strategy still provides remarkable improvements in spectral efficiency over the basic BS and point-to-point transmission schemes.

VI. Conclusions

In this paper, we designed a practical transmission strategy for the Gaussian half-duplex relay channel by using codes as in the LTE standard and by running simulations on an LTE test bench. The scheme uses superposition encoding, decode-and-forward relaying and successive interference cancellation in order to rate split the message into three parts and send them in two time slots from a source to a destination with the help of a relay (which forwards one of the three sub-messages). Comparisons between the theoretical achievable rate with (point-to-point capacity achieving) Gaussian codes and the rate achieved in a practical scenario were provided for an overall system BLER of 10^{-2} for the two-antenna case (and single-antenna case) at the destination. The simulation results shed light on the benefits of physical layer cooperation with respect to direct-link transmission and to the case of a baseline two-hop communication scheme. The half-duplex relay strategy was also shown to provide robust spectral efficiency gains over the proposed baseline two-hop communication scheme using the EPA and ETU channel models.

The results presented in this paper should be used to reconsider the case of relay stations for future 5G deployments where new layers 1 and 2 signaling can be envisaged to reap the benefits from physical-layer relaying. This may prove to be even more effective in millimeter-wave deployments where predominantly line-of-sight links are required and wired access-points will be extremely costly for large scale deployment. In this case we would have a few donor eNBs with an optical backbone to the core network and many wireless relay stations using physical layer relaying to their proximity when possible. For UEs in line-of-sight with both the destination and relay, there will be substantial benefits as described by these results. This study shows strong promise to be deployed in upcoming release standards of LTE as well as 5G systems with respect to advanced relay architectures.

Future work would include the analysis of the proposed three-part-message when channel state information is not available and hence has to be estimated/learned as well as the investigation of resource allocation strategies for dynamic bandwidth assignment. We considered equal duration of the two phases and a fixed value for the superposition factor; for inclusion in real-time systems, these parameters have to be adaptive. More general dimensioning of resources can be made over H-ARQ rounds. Higher-order MIMO configurations can be considered at the relay and source as well. An analysis of the achievable rate, given an outage constraint in a fading scenario would also present an interesting avenue for future research.

Appendix
Performance Guarantee of the Proposed Scheme

A well-known upper bound on the capacity C of the relay channel is the cut-set bound [2] which, for the channel model in (1), reads as (15), shown at the top of the next page, where the different (in)equalities are explained in what follows.

(a): By using the chain rule for the mutual information and because $I(S_r;\cdot) \le H(S_r)$, since S_r is a discrete random variable.

(b): By defining $\gamma := \mathbb{P}[S_r = 0] \in [0,1]$ and $H(S_r) = \mathcal{H}(\gamma) = -\gamma\log(\gamma) - (1-\gamma)\log(1-\gamma) \le \log(2)$.

(c): By letting n_d be the number of antennas at the destination (in our system model in (1) we have $n_d = 2$), since "Gaussian maximizes entropy" and by defining

$$\mathbf{K}_\ell := \mathrm{Cov}\left[\begin{matrix}X_s\\ X_r\end{matrix}\right]\Big|_{S_r=\ell} = \left[\begin{matrix}P_{s|\ell} & \rho_\ell\sqrt{P_{s|\ell}P_{r|\ell}}\\ \rho_\ell^*\sqrt{P_{s|\ell}P_{r|\ell}} & P_{r|\ell}\end{matrix}\right] : |\rho_\ell| \le 1,$$

for some $(P_{s|0}, P_{s|1}, P_{r|0}, P_{r|1}) \in \mathbb{R}_+^4$ satisfying the average power constraint

$$\gamma P_{u|0} + (1-\gamma)P_{u|1} \le 1, \quad u \in \{s,r\}. \tag{14}$$

It is not difficult to see that $|\rho_0| = 0$ is optimal, since the relay is listening. For notation convenience we let $\rho := \rho_1$ in (15), as shown at the top of the next page.

(d): By defining $\upsilon := \frac{\boldsymbol{h}_s^H\boldsymbol{h}_r}{\|\boldsymbol{h}_s\|\|\boldsymbol{h}_r\|} : |\upsilon| \le 1$, $t_s := \|\boldsymbol{h}_s\|^2 P_{s|1}$, $t_r := \|\boldsymbol{h}_r\|^2 P_{r|1}$, we obtain

$$\begin{aligned}
T_1 :&= \log\left|\mathbf{I}_{n_d} + \left[\boldsymbol{h}_s\ \boldsymbol{h}_r\right]\mathbf{K}_1\left[\boldsymbol{h}_s\ \boldsymbol{h}_r\right]^H\right| \\
&\overset{(f)}{=} \log\left|\mathbf{I}_2 + \left[\boldsymbol{h}_s\ \boldsymbol{h}_r\right]^H\left[\boldsymbol{h}_s\ \boldsymbol{h}_r\right]\mathbf{K}_1\right| \\
&= \log\left|\mathbf{I}_2 + \left[\begin{matrix}\|\boldsymbol{h}_s\|^2 & \boldsymbol{h}_s^H\boldsymbol{h}_r\\ \boldsymbol{h}_r^H\boldsymbol{h}_s & \|\boldsymbol{h}_r\|^2\end{matrix}\right]\mathbf{K}_1\right| \\
&= \log\Big(1 + t_s + t_r + t_s t_r\left(1-|\upsilon|^2\right)\left(1-|\rho|^2\right) \\
&\qquad + 2\sqrt{t_r t_s}\,\Re\left\{\upsilon\rho^*\right\}\Big) \\
&\overset{(g)}{\le} \log\Big(1 + t_s + t_r + 2\sqrt{t_s t_r}|\upsilon||\rho| \\
&\qquad + t_s t_r(1-|\upsilon|^2)(1-|\rho|^2)\Big) \\
&\overset{(h)}{\le} \log\left(1 + t_s + t_r + 2\sqrt{t_s t_r}|\upsilon| + t_s t_r(1-|\upsilon|^2)\right) \\
&\overset{(i)}{\le} -2\log(1-\gamma) + \log\Big(1 + x + y + 2\sqrt{xy}|\upsilon| \\
&\qquad + xy(1-|\upsilon|^2)\Big),
\end{aligned}$$

where: (i) the equality in (f) follows from the Sylvester's determinant identity; (ii) the inequality in

$$
\begin{aligned}
\mathsf{C} \;&\leq\; \max_{P_{X_s,X_r,S_r}} \min\left\{I\left(X_s, X_r, S_r; \mathbf{Y}_d\right), I\left(X_s; Y_r, \mathbf{Y}_d | X_r, S_r\right)\right\} \\
&\overset{(a)}{\leq} \max_{P_{X_s,X_r,S_r}} \min\left\{I\left(X_s, X_r; \mathbf{Y}_d | S_r\right) + H\left(S_r\right), I\left(X_s; Y_r, \mathbf{Y}_d | X_r, S_r\right)\right\} \\
&\overset{(b)}{=} \max_{\gamma\in[0,1],\ P_{X_s,X_r|S_r=0},\ P_{X_s,X_r|S_r=1}} \min\Big\{ \\
&\qquad \gamma\, I\left(X_s, X_r; \mathbf{h}_s X_s + Z_d | S_r = 0\right) + (1-\gamma)\, I\left(X_s, X_r; \mathbf{h}_r X_r + \mathbf{h}_s X_s + Z_d | S_r = 1\right) + \mathcal{H}\left(\gamma\right), \\
&\qquad \gamma\, I\left(X_s; h_{rs} X_s + Z_r,\ \mathbf{h}_s X_s + Z_d | X_r, S_r = 0\right) + (1-\gamma)\, I\left(X_s; Z_r,\ \mathbf{h}_s X_s + Z_d | X_r, S_r = 1\right)\Big\} \\
&\overset{(c)}{=} \max_{\gamma\in[0,1],\ |\rho|\leq 1,\ \gamma P_{s|0}+(1-\gamma)P_{s|1}\leq 1,\ \gamma P_{r|0}+(1-\gamma)P_{r|1}\leq 1} \min\Big\{ \\
&\qquad \gamma \log\left(1 + \|\mathbf{h}_s\|^2 P_{s|0}\right) + (1-\gamma)\log\left|\mathbf{I}_{n_d} + \begin{bmatrix}\mathbf{h}_s & \mathbf{h}_r\end{bmatrix} \mathbf{K}_1 \begin{bmatrix}\mathbf{h}_s & \mathbf{h}_r\end{bmatrix}^H\right| + \mathcal{H}\left(\gamma\right), \\
&\qquad \gamma \log\left(1 + \left(\|\mathbf{h}_s\|^2 + |h_{rs}|^2\right) P_{s|0}\right) + (1-\gamma)\log\left(1 + \|\mathbf{h}_s\|^2 P_{s|1}\right)\Big\} \\
&\overset{(d)}{\leq} \max_{\gamma\in[0,1],\ \alpha\in[0,1],\ \beta\in[0,1]} 2\mathcal{H}(\gamma) - (1-\gamma)\log(1-\gamma) + \min\Big\{ \\
&\qquad \gamma \log\left(1 + \|\mathbf{h}_s\|^2\alpha\right) + (1-\gamma)\log\left(1 + x + y + 2\sqrt{xy}|\upsilon| + xy(1-|\upsilon|^2)\right)\Big|_{\substack{x=\|\mathbf{h}_s\|^2(1-\alpha)\\ y=\|\mathbf{h}_r\|^2(1-\beta)\\ \upsilon=\frac{\mathbf{h}_s^H\mathbf{h}_r}{\|\mathbf{h}_s\|\|\mathbf{h}_r\|}}}, \\
&\qquad \gamma \log\left(1 + \left(\|\mathbf{h}_s\|^2 + |h_{rs}|^2\right)\alpha\right) + (1-\gamma)\log\left(1 + \|\mathbf{h}_s\|^2(1-\alpha)\right)\Big\} \\
&\leq \max_{\gamma\in[0,1]} 2.51\log(2) + \min\Big\{ \\
&\qquad \gamma \log\left(1 + \|\boldsymbol{h}_s\|^2\right) + (1-\gamma)\log\left(1 + \|\boldsymbol{h}_s\|^2 + \|\boldsymbol{h}_r\|^2 + 2\|\boldsymbol{h}_s\|\|\boldsymbol{h}_r\||\upsilon| + \|\boldsymbol{h}_s\|^2\|\boldsymbol{h}_r\|^2(1-|\upsilon|^2)\right), \\
&\qquad \gamma \log\left(1 + \|\boldsymbol{h}_s\|^2 + |h_{rs}|^2\right) + (1-\gamma)\log\left(1 + \|\boldsymbol{h}_s\|^2\right)\Big\} \\
&\overset{(e)}{=} 2.51\log(2) + \log\left(1 + \|\boldsymbol{h}_s\|^2\right) + \max_{\gamma\in[0,1]} \min\left\{(1-\gamma)\,\overline{a}, \gamma\,\overline{b}\right\} \\
&= 2.51\log(2) + \log\left(1 + \|\boldsymbol{h}_s\|^2\right) + \frac{\overline{a}\;\overline{b}}{\overline{a}+\overline{b}}, \qquad (15)
\end{aligned}
$$

(g) follows since $\Re\{\upsilon\rho^*\} \leq |\rho||\upsilon|$; (iii) the inequality in (h) follows by using $|\rho| = 1$ in the term increasing in $|\rho|$ and $|\rho| = 0$ in the term decreasing in $|\rho|$; (iv) the inequality in (i) follows by parameterizing the power constraints in (14) as $\gamma\, P_{s|0} = \alpha$, $(1-\gamma)\, P_{s|1} = 1-\alpha$, $\gamma\, P_{r|0} = \beta$, $(1-\gamma)\, P_{r|1} = 1-\beta$, for some $\alpha \in [0,1]$, $\beta \in [0,1]$, and by introducing $x := \|\boldsymbol{h}_s\|^2(1-\alpha)$, $y := \|\boldsymbol{h}_r\|^2(1-\beta)$. By means of similar steps, we have

$$
\begin{aligned}
T_2 &:= \log\left(1 + \|\mathbf{h}_s\|^2 P_{s|0}\right) = \log\left(1 + \|\mathbf{h}_s\|^2 \frac{\alpha}{\gamma}\right) \\
&= \log\left(\gamma + \|\mathbf{h}_s\|^2\alpha\right) - \log\left(\gamma\right) \\
&\leq \log\left(1 + \|\mathbf{h}_s\|^2\alpha\right) - \log\left(\gamma\right).
\end{aligned}
$$

Hence, the first term within the min in the equality in (c) in (15) can be upper bounded as

$$
\begin{aligned}
&\gamma\, T_2 + (1-\gamma)\, T_1 + \mathcal{H}(\gamma) \\
&\leq 2\mathcal{H}(\gamma) - (1-\gamma)\log(1-\gamma) + \gamma \log\left(1 + \|\mathbf{h}_s\|^2\alpha\right) \\
&\quad + (1-\gamma)\log\left(1 + x + y + 2\sqrt{xy}|\upsilon| + xy(1-|\upsilon|^2)\right).
\end{aligned}
$$

Similarly, we can upper bound the second term within the min in the equality in (c) in (15) as

$$
\begin{aligned}
T_3 &:= \gamma \log\left(1 + \left(\|\mathbf{h}_s\|^2 + |h_{rs}|^2\right) P_{s|0}\right) \\
&\qquad + (1-\gamma)\log\left(1 + \|\mathbf{h}_s\|^2 P_{s|1}\right) \\
&= \gamma \log\left(\gamma + \left(\|\mathbf{h}_s\|^2 + |h_{rs}|^2\right)\alpha\right) \\
&\qquad + (1-\gamma)\log\left(1 - \gamma + \|\mathbf{h}_s\|^2(1-\alpha)\right) + \mathcal{H}(\gamma) \\
&\leq \gamma \log\left(1 + \left(\|\mathbf{h}_s\|^2 + |h_{rs}|^2\right)\alpha\right) \\
&\qquad + (1-\gamma)\log\left(1 + \|\mathbf{h}_s\|^2(1-\alpha)\right) + \mathcal{H}(\gamma).
\end{aligned}
$$

Therefore, the expression in the equality in (c) in (15) can be upper bounded as

$$
\begin{aligned}
&\min\left\{\gamma\, T_2 + (1-\gamma)\, T_1 + \mathcal{H}(\gamma), T_3\right\} \\
&\leq 2\mathcal{H}(\gamma) - (1-\gamma)\log(1-\gamma) \\
&\quad + \min\Big\{\gamma \log\left(1 + \|\mathbf{h}_s\|^2\alpha\right) + (1-\gamma)\log\left(1 + x + y\right. \\
&\qquad \left. + 2\sqrt{xy}|\upsilon| + xy(1-|\upsilon|^2)\right), \\
&\qquad \gamma \log\left(1 + \left(\|\mathbf{h}_s\|^2 + |h_{rs}|^2\right)\alpha\right) \\
&\qquad + (1-\gamma)\log\left(1 + \|\mathbf{h}_s\|^2(1-\alpha)\right)\Big\},
\end{aligned}
$$

which is precisely the expression in the inequality in (d).

(e): By defining

$$\overline{a} := \log\left(1 + \frac{\|\boldsymbol{h}_r\|^2 + 2\|\boldsymbol{h}_s\|\|\boldsymbol{h}_r\||\upsilon|}{1+\|\boldsymbol{h}_s\|^2} + \frac{\|\boldsymbol{h}_s\|^2\|\boldsymbol{h}_r\|^2(1-|\upsilon|^2)}{1+\|\boldsymbol{h}_s\|^2}\right) \leq \underline{a} + \log(2), \quad (16)$$

$$\overline{b} := \log\left(1 + \frac{|h_{rs}|^2}{1+\|\boldsymbol{h}_s\|^2}\right) \leq \underline{b} + \log(2), \quad (17)$$

where $\underline{a}$ and $\underline{b}$ are defined in (2b) and (2c), respectively.

Notice that the above analysis holds for the case when the destination is equipped with a general number n_d of antennas.

We next show that the upper bound in (15) and the lower bound in (2) are to within a constant gap of one another. By taking the difference between (15) and (2) we obtain

$$\mathsf{gap} \leq \frac{\overline{a}\,\overline{b}}{\overline{a}+\overline{b}} - \frac{\underline{a}\,\left[\underline{b}\right]^+}{\underline{a}+\left[\underline{b}\right]^+} + 2.51\log(2) \leq 3.51\log(2).$$

Thus, the two-phase three-part-message scheme designed in Section II is at most $3.51\log(2)$ bits/dim from capacity, irrespective of the number of antennas at the destination and of the channel gains.

REFERENCES

[1] A. Nosratinia, T. E. Hunter, and A. Hedayat, "Cooperative communication in wireless networks," *IEEE Commun. Mag.*, vol. 42, no. 10, pp. 74–80, Oct. 2004.

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

[3] Third Generation Partnership Project, "Relay architectures for E-UTRA (LTE-Advanced)(Release 9)," 3GPP, Tech. Rep. 3GPP TR 36.806 v2.0.0, Feb. 2010.

[4] Third Generation Partnership Project, "Further advancements for E-UTRA physical layer aspects(Release 9)," 3GPP, Tech. Rep. 3GPP TR 36.814 v9.0.0, Mar. 2010.

[5] Third Generation Partnership Project, "Physical layer for relaying operation (Rel. 10)," 3GPP, Tech. Rep. 3GPP TS 36.216 V10.3.1, Sep. 2011.

[6] R. Thomas, M. Cardone, R. Knopp, D. Tuninetti, and B. T. Maharaj, "An LTE implementation of a novel strategy for the Gaussian half-duplex relay channel," in *Proc. IEEE ICC*, Jun. 2015, pp. 2209–2214.

[7] M. Cardone, D. Tuninetti, R. Knopp, and U. Salim, "On the Gaussian half-duplex relay channel," *IEEE Trans. Inf. Theory*, vol. 60, no. 5, pp. 2542–2562, May 2014.

[8] M. Iwamura, H. Takashi, and S. Nagata, "Relay technology in LTE-advanced," *NTT Docomo Tech. J.*, vol. 12, no. 2, pp. 29–36, Sep. 2010.

[9] Z. Juan, P. Sartori, and B. Wei, "Performance analysis of layer 1 relays," in *Proc. IEEE Int. Conf. Commun. Commun. Workshop (ICC)*, Jun. 2009, pp. 1–6.

[10] Ericsson, "A discussion on some technology components for LTE-advanced," 3GPP/Ericsson, Tech. Rep. 3GPP R1-082084, May 2008.

[11] T. Beniero, S. Redana, J. Hamalainen, and B. Raaf, "Effect of relaying on coverage in 3GPP LTE-advanced," in *Proc. IEEE VTC-Spring*, Apr. 2009, pp. 1–5.

[12] A. Saleh, S. Redana, B. Raaf, T. Riihonen, J. Hamalainen, and R. Wichman, "Performance of amplify-and-forward and decode-and-forward relays in LTE-advanced," in *Proc. IEEE VTC-Fall*, Sep. 2009, pp. 1–5.

[13] N. Zlatanov, V. Jamali, and R. Schober, "On the capacity of the two-hop half-duplex relay channel," in *Proc. IEEE GLOBECOM*, Dec. 2015, pp. 1–7.

[14] B. Zhao and M. C. Valenti, "Distributed turbo coded diversity for relay channel," *Electron. Lett.*, vol. 39, no. 10, pp. 786–787, May 2003.

[15] K. Ishibashi, K. Ishii, and H. Ochiai, "Dynamic coded cooperation using multiple turbo codes in wireless relay networks," *IEEE J. Sel. Topics Signal Process.*, vol. 5, no. 1, pp. 197–207, Feb. 2011.

[16] X. Liu and T. J. Lim, "Fountain codes over fading relay channels," *IEEE Trans. Wireless Commun.*, vol. 8, no. 6, pp. 3278–3287, Jun. 2009.

[17] L. Chen, "Opportunistic cooperative communications with Reed-Solomon convolutional concatenated codes," in *Proc. IEEE ICCT*, Sep. 2011, pp. 22–26.

[18] H. Yang, W. Meng, B. Li, and G. Wang, "Physical layer implementation of network coding in two-way relay networks," in *Proc. IEEE ICC*, Jun. 2012, pp. 671–675.

[19] E. Atsan, R. Knopp, S. Diggavi, and C. Fragouli, "Towards integrating quantize-map-forward relaying into LTE," in *Proc. IEEE Inf. Theory Workshop (ITW)*, Sep. 2012, pp. 212–216.

[20] G. J. Bradford and J. N. Laneman, "Low latency relaying schemes for next-generation cellular networks," in *Proc. IEEE Int. Conf. Commun. (ICC)*, Jun. 2012, pp. 4294–4299.

[21] Y. Meng, M. You, S. Zhao, and L. Zhang, "A distributed layered modulation for relay downlink cooperative transmission," in *Proc. IEEE Int. Symp. PIMRC*, Sep. 2010, pp. 1–5.

[22] S. Vanka, S. Srinivasa, Z. Gong, P. Vizi, K. Stamatiou, and M. Haenggi, "Superposition coding strategies: Design and experimental evaluation," *IEEE Trans. Wireless Commun.*, vol. 11, no. 7, pp. 2628–2639, Jul. 2012.

[23] J. Zhang, J. Jia, Q. Zhang, and E. M. K. Lo, "Implementation and evaluation of cooperative communication schemes in software-defined radio testbed," in *Proc. IEEE INFOCOM*, Mar. 2010, pp. 1–9.

[24] M. Duarte, A. Sengupta, S. Brahma, C. Fragouli, and S. Diggavi, "Quantize-map-forward (QMF) relaying: An experimental study," in *Proc. ACM Int. Symp. Mobile Ad Hoc Netw. Comput.*, Jul. 2013, pp. 227–236.

[25] S. Brahma, M. Duarte, A. Sengupta, I. H. Wang, C. Fragouli, and S. Diggavi, "QUILT: A decode/quantize-interleave-transmit approach to cooperative relaying," in *Proc. IEEE INFOCOM*, Apr. 2014, pp. 2508–2516.

[26] F. Kaltenberger. (Feb. 2016). OpenAirInterface project. Eurecom, accessed on Mar. 23, 2016. [Online]. Available: https://gitlab.eurecom.fr/oai/openairinterface5g/wikis/home

[27] C. N. I. Manchon, L. Deneire, P. Mogensen, and T. B. Sorensen, "On the design of a MIMO-SIC receiver for LTE downlink," in *Proc. IEEE 68th Veh. Technol. Conf. VTC*, Sep. 2008, pp. 1–5.

[28] "LTE-advanced the advanced LTE toolbox for more efficient delivery of better user experience," Nokia, Tech. Rep. White Paper, Sep. 2014.

[29] E. Lang, S. Redana, and B. Raaf, "Business impact of relay deployment for coverage extension in 3GPP LTE-advanced," in *Proc. IEEE ICC Workshops*, Jun. 2009, pp. 1–5.

[30] S. Sesia, I. Toufik, and M. Baker, *LTE—The UMTS Long Term Evolution: From Theory to Practice*, 2nd ed. Hoboken, NJ, USA, Wiley, 2011.

[31] Y. Polyanskiy, "Channel coding: Non-asymptotic fundamental limits," Ph.D. dissertation, Dept. Elect. Eng., Princeton University, Princeton, NJ, Nov. 2010.

[32] Third Generation Partnership Project, "Multiplexing and channel coding (release 11)," 3GPP, Tech. Rep. 3GPP TS 36.212 v11.3.0, Jun. 2013.

[33] R. Ghaffar and R. Knopp, "Spatial interference cancellation algorithm," in *Proc. IEEE Wireless Commun. Netw. Conf.*, Apr. 2009, pp. 1–5.

[34] R. Ghaffar and R. Knopp, "Interference suppression for next generation wireless systems," in *Proc. IEEE VTC Spring*, Apr. 2009, pp. 1–5.

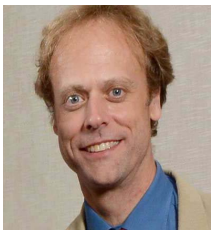
[35] R. Ghaffar and R. Knopp, "Interference suppression strategy for cell-edge users in the downlink," *IEEE Trans. Wireless Commun.*, vol. 11, no. 1, pp. 154–165, Jan. 2012.

[36] Third Generation Partnership Project, "Phyisical layer procedures (Release 11)," 3GPP, Tech. Rep. 3GPP TS 36.213 v11.3.0, Jun. 2013.lo, org, au

[37] Third Generation Partnership Project, "Base station (BS) radio transmission and reception," 3GPP, Tech. Rep. 3GPP TR 36.104 v8.2.0, May 2008.

Robin Rajan Thomas received the B.Sc. degree in electrical information engineering from the University of Witwatersrand, Johannesburg, South Africa, and the M.Eng. degree (Hons.) in electronic engineering from the University of Pretoria, Pretoria, South Africa, in 2008 and 2012, respectively. He is currently pursuing the Ph.D. degree with the Communication Systems Group, Eurecom, Sophia Antipolis, France, and Télécom ParisTech, Paris, France. He was part of the Sentech Chair with the Broadband Wireless Multimedia Communications Research Group, University of Pretoria. His current research focuses on low-complexity advanced receiver and relay architecture design, interference-aware management techniques for mobile communication systems, heterogeneous networks, 3GPP LTE networks, device-to-device communication, location awareness techniques for mobile positioning applications, cognitive radio, and dynamic spectrum access techniques.

Martina Cardone received the B.S. and M.S. (*summa cum laude*) degrees in telecommunications engineering from the Politecnico di Torino, Italy, in 2009 and 2011, respectively, and the M.S. degree in telecommunications engineering, as part of a double degree program, and the Ph.D. degree in electronics and communications (with work done at Eurecom, Sophia Antipolis, France) from Télécom ParisTech, Paris, France, in 2011 and 2015, respectively. She is currently a Post-Doctoral Research Fellow with the Electrical Engineering Department, UCLA. Her main research interests are in network information theory, wireless networks,security and network coding. She received the second prize in the Outstanding Ph.D. Award at Télécom ParisTech, and the Qualcomm Innovation Fellowship in 2014.



Raymond Knopp received the B.Eng. (Hons.) and the M.Eng. degrees in electrical engineering from McGill University, Montreal, Canada, in 1992 and 1993, respectively. He is currently pursuing the Ph.D. degree in communication systems with the Swiss Federal Institute of Technology (EPFL), Lausanne. From 1993 to 1997, he was a Research Assistant with the Mobile Communications Department, Eurecom, Sophia Antipolis, France. From 1997 to 2000 he was a Research Associate with the Mobile Communications Laboratory, Communication Systems Department, EPFL. He is also a Professor with the Communication Systems Group, Eurecom, and the General Secretary of the OpenAirInterface Software Alliance, which aims to bridge the gap between cutting-edge theoretical advances in wireless communications and practical designs. His current research and teaching interests are in the area of digital communications, software radio architectures, and implementation aspects of signal processing systems, and real-time wireless networking protocols. He has a proven track record in managing both fundamental and experimental research projects at an international level.

Daniela Tuninetti received the Ph.D. degree in electrical engineering from ENST/Télécom ParisTech, Paris, France, in 2002 (with work done at the Eurecom Institute, Sophia Antipolis, France). She was a Post-Doctoral Research Associate with the School of Communication and Computer Science, Swiss Federal Institute of Technology, Lausanne, Switzerland, from 2002 to 2004. She has been a Professor with the Department of Electrical and Computer Engineering, The University of Illinois at Chicago (UIC), since 2005. Her research interests are in the ultimate performance limits of wireless interference networks (with special emphasis on cognition and user cooperation), coexistence between radar and communication systems, multirelay networks, content-type coding, and caching. She was the Editor-in-Chief of the IEEE Information Theory Society Newsletter from 2006 to 2008, an Editor of the IEEE COMMUNICATION LETTERS from 2006 to 2009, and the IEEE TRANSACTIONS ON WIRELESS COMMUNICATIONS from 2011 to 2014. She is currently an Associate Editor of the IEEE TRANSACTIONS ON INFORMATION THEORY. She was a recipient of a Best Paper Award at the European Wireless Conference in 2002, an NSF CAREER Award in 2007, and was named a UIC University Scholar in 2015.

Bodhaswar T. Maharaj received the B.Sc. and M.Sc. degrees in electronic engineering from the University of KwaZulu-Natal, Durban, South Africa, the M.Sc. degree (Hons.) in operational telecommunications from the University of Coventry, and the Ph.D. degree in wireless communications from the University of Pretoria. He is currently the Dean of the Faculty of Engineering, Built Environment and IT, University of Pretoria, and also holds the position of Sentech Chair in Broadband Wireless Multimedia Communications with the Department of Electrical, Electronic, and Computer Engineering. His research interests are in MIMO channel modeling, OFDM-MIMO systems, and cognitive radio.