EURASIP Journal on Advances in Signal Processing

Wireless Cooperative Networks

Guest Editors: Andrea Conti, Jiangzhou Wang,
Hyundong Shin, Ramesh Annavajjala, and Moe Z. Win

Wireless Cooperative Networks

EURASIP Journal on Advances in Signal Processing

Wireless Cooperative Networks

Guest Editors: Andrea Conti, Jiangzhou Wang,
Hyundong Shin, Ramesh Annavajjala, and Moe Z. Win

Contents

Hindawi Publishing Corporation
EURASIP Journal on Advances in Signal Processing
Volume 2008, Article ID 810149, 2 pages
doi:10.1155/2008/810149

Editorial
Wireless Cooperative Networks

Andrea Conti,[1] Jiangzhou Wang,[2] Hyundong Shin,[3] Ramesh Annavajjala,[4] and Moe Z. Win[5]

[1] *Engineering Department in Ferrara (ENDIF), University of Ferrara, 44100 Ferrara, Italy*
[2] *Department of Electronics, University of Kent, Canterbury, Kent CT2 7NT, UK*
[3] *Department of Radio Communication Engineering, School of Electronics and Information, Kyung Hee University, Gueonggi-Do 449-701, South Korea*
[4] *Mitsubishi Electric Research Laboratories, 201 Broadway, Cambridge, MA 02139, USA*
[5] *Department of Aeronautics and Astronautics, Massachusetts Institute of Technology, Cambridge, MA 02139, USA*

Correspondence should be addressed to Andrea Conti, a.conti@ieee.org

Received 29 July 2008; Accepted 29 July 2008

Cooperative networks are gaining an increasing interest in information and communications technologies since such networks can improve communication capability and provide a fertile environment for the development of context-aware services. Cooperative communications and networking represent a new paradigm which involves both transmission and distributed processing, promising significant increase of capacity and diversity gain in wireless networks. From one hand, the integration of long-range and short-range wireless communication networks (e.g., infrastructured networks such as 3G, wireless ad hoc networks, and wireless sensor networks) improves the performance in terms of both area coverage and quality of service (QoS). On the other hand, the cooperation among nodes, as in the case of wireless sensor networks, allows a distributed space-time signal processing which enables environmental monitoring, localization techniques, distributed measurements, and others, with a reduced complexity or energy consumption per node. The relevance of this topic is also reflected by numerous technical sessions in current international conferences as well as by the increasing number of national and international projects on these aspects.

This special issue aims to collect cutting-edge research achievements in this area. We solicited papers that present original and unpublished work on topics including, but not limited to, the following: physical layer models, for example, channel models (statistics, fading, MIMO, feedback); device constraints (power, energy, multiple access, synchronization) and resource management; distributed processing for cooperative networks (e.g., distributed compression in wireless sensor networks, channel and network codes design); performance metrics (e.g., capacity, cost, outage, delay, energy, scaling laws); cross-layer issues, for example, PHY/MAC/NET interactions, joint source-channel coding, separation theorems; multiterminal information theory; multihop communications; integration of wireless heterogeneous (long- and short-range) systems.

In "Asymptotic analysis of large cooperative relay networks using random matrix theory" by H. Li et al., cooperative relay networks with large number of nodes are analyzed, and in particular the asymptotic performance improvement of cooperative transmission over direct transmission and relay transmission is analyzed using random matrix theory. The key idea is to investigate the eigenvalue distributions related to channel capacity and to analyze the moments of this distribution in large wireless networks. The analysis in this paper provides important tools for the understanding and the design of large cooperative wireless networks.

H. Van Khuong and T. Le-Ngoc propose, in the paper "Bandwidth-efficient cooperative relaying schemes with multi-antenna relay," coded cooperative relaying schemes in which all successfully decoded signals from multiple sources are forwarded simultaneously by a multiantenna relay to a common multiantenna destination to increase bandwidth efficiency. These schemes facilitate various retransmission strategies at relay together with single-user and multiuser iterative decoding techniques at destination, suitable for tradeoffs between performance, latency, and complexity.

The problem of choosing the best relay node in relaying networks is addressed in "Performance of multiple-relay cooperative diversity systems with best-relay selection over rayleigh fading channels" by S. S. Ikki and M. H. Ahmed. They consider an amplify-and-forward (AF) cooperative diversity system where a source node communicates with a

destination node directly and indirectly (through multiple relays). It is shown that the best-relay selection reduces the amount of required resources while improving the performance. Authors derive closed form expressions for tight lower bounds on the symbol error probability and outage probability.

A double-differential modulation for the amplify-and-forward protocol over Nakagami-m fading channels with carrier offsets is proposed by M. R. Bhatnagar et al. in "Cooperative communications over flat fading channels with carrier offsets: a double-differential modulation approach." They propose an emulated maximum ratio combining (EMRC) decoder, which could be used by the double-differential receiver in the absence of exact channel knowledge. Approximate bit error rate (BER) analysis is performed for the double-differential modulation-based cooperative communication system. They propose a double-differential system which is immune to random carrier offsets, whereas the conventional single-differential modulation-based cooperative system breaks down, and perform better than training-based cooperative system which utilizes training data to estimate carrier offsets and channel gains.

In "Delay optimization in cooperative relaying with cyclic delay diversity," S. B. Slimane et al. propose to inserting random delays at the nonregenerative fixed relays to further improve the system performance. However, random delays result in limited performance gain from multipath diversity. In this paper, two promising delay optimization schemes are introduced for a multicellular OFDM system with cooperative relaying, stationary multiple users, and fixed relays.

A. Conti et al. address the construction of space-time codes for cooperative communications over block fading channels in the paper "Pragmatic space-time codes for cooperative relaying in block fading channels." They consider a pragmatic approach based on the concatenation of convolutional codes and BPSK/QPSK modulation to obtain cooperative codes for relay networks. The pairwise error probability, an asymptotic bound on the frame error probability, and a design criterion to optimize both diversity and coding gain are also derived. While the implementation of pragmatic space-time codes only requires common convolutional encoders and Viterbi decoders with suitable generators, rate, and branch metric, they perform well in block fading channels, including quasistatic channel, even with a low number of states and relays.

In "Interference mitigation in cooperative SFBC-OFDM," D. Sreedhar and A. Chockalingam consider cooperative space-frequency block-coded orthogonal frequency-division multiplexing (SFBC-OFDM) networks with amplify-and-forward (AF) and decode-and-forward (DF) protocols at the relays. They propose an interference cancellation algorithm for this system at the destination node, and show that the proposed algorithm effectively mitigates the intersymbol interference and intercarrier interference effects.

For what concerns the MAC layer, J. Alonso-Zrate et al., in "Persistent RCSMA: a MAC protocol for a distributed cooperative ARQ scheme in wireless networks," present the persistent relay carrier sensing multiple access (PRCSMA) protocol that allows for the execution of a distributed cooperative automatic retransmission request (ARQ) scheme in IEEE 802.11 wireless networks. The underlying idea of the PRCSMA protocol is to modify the basic rules of the IEEE 802.11 MAC protocol to execute a distributed cooperative ARQ scheme in wireless networks to enhance their performance and to extend coverage.

At the scheduling level, "Optimally joint subcarrier matching and power allocation in OFDM multihop system" by W. Wang et al. propose an optimally joint subcarrier matching and power allocation scheme to maximize the channel capacity under total system power constrain of OFDM systems. The problem is formulated as a mixed binary integer programming problem (which is prohibitive to find the global optimum in terms of complexity) and then a low-complexity scheme by making use of the equivalent channel power gain for any matched subcarrier pair is proposed.

W. Mesbah and T. N. Davidson study, in "Power and resource allocation for orthogonal multiple access relay systems," the problem of joint power and channel resource allocation for orthogonal multiple access relay (MAR) systems to maximize the achievable rate region. The authors consider four relaying strategies and show that the problem can be formulated as a quasiconvex problem in several cases. Therefore, efficient algorithms can be derived for joint optimal power and channel resource allocation.

At networking level, the paper "Resource sharing via planed relay for HWN" by C. Shen et al. presents an improved version of adaptive distributed cross-layer routing algorithm for hybrid wireless network with dedicated relay stations. They verify that the performance of routing protocol benefits of the hybrid wireless networks nature.

Collaboration in heterogeneous wireless networks is addressed by A. Bazzi et al. a in "Multi radio resource management: parallel transmission for higher throughput." Mobile communication systems beyond the third generation will see the interconnection of heterogeneous radio access networks (UMTS, WiMax, wireless local area networks, etc.) to always provide the best QoS to users with multimode terminals. The issue of parallel transmission over multiple radio access technologies (RATs) is investigated focusing the attention on the QoS perceived by the end users. It shows the real benefit of parallel transmission over multiple RATs and how it is conditioned to the fulfilment of some requirements related to the particular kind of RATs, the multiradio resource management strategy, and the transport level protocol behavior.

Andrea Conti
Jiangzhou Wang
Hyundong Shin
Ramesh Annavajjala
Moe Z. Win

Hindawi Publishing Corporation
EURASIP Journal on Advances in Signal Processing
Volume 2008, Article ID 235867, 15 pages
doi:10.1155/2008/235867

Research Article

Asymptotic Analysis of Large Cooperative Relay Networks Using Random Matrix Theory

Husheng Li,[1] Z. Han,[2] and H. Poor[3]

[1] *Department of Electrical Engineering and Computer Science, The University of Tennessee, Knoxville, TN 37996-2100, USA*
[2] *Department of Electrical and Computer Engineering, Boise State University, Boise, ID 83725, USA*
[3] *Department of Electrical Engineering, Princeton University, Princeton, NJ 08544, USA*

Correspondence should be addressed to Husheng Li, husheng@ece.utk.edu

Received 29 November 2007; Accepted 22 February 2008

Recommended by Andrea Conti

Cooperative transmission is an emerging communication technology that takes advantage of the broadcast nature of wireless channels. In cooperative transmission, the use of relays can create a virtual antenna array so that multiple-input/multiple-output (MIMO) techniques can be employed. Most existing work in this area has focused on the situation in which there are a small number of sources and relays and a destination. In this paper, cooperative relay networks with large numbers of nodes are analyzed, and in particular the asymptotic performance improvement of cooperative transmission over direction transmission and relay transmission is analyzed using random matrix theory. The key idea is to investigate the eigenvalue distributions related to channel capacity and to analyze the moments of this distribution in large wireless networks. A performance upper bound is derived, the performance in the low signal-to-noise-ratio regime is analyzed, and two approximations are obtained for high and low relay-to-destination link qualities, respectively. Finally, simulations are provided to validate the accuracy of the analytical results. The analysis in this paper provides important tools for the understanding and the design of large cooperative wireless networks.

1. INTRODUCTION

In recent years, cooperative transmission [1, 2] has gained considerable attention as a potential transmit strategy for wireless networks. Cooperative transmission efficiently takes advantage of the broadcast nature of wireless networks, and also exploits the inherent spatial and multiuser diversities of the wireless medium. The basic idea of cooperative transmission is to allow nodes in the network to help transmit/relay information for each other, so that cooperating nodes create a virtual multiple-input/multiple-output (MIMO) transmission system. Significant research has been devoted to the design of cooperative transmission schemes and the integration of this technique into cellular, WiFi, Bluetooth, ultrawideband, Worldwide Interoperability for Microwave Access (WiMAX), and ad hoc and sensor networks. Cooperative transmission is also making its way into wireless communication standards, such as IEEE 802.16j.

Most current research on cooperative transmission focuses on protocol design and analysis, power control, relay selection, and cross-layer optimization. Examples of representative work are as follows. In [3], transmission protocols for cooperative transmission are classified into different types and their performance is analyzed in terms of outage probabilities. The work in [4] analyzes more complex transmitter cooperative schemes involving dirty paper coding. In [5], centralized power allocation schemes are presented, while energy-efficient transmission is considered for broadcast networks in [6]. In [7], oversampling is combined with the intrinsic properties of orthogonal frequency division multiplexing (OFDM) symbols, in the context of maximal ratio combining (MRC) and amplify-and-forward relaying, so that the rate loss of cooperative transmission can be overcome. In [8], the authors evaluate cooperative-diversity performance when the best relay is chosen according to the average signal-to-noise ratio (SNR), and the outage probability of relay selection based on the instantaneous SNR. In [9], the authors propose a distributed relay selection scheme that requires limited network knowledge and is based on instantaneous SNRs. In [10], sensors are assigned for cooperation so as to reduce power consumption. In [11], cooperative transmission is used to create new paths

so that energy depleting critical paths can be bypassed. In [12], it is shown that cooperative transmission can improve the operating point for multiuser detection so that multiuser efficiency can be improved. Moreover, network coding is also employed to improve the diversity order and bandwidth efficiency. In [13], a buyer/seller game is proposed to circumvent the need for exchanging channel information to optimize the cooperative communication performance. In [14], it is demonstrated that boundary nodes can help backbone nodes' transmissions using cooperative transmission as future rewards for packet forwarding. In [15], auction theory is explored for resource allocation in cooperative transmission.

Most existing work in this area analyzes the performance gain of cooperative transmission protocols assuming small numbers of source-relay-destination combinations. In [16], large relay networks are investigated without combining of source-destination and relay-destination signals. In [17], transmit beamforming is analyzed asymptotically as the number of nodes increases without bound. In this paper, we analyze the asymptotic (again, as the number of nodes increases) performance improvement of cooperative transmission over direct transmission and relay transmission. Relay nodes are considered in this paper while only beamforming in point-to-point communication is considered in [17]. Unlike [16], in which only the indirect source-relay-destination link is considered, we consider the direct link from source nodes to destination nodes. The primary tool we will use is random matrix theory [18, 19]. The key idea is to investigate the eigenvalue distributions related to capacity and to analyze their moments in the asymptote of large wireless networks. Using this approach, we derive a performance upper bound, we analyze the performance in the low signal-to-noise-ratio regime, and we obtain approximations for high and low relay-to-destination link qualities. Finally, we provide simulation results to validate the analytical results.

This paper is organized as follows. In Section 2, the system model is given, while the basics of random matrix theory are discussed in Section 3. In Section 4, we analyze the asymptotic performance and construct an upper bound for cooperative relay networks using random matrix theory. Some special cases are analyzed in Section 5, and simulation results are discussed in Section 6. Finally, conclusions are drawn in Section 7.

2. SYSTEM MODEL

We consider the system model shown in Figure 1. Suppose there are M source nodes, M destination nodes, and K relay nodes. Denote by $\mathbf{H}$, $\mathbf{F}$, and $\mathbf{G}$ the channel matrices of source-to-relay, relay-to-destination, and source-to-destination links, respectively, so that $\mathbf{H}$ is $M \times K$, $\mathbf{F}$ is $K \times M$, and $\mathbf{G}$ is $M \times M$. Transmissions take place in two stages. Further denote the thermal noise at the relays by the K-vector $\mathbf{z}$, the noise in the first stage at the destination by the M-vector $\mathbf{w}_1$ and the noise in the second stage at the destination by the M-vector $\mathbf{w}_2$. For simplicity of notation, we assume that all of the noise variables have the same power

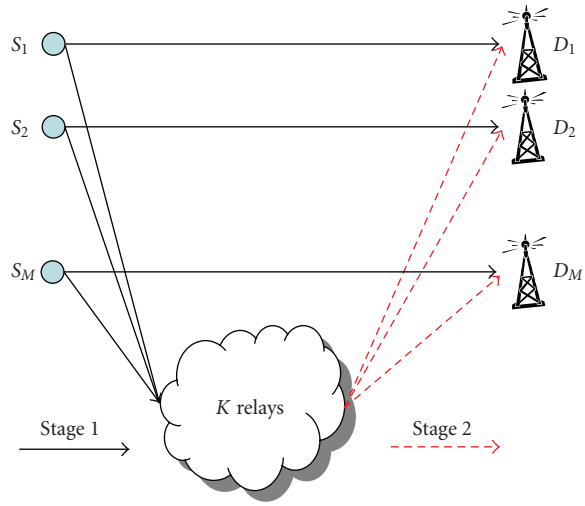


FIGURE 1: Cooperative transmission system model.

and denote this common value by σ_n^2, the more general case being straightforward. The signals at the source nodes are collected into the M-vector $\mathbf{s}$. We assume that the transmit power of each source node and each relay node is given by P_s and P_r, respectively. For simplicity, we further assume that matrices $\mathbf{H}$, $\mathbf{F}$, and $\mathbf{G}$ have independent and identically distributed (i.i.d.) elements whose variances are normalized to $1/K$, $1/M$, and $1/M$, respectively. Thus, the average norm of each column is normalized to 1; otherwise the receive SNR at both relay nodes and destination nodes will diverge in the large system limit. (Note that we do not specify the distribution of the matrix elements since the large system limit is identical for most distributions, as will be seen later.) The average channel power gains, determined by path loss, of source-to-relay, source-to-destination, and relay-to-destination links are denoted by g_{sr}, g_{sd}, and g_{rd}, respectively.

Using the above definitions, the received signal at the destination in the first stage can be written as

$$\mathbf{y}_{\mathrm{sd}} = \sqrt{g_{\mathrm{sd}} P_s}\,\mathbf{G}\mathbf{s} + \mathbf{w}_1, \tag{1}$$

and the received signal at the relays in the first stage can be written as

$$\mathbf{y}_{\mathrm{sr}} = \sqrt{g_{\mathrm{sr}} P_s}\,\mathbf{H}\mathbf{s} + \mathbf{z}. \tag{2}$$

If an amplify-and-forward protocol [16] is used, the received signal at the destination in the second stage is given by

$$\mathbf{y}_{\mathrm{rd}} = \sqrt{\frac{g_{\mathrm{rd}}\, g_{\mathrm{sr}} P_r P_s}{P_0}}\,\mathbf{F}\mathbf{H}\mathbf{s} + \sqrt{\frac{g_{\mathrm{rd}} P_r}{P_0}}\,\mathbf{F}\mathbf{z} + \mathbf{w}_2, \tag{3}$$

where

$$P_0 = \frac{g_{\mathrm{sr}} P_s}{K}\mathrm{trace}\left(\mathbf{H}\mathbf{H}^H\right) + \sigma_n^2, \tag{4}$$

namely, the average received power at the relay nodes, which is used to normalize the received signal at the relay nodes so that the average relays transmit power equals P_r. To see this,

we can deduce the transmitted signal at the relays, which is given by

$$\mathbf{t}_{\mathrm{rd}} = \sqrt{\frac{g_{\mathrm{sr}}P_rP_s}{P_0}}\mathbf{H}\mathbf{s} + \sqrt{\frac{P_r}{P_0}}\mathbf{z}. \tag{5}$$

Then, the average transmit power is given by

$$\begin{aligned}\frac{1}{K}\mathrm{trace}[E[\mathbf{t}_{\mathrm{rd}}\mathbf{t}_{\mathrm{rd}}^H]] &= \frac{1}{K}\mathrm{trace}\left[\frac{g_{\mathrm{sr}}P_rP_s}{P_0}\mathbf{H}\mathbf{H}^H + \frac{P_r\sigma_n^2}{P_0}\mathbf{I}\right]\\ &= \frac{P_r}{KP_0}\mathrm{trace}[g_{\mathrm{sr}}P_s\mathbf{H}\mathbf{H}^H + \sigma_n^2\mathbf{I}]\\ &= P_r,\end{aligned} \tag{6}$$

where the last equation is due to (4).

Combining the received signal in the first and second stages, the total received signal at the destination is a $2M$-vector:

$$\mathbf{y} = \mathbf{T}\mathbf{s} + \mathbf{w}, \tag{7}$$

where

$$\begin{aligned}\mathbf{T} &= \begin{pmatrix}\sqrt{g_{\mathrm{sd}}P_s}\mathbf{G}\\ \sqrt{\frac{g_{\mathrm{sr}}g_{\mathrm{rd}}P_rP_s}{P_0}}\mathbf{F}\mathbf{H}\end{pmatrix},\\ \mathbf{w} &= \begin{pmatrix}\mathbf{w}_1\\ \sqrt{\frac{g_{\mathrm{rd}}P_r}{P_0}}\mathbf{F}\mathbf{z} + \mathbf{w}_2\end{pmatrix}.\end{aligned} \tag{8}$$

The sum capacity of this system is given by

$$\begin{aligned}&C_{\mathrm{sum}}\\ &= \log\det\left(\mathbf{I} + \mathbf{T}^H E^{-1}[\mathbf{w}\mathbf{w}^H]\mathbf{T}\right)\\ &= \log\det\left[\mathbf{I} + \left(\sqrt{g_{\mathrm{sd}}P_s}\mathbf{G}^H, \sqrt{\frac{g_{\mathrm{sr}}g_{\mathrm{rd}}P_rP_s}{P_0}}\mathbf{H}^H\mathbf{F}^H\right)\right.\\ &\qquad\left.\times\begin{pmatrix}\sigma_n^2\mathbf{I} & 0\\ 0 & \sigma_n^2\left(\mathbf{I} + \frac{g_{\mathrm{rd}}P_r}{P_0}\mathbf{F}\mathbf{F}^H\right)\end{pmatrix}^{-1}\begin{pmatrix}\sqrt{g_{\mathrm{sd}}P_s}\mathbf{G}\\ \sqrt{\frac{g_{\mathrm{sr}}g_{\mathrm{rd}}P_rP_s}{P_0}}\mathbf{F}\mathbf{H}\end{pmatrix}\right]\\ &= \log\det\left[\mathbf{I} + \frac{g_{\mathrm{sd}}P_s}{\sigma_n^2}\mathbf{G}^H\mathbf{G}\right.\\ &\qquad\left. + \frac{g_{\mathrm{sr}}g_{\mathrm{rd}}P_rP_s}{P_0\sigma_n^2}\mathbf{H}^H\mathbf{F}^H\left(\mathbf{I} + \frac{g_{\mathrm{rd}}P_r}{P_0}\mathbf{F}\mathbf{F}^H\right)^{-1}\mathbf{F}\mathbf{H}\right]\\ &= \log\det\left[\mathbf{I} + \gamma_1\mathbf{G}^H\mathbf{G} + \beta\gamma_2\mathbf{H}^H\mathbf{F}^H\left(\mathbf{I} + \beta\mathbf{F}\mathbf{F}^H\right)^{-1}\mathbf{F}\mathbf{H}\right].\end{aligned} \tag{9}$$

Here $\gamma_1 \triangleq g_{\mathrm{sd}}P_s/\sigma_n^2$ and $\gamma_2 \triangleq g_{\mathrm{sr}}P_s/\sigma_n^2$ represent the SNRs of the source-to-destination and source-to-relay links, respectively, and $\beta \triangleq g_{\mathrm{rd}}P_r/P_0$ is the amplification ratio of the relay.

We use a simpler notation for (9), which is given by

$$C_{\mathrm{sum}} = \log\det(\mathbf{I} + \mathbf{\Omega}) = \log\det\left(\mathbf{I} + \mathbf{\Omega}_s + \mathbf{\Omega}_r\right), \tag{10}$$

where $\mathbf{\Omega}_s \triangleq \gamma_1\mathbf{G}^H\mathbf{G}$ corresponds to the direct channel from the source to the destination; and

$$\mathbf{\Omega}_r \triangleq \beta\gamma_2\mathbf{H}^H\mathbf{F}^H\left(\mathbf{I} + \beta\mathbf{F}\mathbf{F}^H\right)^{-1}\mathbf{F}\mathbf{H} \tag{11}$$

corresponds to the signal relayed to the destination by the relay nodes. On denoting the eigenvalues of the matrix $\mathbf{\Omega}$ by $\{\lambda_m^{\Omega}\}_{m=1,2,\ldots}$, the sum capacity C_{sum} can be written as

$$C_{\mathrm{sum}} = \sum_{m=1}^{M}\log\left(1 + \lambda_m^{\Omega}\right). \tag{12}$$

In the following sections, we obtain expressions or approximations for C_{sum} by studying the distribution of λ_m^{Ω}.

We are interested in the average channel capacity of the large relay network, which is defined as

$$C_{\mathrm{avg}} \triangleq \frac{1}{M}C_{\mathrm{sum}}. \tag{13}$$

In this paper, we focus on analyzing C_{avg} in the large system scenario, namely, $K, M \to \infty$ while $\alpha \triangleq M/K$ is held constant, which is similar to the large system analysis arising in the study of code division multiple access (CDMA) systems [20]. Therefore, we place the following assumption on C_{avg}.

Assumption 1.

$$C_{\mathrm{avg}} \longrightarrow E[\log\left(1 + \lambda^{\Omega}\right)], \text{ almost surely,} \tag{14}$$

where λ^{Ω} is a generic eigenvalue of Ω, as $K, M \to \infty$.

This assumption will be validated by the numerical result in Section 6, which shows that the variance of C_{avg} decreases to zero as K and M increase. In the remaining part of this paper, we consider C_{avg} to be a constant in the sense of the large system limit, unless noted otherwise.

3. BASICS OF LARGE RANDOM MATRIX THEORY

In this section, we provide some basics of random matrix theory, including the notions of noncrossing partitions, isomorphic decomposition, combinatorial convolution, and free cumulants, which provide analytical machinery for characterizing the average channel capacity when the system dimensions increase asymptotically.

3.1. Freeness

Below is the abstract definition of freeness, which is originated by Voiculescu [21–23].

Definition 1. Let $\mathcal{A}$ be a unital algebra equipped with a linear functional $\psi : \mathcal{A} \to \mathcal{C}$, which satisfies $\psi(1) = 1$. Let $p_1, \ldots, p_k$ be one-variable polynomials. We call elements $a_1, \ldots, a_m \in \mathcal{A}$ *free* if for all $i_1 \neq i_2 \neq \cdots \neq i_k$, we have

$$\psi[p_1(a_{i_1})\cdots p_k(a_{i_k})] = 0, \tag{15}$$

whenever

$$\psi[p_j(a_{i_j})] = 0, \quad \forall j = 1, \ldots, k. \tag{16}$$

In the theory of large random matrices, we can consider random matrices as elements $a_1, \ldots, a_m$, and the linear functional ψ maps a random matrix $\mathbf{A}$ to the expectation of eigenvalues of $\mathbf{A}$.

3.2. Noncrossing partitions

A *partition* of a set $\{1, \ldots, p\}$ is defined as a division of the elements into a group of disjoint subsets, or *blocks* (a block is termed an i-block when the block size is i). A partition is called an r-partition when the number of blocks is r.

We say that a partition of a p-set is *noncrossing* if, for any two blocks $\{u_1, \ldots, u_s\}$ and $\{v_1, \ldots, v_t\}$, we have

$$u_k < v_1 < u_{k+1} \Longleftrightarrow u_k < v_t < u_{k+1}, \quad \forall k = 1, \ldots, s, \tag{17}$$

with the convention that $u_{s+1} = u_1$. For example, for the set $\{1, 2, 3, 4, 5, 6, 7, 8\}$, $\{\{1, 4, 5, 6\}, \{2, 3\}, \{7\}, \{8\}\}$ is noncrossing, while $\{\{1, 3, 4, 6\}, \{2, 5\}, \{7\}, \{8\}\}$ is not. We denote the set of noncrossing partitions on the set $\{1, 2, \ldots, p\}$ by $\mathbb{NC}_p$.

3.3. Isomorphic decomposition

The set of noncrossing partitions in $\mathbb{NC}_p$ has a partial ordering structure, in which $\pi \leq \sigma$ if each block of π is a subset of a corresponding block of σ. Then, for any $\pi \leq \sigma \in \mathbb{NC}_p$, we define the interval between π and σ as

$$[\pi, \sigma] \triangleq \{\psi \in \mathbb{NC}_p \mid \pi \leq \psi \leq \sigma\}. \tag{18}$$

It is shown in [21] that, for all $\pi \leq \sigma \in \mathbb{NC}_p$, there exists a canonical sequence of positive integers $\{k_i\}_{i \in \mathbb{N}}$ such that

$$[\pi, \sigma] \cong \prod_{j \in \mathbb{N}} \mathbb{NC}_j^{k_j}, \tag{19}$$

where $\cong$ is an isomorphism (the detailed mapping which can be found in the proof of Proposition 1 in [21]), the product is the Cartesian product, and $\{k_j\}_{j \in \mathbb{N}}$ is called the *class* of $[\pi, \sigma]$.

3.4. Incidence algebra, multiplicative function, and combinatorial convolution

The *incidence algebra* on the partial ordering structure of $\mathbb{NC}_p$ is defined as the set of all complex-valued functions $f(\psi, \sigma)$ with the property that $f(\psi, \sigma) = 0$ if $\psi \not\leq \sigma$ [20].

The *combinatorial convolution* between two functions f and g in the incidence algebra is defined as

$$f \star g(\pi, \sigma) \triangleq \sum_{\pi \leq \psi \leq \sigma} f(\pi, \psi) g(\psi, \sigma), \quad \forall \pi \leq \sigma. \tag{20}$$

An important subset of the incidence algebra is the set of *multiplicative functions* f on $[\pi, \sigma]$, which are defined by the property

$$f(\pi, \sigma) \triangleq \prod_{j \in \mathbb{N}} a_j^{k_j}, \tag{21}$$

where $\{a_j\}_{j \in \mathbb{N}}$ is a series of constants associated with f, and the class of $[\pi, \sigma]$ is $\{k_j\}_{j \in \mathbb{N}}$. We denote by f_a the multiplicative function with respect to $\{a_j\}_{j \in \mathbb{N}}$. An important function in the incidence algebra is the zeta function ζ, which is defined as

$$\zeta(\pi, \sigma) \triangleq \begin{cases} 1, & \text{if } \psi \leq \sigma, \\ 0, & \text{else.} \end{cases} \tag{22}$$

Further, the unit function I on the incidence algebra is defined as

$$I(\pi, \sigma) \triangleq \begin{cases} 1, & \text{if } \psi = \sigma, \\ 0, & \text{else.} \end{cases} \tag{23}$$

The inverse of the ζ function, denoted by μ, with respect to combinatorial convolution, namely, $\mu \star \zeta = I$, is termed the *Möbius* function.

3.5. Moments and free cumulants

Denote the pth moment of the (random) eigenvalue λ by $m_p \triangleq E[\lambda^p]$. We introduce a family of quantities termed *free cumulants* [22] denoted by $\{k_p\}$ for Ω where $p$denotes the order. We will use a superscript to indicate the matrix for which the moments and free cumulants are defined. The relationship between moments and free cumulants is given by combinatorial convolution in the incidence algebra [21, 22], namely,

$$\begin{aligned} f_m &= f_k \star \zeta, \\ f_k &= f_m \star \mu, \end{aligned} \tag{24}$$

where the multiplicative functions f_m (characterizing the moments), f_k (characterizing the free cumulants), zeta function ζ, Möbius function μ, and combinatorial convolution $\star$ are defined above.

By applying the definition of a noncrossing partition, (24), can be translated into the following explicit forms for the first three moments and free cumulants:

$$\begin{aligned} m_1 &= k_1, \\ m_2 &= k_2 + k_1^2, \\ m_3 &= k_3 + 3k_1k_2 + k_1^3, \\ k_1 &= m_1, \\ k_2 &= m_2 - m_1^2, \\ k_3 &= m_3 - 3m_1m_2 + 2m_1^3. \end{aligned} \tag{25}$$

The following lemma provides the rules for the addition [22] (see (B.4)) and product [22] (see (D.9)) of two free matrices.

Lemma 1. *If matrices A and B are mutually free, one has*

$$f_{k^{A+B}} = f_{k^A} + f_{k^B}, \tag{26}$$

$$f_{k^{AB}} = f_{k^A} \star f_{k^B}. \tag{27}$$

4. ANALYSIS USING RANDOM MATRIX THEORY

It is difficult to obtain a closed-form expression for the asymptotic average capacity C_{avg} in (13). In this section, using the theory of random matrices introduced in the last section, we first analyze the random variable λ^{Ω} by characterizing its moments and providing an upper bound for C_{avg}. Then, we can rewrite C_{avg} in terms of a moment series, which facilitates the approximation.

4.1. Moment analysis of λ^{Ω}

In contrast to [16], we analyze the random variable C_{avg} via its moments, instead of its distribution function, because moment analysis is more mathematically tractable. For simplicity, we denote $\beta\mathbf{F}^H(\mathbf{I}+\beta\mathbf{F}\mathbf{F}^H)^{-1}\mathbf{F}$ by $\mathbf{\Gamma}$, which is obviously Hermitian. Then, the matrix $\mathbf{\Omega}$ is given by

$$\mathbf{\Omega} = \gamma_1\mathbf{G}^H\mathbf{G} + \gamma_2\mathbf{H}^H\mathbf{\Gamma}\mathbf{H}. \tag{28}$$

In order to apply free probability theory, we need as a prerequisite that $\mathbf{G}^H\mathbf{G}$, $\mathbf{H}^H\mathbf{H}$, and $\mathbf{F}^H(\mathbf{I}+\beta\mathbf{F}\mathbf{F}^H)^{-1}\mathbf{F}$ be mutually free (the definition of freeness can be found in [23]). It is difficult to prove the freeness directly. However, the following proposition shows that the result obtained from the freeness assumption coincides with [24, Theorem 1.1] (same as in (29)) in [24], which is obtained via an alternative approach.

Proposition 1. *Suppose $\gamma_1 = \gamma_2 = 1$ (note that the assumption $\gamma_1 = \gamma_2 = 1$ is for convenience of analysis; it is straightforward to extend the proposition to general cases). Based on the freeness assumption, the Stieltjes transform of the eigenvalues in the matrix Ω satisfies the following Marcĕnko-Pastur equation:*

$$m_{\Omega}(z) = m_{G^HG}\left[z - \frac{1}{\alpha}\int\frac{\tau d\mathcal{F}(\tau)}{1+\tau(z)m_{\Omega}(z)}\right], \tag{29}$$

where $\mathcal{F}$ is the probability distribution function of the eigenvalues of the matrix $\mathbf{\Gamma}$, and $m(z)$ denotes the Stieltjes transform [20].

Proof. See Appendix A. □

Therefore, we assume that these matrices are mutually free (the *freeness assumption*) since this assumption yields the same result as a rigorously proved conclusion. The validity of the assumption is also supported by numerical results included in Section 6. Note that the reason why we do not apply the conclusion in Proposition 1 directly is that it is easier to manipulate using the moments and free probability theory.

Using the notion of multiplicative functions and Lemma 1, the following proposition characterizes the free cumulants of the matrix Ω, based upon which we can compute the eigenvalue moments of Ω from (24) (or (25) explicitly for the first three moments).

Proposition 2. *The free cumulants of the matrix Ω in (28) are given by*

$$f_{k^{\Omega}} = f_{k^{\Omega_s}} + (((f_{k^{\Gamma}} \star f_{k^{\tilde{H}}}) \star \zeta) \star \delta_{1/\alpha}) \star \mu, \tag{30}$$

where $k_p^{\Omega_s} = 1$, the free cumulant of $k_p^{\tilde{H}} = \gamma_2^p/\alpha$, $\forall p \in \mathcal{N}$, $\tilde{H} = \gamma_2\mathbf{H}\mathbf{H}^H$, and the multiplicative function $\delta_{1/\alpha}$ is defined as

$$\delta_{1/\alpha}(\tau,\pi) = \begin{cases} \dfrac{1}{\alpha}, & \text{if } \tau = \pi; \\ 0, & \text{if } \tau \neq \pi. \end{cases} \tag{31}$$

Proof. The proof is straightforward by applying the relationship between free cumulants and moments. The reasoning is given as follows:

(i) $f_{k^{\Gamma}} \star f_{k^{\tilde{H}}}$ represents the free cumulants of the matrix $\gamma_2\mathbf{\Gamma}\mathbf{H}\mathbf{H}^H$ (applying Lemma 1);
(ii) $(f_{k^{\Gamma}} \star f_{k^{\tilde{H}}}) \star \zeta$ represents the moments of the matrix $\gamma_2\mathbf{\Gamma}\mathbf{H}\mathbf{H}^H$;
(iii) $((f_{k^{\Gamma}} \star f_{k^{\tilde{H}}}) \star \zeta) \star \delta_{1/\alpha}$ represents the moments of the matrix $\gamma_2\mathbf{H}^H\mathbf{\Gamma}\mathbf{H}$;
(iv) $(((f_{k^{\Gamma}} \star f_{k^{\tilde{H}}}) \star \zeta) \star \delta_{1/\alpha}) \star \mu$ represents the free cumulants of the matrix $\gamma_2\mathbf{H}^H\mathbf{\Gamma}\mathbf{H}$;
(v) the final result is obtained by applying Lemma 1.

□

4.2. Upper bound of average capacity

Although in Section 4.1 we obtained all moments of λ^{Ω}, we did not obtain an explicit expression for the average channel capacity. However, we can provide an upper bound on this quantity by applying Jensen's inequality, which we summarize in the following proposition.

Proposition 3. *The average capacity satisfies*

$$C_{\text{avg}}^{(u)} \leq \log\left(1+\gamma_1+\frac{\alpha\beta\gamma_2}{\alpha+\beta}\right). \tag{32}$$

Proof. By applying Jensen's inequality, we have

$$\begin{aligned} E[\log(1+\lambda^{\Omega})] &\leq \log(1+E[\lambda^{\Omega}]) \\ &= \log(1+E[\lambda^{\Omega_s}]+E[\lambda^{\Omega_r}]). \end{aligned} \tag{33}$$

From [20], we obtain

$$E[\lambda^{\Omega_s}] = \gamma_1. \tag{34}$$

For $\mathbf{\Omega}$, we can show

$$E[\lambda^{\Omega_r}] = \frac{1}{\alpha}E[\lambda^{\Omega_r'}], \tag{35}$$

where

$$\Omega_r' = \beta\gamma_2\mathbf{F}^H(\mathbf{I}+\beta\mathbf{F}\mathbf{F}^H)^{-1}\mathbf{F}\mathbf{H}\mathbf{H}^H. \tag{36}$$

By applying the law of matrix product in Lemma 1, we can further simplify (35) to

$$E[\lambda^{\Omega_r}] = \frac{\gamma_2}{\alpha}E[\lambda^{HH^H}]E[\lambda^{\Gamma}] = \gamma_2 E[\lambda^{HH^H}]E\left[\frac{\beta\lambda^{FF^H}}{1+\beta\lambda^{FF^H}}\right]. \tag{37}$$

By applying Jensen's inequality again, we have

$$E[\lambda^{\Omega_r'}] \le \gamma_2 E[\lambda^{HH^H}] \frac{\beta E[\lambda^{FF^H}]}{1+\beta E[\lambda^{FF^H}]} = \frac{\alpha\beta\gamma_2}{\alpha+\beta}, \tag{38}$$

where we have applied the facts $E[\lambda^{HH^H}] = \alpha$ and $E[\lambda^{FF^H}] = 1/\alpha$.

Combining the above equations yields the upper bound in (32). □

4.3. Expansion of average capacity

In addition to providing an upper bound on the average capacity, we can also expand C_{avg} into a power series so that the moment expressions obtained from Proposition 2 can be applied. Truncating this power series yields approximations for the average capacity.

In particular, by applying a Taylor series expansion around a properly chosen constant x_0, C_{avg} can be written as

$$C_{\text{avg}} = \log(1+x_0) + \sum_{k=1}^{\infty} (-1)^{k-1} E\left[\frac{(\lambda - x_0)^k}{k(1+x_0)^k}\right]. \tag{39}$$

Taking the first two terms of the series yields the approximation

$$C_{\text{avg}} \approx \log(1+x_0) + \frac{m_1 - x_0}{1+x_0} - \frac{m_2 - 2x_0 m_1 + x_0^2}{2(1+x_0)^2}. \tag{40}$$

We can set $x_0 = \gamma_1 + \alpha\beta\gamma_2/(\alpha+\beta)$, which is an upper bound for $E[\lambda^{\Omega}]$ as shown in Proposition 3. We can also set $x_0 = 0$ and obtain an approximation when λ^{Ω} is small.Equations (40) will be a useful approximation for C_{avg} in Sections 5.2 and 5.3 when β is large or small or when SNR is small.

5. APPROXIMATIONS OF C_{avg}

In this section, we provide explicit approximations to C_{avg} for several special cases of interest. The difficulty in computing C_{avg} lies in determining the moments of the matrix $\mathbf{\Gamma}$. Therefore, in the low SNR region (Section 5.1), we consider representing C_{avg} in terms of the average capacities of the source-destination link and the source-relay-destination link. Then, we consider the region of high (Section 5.2) or low β (Section 5.3), where $\mathbf{\Gamma}$ can be simplified; thus we will obtain approximations in terms of α, β, γ_1, and γ_2. Finally, higher-order approximation will be studied in Section 5.4.

5.1. Approximate analysis in the low SNR regime

Unlike Section 4 which deals with general cases, we assume here that both the source-to-destination and relay-to-destination links within the low SNR regime, that is, P_s/σ_n^2 and P_r/σ_n^2 are small. Such an assumption is reasonable when both source nodes and relay nodes are far away from the destination nodes.

Within the low-SNR assumption, the asymptotic average capacity can be expanded in the Taylor series expansion about $x_0 = 0$ in (40), which is given by

$$C_{\text{avg}} = E[\log(1+\lambda^{\Omega})] = \sum_{i=1}^{\infty} (-1)^{i+1} \frac{m_i^{\Omega}}{i}. \tag{41}$$

We denote the pth-order approximation of C_{avg} by

$$C_p = \sum_{i=1}^{p} (-1)^{i+1} \frac{m_i^{\Omega}}{i}, \tag{42}$$

which implies

$$m_i^{\Omega} = (-1)^{i+1} i (C_i - C_{i-1}). \tag{43}$$

We denote by $\{C_p^s\}$ and $\{C_p^r\}$ the average capacity approximations (the same as in (42)) for the source-destination link and the source-relay-destination link, respectively. Our target is to represent the average capacity approximations $\{C_p\}$ by using $\{C_p^s\}$ and $\{C_p^r\}$ under the low-SNR assumption, which reveals the mechanism of information combining of the two links.

By combining (25), (26), and (43), we can obtain

$$\begin{aligned} C_1 &= C_1^s + C_1^r, \\ C_2 &= C_2^s + C_2^r - C_1^s C_1^r, \\ C_3 &= C_3^s + C_3^r - C_1^s C_1^r + 4C_1^s C_1^r - 2C_1^s C_2^r - 2C_1^r C_2^s, \end{aligned} \tag{44}$$

where C_p^s and C_p^r denote the pth-order approximations of the average capacity of the source-destination link and the source-relay-destination link, respectively.

Equation (44) shows that, to a first-order approximation, the combined effect of the source-destination and source-relay-destination links is simply a linear addition of average channel capacities, when the low-SNR assumption holds. For the second-order approximation in (44), the average capacity is reduced by a nonlinear term $C_1^s C_1^r$. The third-order term in (44) is relatively complicated to interpret.

5.2. High β region

In the high β region, the relay-destination link has a better channel than that of the source-relay link. The following proposition provides the first two moments of the eigenvalues λ in Ω in this case.

Proposition 4. *As $\beta \to \infty$, the first two moments of the eigenvalues λ in Ω converge to*

$$\begin{aligned} m_1 &= \begin{cases} \gamma_1 + \alpha\gamma_2, & \text{if } \alpha \le 1, \\ \gamma_1 + \gamma_2, & \text{if } \alpha > 1, \end{cases} \\ m_2 &= \begin{cases} 2(\gamma_1^2 + \alpha\gamma_2^2 + \alpha\gamma_1\gamma_2), & \text{if } \alpha \le 1, \\ 2\gamma_1^2 + 2\gamma_1\gamma_2 + \gamma_2^2(1+\alpha), & \text{if } \alpha > 1. \end{cases} \end{aligned} \tag{45}$$

Proof. See Appendix B. □

5.3. Low β region

In the low β region, the source-relay link has a better channel than the relay-destination link does. Similar to the result of Section 5.2, the first two eigenvalue moments of Ω are provided in the following proposition, which can be used to approximate C_{avg} in (40).

Proposition 5. *Suppose $\beta\gamma_2 = D$. As $\beta \to 0$ and D remains a constant, the first two moments of the eigenvalues λ in Ω converge to*

$$\begin{aligned} m_1 &= \gamma_1 + D, \\ m_2 &= 2\gamma_1^2 + 2\gamma_1 D + D^2(\alpha + 2). \end{aligned} \tag{46}$$

Proof. See Appendix C. □

5.4. Higher-order approximations for high and low β regions

In the previous two subsections, taking a first order approximation of the matrix $\mathbf{\Gamma} = \beta\mathbf{F}^H(\mathbf{I} + \beta\mathbf{F}\mathbf{F}^H)^{-1}\mathbf{F}$ resulted in simple expressions for the moments. We can also consider higher-order approximations, which provide finer expressions for the moments. These results are summarized in the following proposition, a proof of which is given in Appendix D. Note that m_1 and m_2 denote the first-order approximations given in Propositions 4 and 5, and $\tilde{m}_1$ and $\tilde{m}_2$ denote the expressions after considering higher-order terms. Note that, when β is large, we do not consider the case $\alpha = 1$ since the matrix $\mathbf{F}\mathbf{F}^H$ is at a critical point in this case, that is, for any $\alpha < 1$, $\mathbf{F}\mathbf{F}^H$ is of full rank almost surely; for any $\alpha > 1$, $\mathbf{F}\mathbf{F}^H$ is singular.

Proposition 6. *For sufficiently small β, one has*

$$\begin{aligned} \tilde{m}_1 &= m_1 - \gamma_2\beta^2\left(1 + \frac{1}{\alpha}\right) + o(\beta^2), \\ \tilde{m}_2 &= m_2 - 2\gamma_2\beta^2(\gamma_1 + \beta\gamma_2)\left(1 + \frac{1}{\alpha}\right) + o(\beta^2). \end{aligned} \tag{47}$$

For sufficiently large β and $\alpha < 1$, one has

$$\begin{aligned} \tilde{m}_1 &= m_1 - \frac{\gamma_2\alpha^2}{\beta(1-\alpha)} + o\left(\frac{1}{\beta}\right), \\ \tilde{m}_2 &= m_2 - \frac{2\gamma_2\alpha^2(\gamma_1 + \alpha\gamma_2)}{\beta(1-\alpha)} + o\left(\frac{1}{\beta}\right). \end{aligned} \tag{48}$$

For sufficiently large β and $\alpha > 1$, one has

$$\begin{aligned} \tilde{m}_1 &= m_1 - \frac{\alpha\gamma_2}{\beta(\alpha-1)} + o\left(\frac{1}{\beta}\right), \\ \tilde{m}_2 &= m_2 - \frac{2\gamma_2\alpha(\gamma_1 + \gamma_2)}{\beta(\alpha-1)} + o\left(\frac{1}{\beta}\right). \end{aligned} \tag{49}$$

Proof. See Appendix D. □

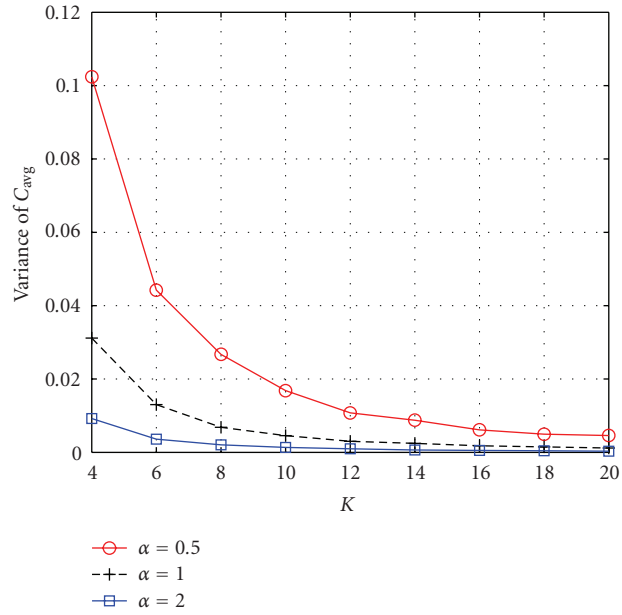


FIGURE 2: Variance of C_{avg} versus different K.

6. SIMULATION RESULTS

In this section, we provide simulation results to validate the analytical results derived in the previous sections. Figure 2 shows the variance of C_{avg} normalized by $E^2[C_{\text{avg}}]$ versus K. The configuration used here is $\gamma_1 = 1$, $\gamma_2 = 10$, $\beta = 1$, and $\alpha = 0.5/1/2$. For each value of K, we obtain the variance of C_{avg} by averaging over 1000 realizations of the random matrices, in which the elements are mutually independent complex Gaussian random variables. We can observe that the variance decreases rapidly as K increases. When K is larger than 10, the variance of C_{avg} is very small. This supports the validity of Assumption 1.

In the following simulations, we fix the value of K to be 40. All accurate values of average capacities C_{avg} are obtained from 1000 realizations of the random matrices. Again, the elements in these random matrices are mutually independent complex Gaussian random variables. All performance bounds and approximations are computed by the analytical results obtained in this paper.

Figure 3 compares the accurate average capacity obtained from (9) and the first three orders of approximation given in (44) with γ_1 ranging from 0.01 to 0.1. We set $\gamma_2 = \gamma_1$ and $\beta = 1$. From Figure 3, we observe that, in the low-SNR region, the approximations approach the correct values quite well. The reason is that the average capacity is approximately linear in the eigenvalues when SNR is small, which makes our expansions more precise. When the SNR becomes larger, the approximations can be used as bounds for the accurate values. (Notice that the odd orders of approximation provide upper bounds while the even ones provide lower bounds.)

In Figure 4, we plot the average capacity versus α, namely the ratio between the number of source nodes (or equivalently, destination nodes) and the number of relay nodes. The configuration is $\gamma_1 = 0.1$, $\gamma_2 = 10$, and $\beta = 10$.

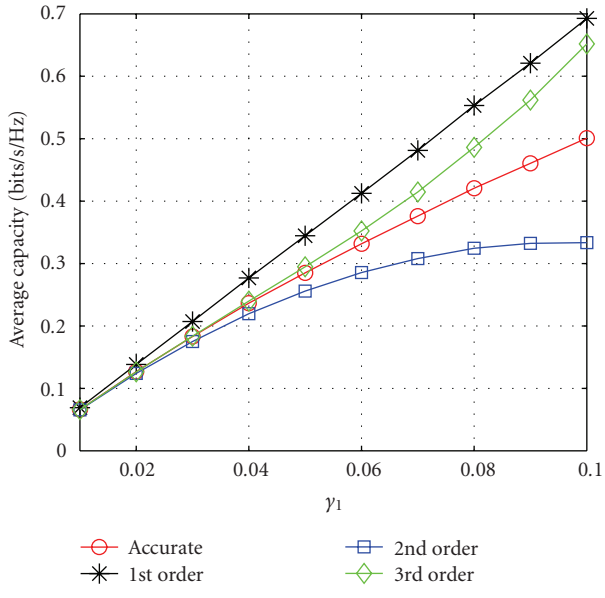


Figure 3: Comparison of different orders of approximation.

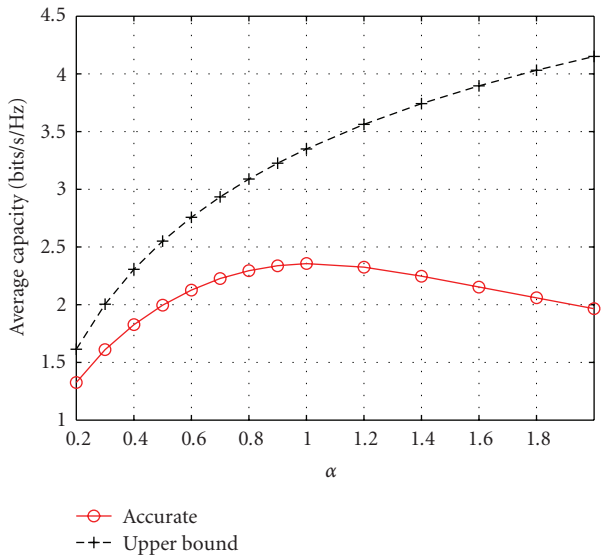


Figure 4: Performance versus various α.

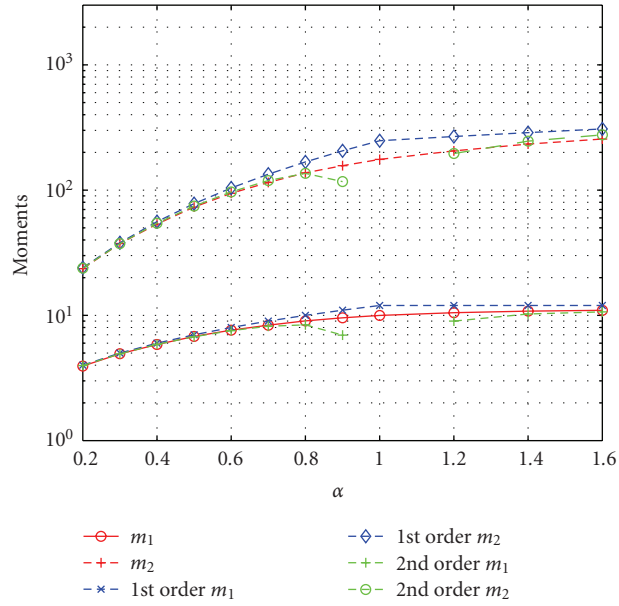


Figure 5: Eigenvalue moments versus various α in the high β region.

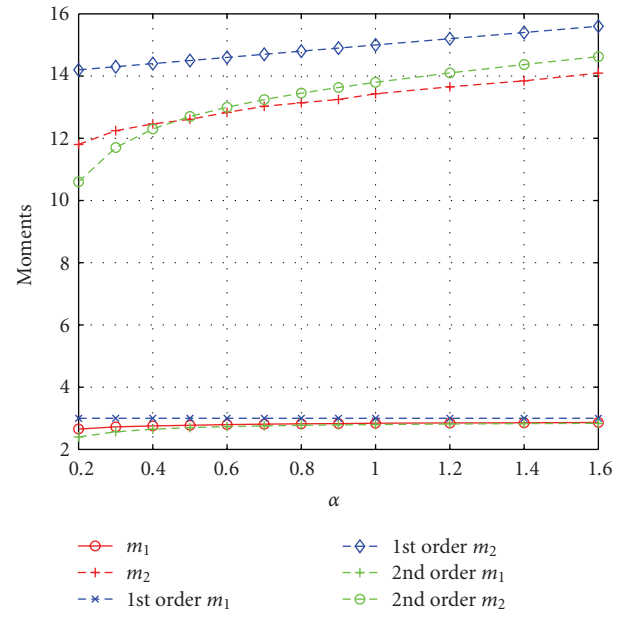


Figure 6: Eigenvalue moments versus various α in the low β region.

We observe that the average capacity achieves a maximum when $\alpha = 1$, namely, when using the same number of relay nodes as the source/destination nodes. A possible reason for this phenomenon is the normalization of elements in $\mathbf{H}$. (Recall that the variance of elements in $\mathbf{H}$ is $1/K$ such that the norms of column vectors in $\mathbf{H}$ are 1.) Now, suppose that M is fixed. When α is small, that is, K is large, the receive SNR at each relay node is small, which impairs the performance. When α is large, that is, K is small, we lose degrees of freedom. Therefore, $\alpha = 1$ achieves the optimal tradeoff. However, in practical systems, when the normalization is removed, it is always better to have more relay nodes if the corresponding cost is ignored. We also plot the upper bound in (32), which provides a loose upper bound here.

In Figures 5 and 6, we plot the precise values of m_1 and m_2 obtained from simulations and the corresponding first- and second-order approximations. The configuration is $\beta = 10$ (Figure 5) or $\beta = 0.1$ (Figure 6), $\gamma_1 = 2$ and $\gamma_2 = 10$. We can observe that the second-order approximation outperforms the first-order approximation except when α is close to 1 and β is large. (According to Proposition 6, the approximation diverges as $\alpha \to 1$ and $\beta \to \infty$.)

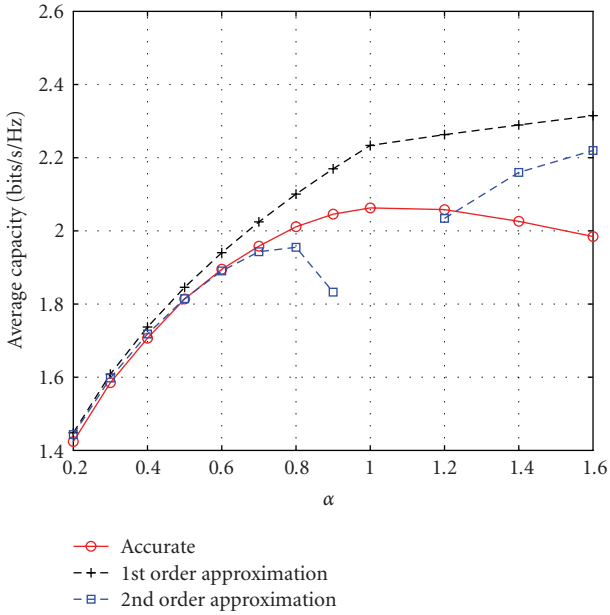


FIGURE 7: Performance versus various α in the high β region.

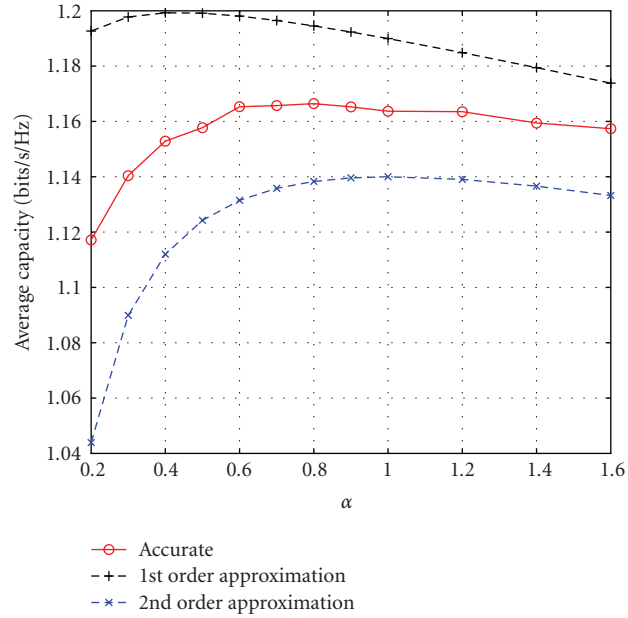


FIGURE 8: Performance versus various α in the low β region.

In Figure 7, we plot the average capacity versus α in the high β region, with configuration $\beta = 10$, $\gamma_1 = 2$, and $\gamma_2 = 10$. We can observe that the Taylor expansion provides a good approximation when α is small. Similar to Figure 7, the second-order approximation outperforms the first-order one except when α is close to 1. In Figure 8, we plot the average capacity versus α in the low β region. The configuration is the same as that in Figure 7 except that $\beta = 0.1$. We can observe that the Taylor expansion provides a good approximation for both small and large α. However, unlike the moment approximation, the error of the second-order approximation is not better than that of the first-order approximation. This is because (40) is also an approximation, and better approximation of the moments does not necessarily lead to a more precise approximation for the average capacity.

In Figure 9, we plot the ratio between the average capacity in (9) and the average capacity when the signal from the source to the destination in the first stage is ignored, as a function of the ratio γ_1/γ_2. We test four combinations of γ_2 and β. (Note that $\alpha = 0.5$.) We observe that the performance gain increases with the ratio γ_1/γ_2 (the channel gain ratio between source-destination link and source-relay link). The performance gain is substantially larger in the low-SNR regime ($\gamma_2 = 1$) than in the high-SNR regime ($\gamma_2 = 10$). When the amplification ratio β decreases, the performance gain is improved. Therefore, substantial performance gain is obtained by incorporating the source-destination link when the channel conditions of the source-destination link are comparable to those of the relay-destination link and the source-relay link, particularly in the low-SNR region. In other cases, we can simply ignore the source-destination link since it achieves marginal gain at the cost of having to process a high-dimensional signal.

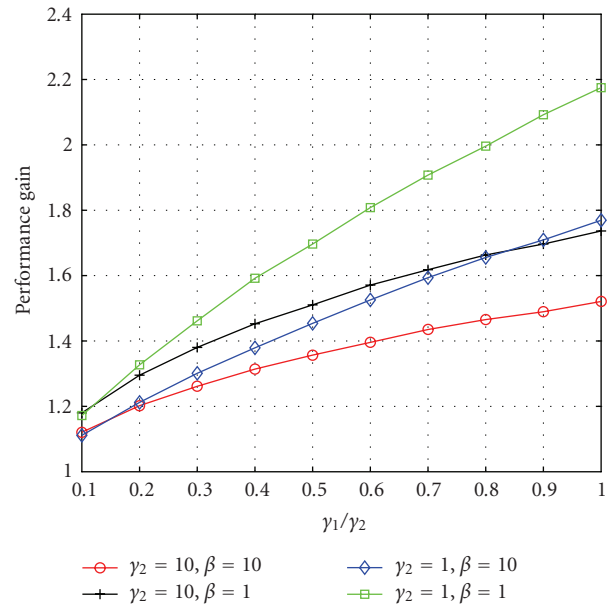


FIGURE 9: Performance gain by incorporating the source-destination link.

7. CONCLUSIONS

In this paper, we have used random matrix theory to analyze the asymptotic behavior of cooperative transmission with a large number of nodes. Compared to prior results of [23], we have considered the combination of relay and direct transmission, which is more complicated than considering relay transmission only. We have constructed a performance upper-bound for the low signal-to-noise-ratio regime, and

have derived approximations for high and low relay-to-destination link qualities, respectively. The key idea has been to investigate the eigenvalue distributions related to capacity and to analyze eigenvalue moments for large wireless networks. We have also conducted simulations which validate the analytical results. Particularly, the numerical simulation results show that incorporating the direct link between the source nodes and destination nodes can substantially improve the performance when the direct link is of high quality. These results provide useful tools and insights for the design of large cooperative wireless networks.

APPENDICES

A. PROOF OF PROPOSITION 1

We first define some useful generating functions and transforms [22], and then use them in the proof by applying some conclusions of free probability theory [23].

A.1. Generating functions and transforms

For simplicity, we rewrite the matrix $\mathbf{\Omega}$ as

$$\Omega = \mathbf{G}^H\mathbf{G} + \mathbf{\Xi}\mathbf{\Gamma}\mathbf{\Xi}^H, \tag{A.1}$$

where $\mathbf{\Xi} \triangleq (1/\alpha)\mathbf{H}^H$ is an $M \times K$ matrix, in which the elements are independent random variables with variance $1/M$.

For a large random matrix with eigenvalue moments $\{m_i\}_{i=1,2,\ldots}$ and free cumulants $\{k_j\}_{j=1,2,\ldots}$, we define the following generating functions:

$$\Lambda(z) = 1 + \sum_{i=1}^{\infty} m_i z^i, \qquad C(z) = 1 + \sum_{j=1}^{\infty} k_j z^j. \tag{A.2}$$

We define the Stieltjes transform

$$m(z) = E\left[\frac{1}{\lambda - z}\right], \tag{A.3}$$

where λ is a generic (random) eigenvalue.

We also define a "Fourier transform" given by

$$D(z) = \frac{1}{z}\left(C(z) - 1\right)^{\langle -1\rangle}, \tag{A.4}$$

which was originally defined in [25].

The following lemma provides some fundamental relations among the above functions and transforms.

Lemma 2. *For the generating functions and transforms in* (A.2)–(A.4)*, the following equations hold:*

$$\Lambda\left[\frac{zD(z)}{z+1}\right] = z + 1, \tag{A.5}$$

$$m\left[\frac{C(z)}{z}\right] = -z, \tag{A.6}$$

$$C(-m(z)) = -zm(z), \tag{A.7}$$

$$\Lambda(z) = -\frac{m(z^{-1})}{z}. \tag{A.8}$$

Note that we use subscripts to indicate the matrix for which the generating functions and transforms are defined. For example, for the matrix $\mathbf{M}$, the eigenvalue moment generating function is denoted by $\Lambda_{\mathbf{M}}(z)$.

A.2. Proof of Proposition 1

We first study the matrix $\mathbf{\Xi}\mathbf{\Gamma}\mathbf{\Xi}^H$ in (A.1). In order to apply the conclusions about matrix products, we can work on the matrix $\mathbf{J} = \mathbf{\Gamma}\mathbf{\Xi}^H\mathbf{\Xi}$ instead since we have the following lemma.

Lemma 3.

$$\Lambda_{\Xi\Gamma\Xi^H}(z) - 1 = \frac{1}{\alpha}\left(\Lambda_{\Gamma\Xi^H\Xi}(z) - 1\right). \tag{A.9}$$

Proof. For any $n \in \mathcal{N}$, we have

$$\begin{aligned}\frac{1}{M}\mathrm{trace}\left(\left(\Xi\Gamma\Xi^H\right)^n\right) &= \frac{1}{M}\mathrm{trace}\left(\left(\Gamma\Xi^H\Xi\right)^n\right)\\ &= \frac{K}{M}\frac{1}{K}\mathrm{trace}\left(\left(\Gamma\Xi^H\Xi\right)^n\right).\end{aligned} \tag{A.10}$$

Letting $K, M \to \infty$, we obtain

$$m_n^{\Xi\Gamma\Xi^H} = \frac{1}{\alpha}m_n^{\Gamma\Xi^H\Xi}. \tag{A.11}$$

Then, we have

$$\begin{aligned}\Lambda_{\Xi\Gamma\Xi^H}(z) - 1 &= \sum_{j=1}^{\infty} m_n^{\Xi\Gamma\Xi^H} z^n\\ &= \frac{1}{\alpha}\sum_{j=1}^{\infty} m_n^{\Gamma\Xi^H\Xi} z^n\\ &= \frac{1}{\alpha}\left(\Lambda_{\Gamma\Xi^H\Xi}(z) - 1\right).\end{aligned} \tag{A.12}$$

□

On denoting $\mathbf{\Xi}^H\mathbf{\Xi}$ by $\mathbf{B}$, the following lemma discloses the law of matrix product[22] and is equivalent to (27).

Lemma 4. *Based on the freeness assumption, for the matrix* $\mathbf{J} = \mathbf{\Gamma}\mathbf{B}$*, we have*

$$D_J(z) = D_\Gamma(z)D_B(z). \tag{A.13}$$

In order to use the "Fourier Transform," we need the following lemma.

Lemma 5. *For the matrix* $\mathbf{B}$*, we have*

$$D_B(z) = \frac{\alpha}{z + \alpha}. \tag{A.14}$$

Proof. Due to the definition of $\mathbf{\Xi}$, we have

$$\mathbf{\Xi}^H\mathbf{\Xi} = \frac{1}{\alpha}\mathbf{H}\mathbf{H}^H. \tag{A.15}$$

Then, it is easy to check that

$$\begin{aligned} m_n^{\Xi^H\Xi} &= \left(\frac{1}{\alpha}\right)^n m_n^{\mathbf{H}\mathbf{H}^H},\\ k_n^{\Xi^H\Xi} &= \left(\frac{1}{\alpha}\right)^n k_n^{\mathbf{H}\mathbf{H}^H},\end{aligned} \tag{A.16}$$

which is equivalent to

$$C_{\mathbf{\Xi}^H\mathbf{\Xi}}(z) = C_{\mathbf{H}\mathbf{H}^H}\left(\frac{z}{\alpha}\right). \tag{A.17}$$

By applying the conclusion in [20], all free cumulants in $\mathbf{H}\mathbf{H}^H$ are equal to α. Therefore,

$$C_{\mathbf{\Xi}^H\mathbf{\Xi}}(z) = C_{\mathbf{H}\mathbf{H}^H}(z) = 1 + \frac{\alpha z}{1-z}. \tag{A.18}$$

The conclusion follows from computing the inverse function of $C_{\mathbf{\Xi}^H\mathbf{\Xi}}(z) - 1 = \alpha z/(\alpha - z)$. □

The following lemma relates $\Lambda_{\mathbf{\Gamma}}(z)$ to $\mathcal{F}$. (Recall that $\mathcal{F}$ is the distribution of eigenvalues of the matrix $\mathbf{\Gamma}$.)

Lemma 6. *For the matrix* $\mathbf{\Gamma}$, *the following equation holds:*

$$\Lambda_{\mathbf{\Gamma}}(z) - 1 = \int \frac{\tau z}{1-\tau z} d\mathcal{F}(\tau). \tag{A.19}$$

Proof. Based on the definition of $\Lambda_{\mathbf{\Gamma}}(z)$, we have

$$\Lambda_{\mathbf{\Gamma}}(z) - 1 = \sum_{j=1}^{\infty} m_j z^j = \sum_{j=1}^{\infty} E[\lambda^j z^j] = E\left[\sum_{j=1}^{\infty} (\lambda z)^j\right] = E\left[\frac{\lambda z}{1-\lambda z}\right], \tag{A.20}$$

from which the conclusion follows. □

Based on the above lemmas, we can show the following important lemma.

Lemma 7. *Based on the freeness assumption, for the matrix* $\mathbf{\Xi}\mathbf{\Gamma}\mathbf{\Xi}^H$, *we have*

$$C_{\mathbf{\Xi}\mathbf{\Gamma}\mathbf{\Xi}^H}(z) = 1 + \frac{1}{\alpha}\int \frac{z\tau}{1-z\tau} d\mathcal{F}(\tau). \tag{A.21}$$

Proof. The lemma can be proved by showing the following series of equivalent equations:

$$C_{\mathbf{\Xi}\mathbf{\Gamma}\mathbf{\Xi}^H}(z) = 1 + \frac{1}{\alpha}\int \frac{z\tau}{1-z\tau} d\mathcal{F}(\tau) \tag{A.22}$$

$$\Longleftrightarrow m_{\mathbf{\Xi}\mathbf{\Gamma}\mathbf{\Xi}^H}(z) = \frac{1}{-z + (1/\alpha)\int(\tau/1+\tau m_{\mathbf{\Xi}\mathbf{\Gamma}\mathbf{\Xi}^H}(z))d\mathcal{F}(\tau)} \tag{A.23}$$

$$\Longleftrightarrow \Lambda_{\mathbf{\Xi}\mathbf{\Gamma}\mathbf{\Xi}^H}(z) = \frac{1}{1-(1/\alpha)\int(z\tau/1-\tau z\Lambda_{\mathbf{\Xi}\mathbf{\Gamma}\mathbf{\Xi}^H}(z))d\mathcal{F}(\tau)} \tag{A.24}$$

$$\Longleftrightarrow \Lambda_{\mathbf{\Xi}\mathbf{\Gamma}\mathbf{\Xi}^H}(z) - \frac{1}{\alpha}\int \frac{z\tau\Lambda_{\mathbf{\Xi}\mathbf{\Gamma}\mathbf{\Xi}^H}(z)}{1-\tau z\Lambda_{\mathbf{\Xi}\mathbf{\Gamma}\mathbf{\Xi}^H}(z)} d\mathcal{F}(\tau) = 1 \tag{A.25}$$

$$\Longleftrightarrow \Lambda_{\mathbf{\Xi}\mathbf{\Gamma}\mathbf{\Xi}^H}(z) - 1 = \frac{1}{\alpha}\left(\Lambda_{\mathbf{\Gamma}}(z\Lambda_{\mathbf{\Xi}\mathbf{\Gamma}\mathbf{\Xi}^H}(z)) - 1\right) \tag{A.26}$$

$$\Longleftrightarrow \Lambda_{\mathbf{\Gamma}\mathbf{\Xi}^H\mathbf{\Xi}}(z) = \Lambda_{\mathbf{\Gamma}}\left(z\left(\frac{1}{\alpha}\left(\Lambda_{\mathbf{\Gamma}\mathbf{\Xi}^H\mathbf{\Xi}}(z) - 1\right)\right) + 1\right) \tag{A.27}$$

$$\Longleftrightarrow z + 1 = \Lambda_{\mathbf{\Gamma}}\left(\frac{zD_{\mathbf{\Gamma}\mathbf{\Xi}^H\mathbf{\Xi}}(z)}{z+1}\left(\frac{1}{\alpha}z + 1\right)\right) \tag{A.28}$$

$$\Longleftrightarrow z + 1 = \Lambda_{\mathbf{\Gamma}}\left(\frac{zD_{\mathbf{\Gamma}}(z)}{z+1}\right). \tag{A.29}$$

The equivalence of the above equations is explained as follows:

(i) substituting (A.6) into (A.22) yields (A.23);
(ii) substituting (A.8) into (A.23) yields (A.24);
(iii) equations (A.25) and (A.26) are equivalent due to Lemma 6;
(iv) equations (A.26) and (A.27) are equivalent due to Lemma 3;
(v) equations (A.27) and (A.28) are equivalent by substituting $z = zD_{\mathbf{\Gamma}\mathbf{\Xi}^H\mathbf{\Xi}}(z)/(z+1)$ into (A.27) and applying (A.5);
(vi) equations (A.28) and (A.29) are equivalent due to Lemmas 4 and 5;
(vii) equation (A.29) holds due to (A.5). □

Based on Lemma 7, we can prove Proposition 1.

Proof. By applying (26) and the freeness assumption, we have

$$C_{\Omega}(z) = C_{G^HG}(z) + C_{\mathbf{\Xi}\mathbf{\Gamma}\mathbf{\Xi}^H}(z)(z) - 1, \tag{A.30}$$

which implies

$$\frac{C_{G^HG}(z)}{z} = \frac{C_{\Omega}(z)}{z} - \frac{C_{\mathbf{\Xi}\mathbf{\Gamma}\mathbf{\Xi}^H}(z)}{z} + \frac{1}{z}. \tag{A.31}$$

Taking both sides of (A.31) as arguments of $m_{G^HG}(z)$, we have

$$-z = m_{G^HG}\left(\frac{C_{\Omega}(z)}{z} - \frac{C_{\mathbf{\Xi}\mathbf{\Gamma}\mathbf{\Xi}^H}(z)}{z} + \frac{1}{z}\right), \tag{A.32}$$

where the left-hand side is obtained from (A.6).

Letting $z = -m_{\Omega}(t)$ in (A.32), we have

$$\begin{aligned} m_{\Omega}(t) &= m_{G^HG}\left(\frac{C_{\Omega}(-m(t))}{-m(t)} - \frac{1+(1/\alpha)\int(m_{\Omega}(t)\tau/(1+m_{\Omega}(t)\tau))d\mathcal{F}(\tau)}{-m_{\Omega}(t)} - \frac{1}{m_{\Omega}(t)}\right) \\ &= m_{G^HG}\left(t - \frac{1}{\alpha}\int \frac{\tau}{1+m_{\Omega}(t)\tau} d\mathcal{F}(\tau)\right), \end{aligned} \tag{A.33}$$

where the first equation is based on (A.7). □

B. PROOF OF PROPOSITION 4

Proof. We first consider the matrix $\mathbf{\Gamma}' = \beta(\mathbf{I} + \beta\mathbf{F}\mathbf{F}^H)^{-1}\mathbf{F}\mathbf{F}^H$. When $K \geq M$, it is easy to check that $\mathbf{F}\mathbf{F}^H$ is invertible almost surely since $\mathbf{F}$ is an $M \times K$ matrix. Then

$$\mathbf{\Gamma}' \longrightarrow \mathbf{I}, \tag{B.1}$$

as $\beta \to \infty$. Therefore, $m_p^{\mathbf{\Gamma}'} = 1$, $\forall\, p \in \mathcal{N}$.

When $K \le M$, let $\mathbf{F}\mathbf{F}^H = \mathbf{U}^H\mathbf{\Lambda}\mathbf{U}$, where $\mathbf{U}$ is unitary and $\mathbf{\Lambda}$ is diagonal. Then, we have

$$\begin{aligned} m_p^{\Gamma'} &= \frac{1}{M}\text{trace}\left[(\mathbf{\Gamma}')^p\right] \\ &= \frac{1}{M}\text{trace}\left[\beta(\mathbf{I}+\beta\mathbf{\Lambda})^{-p}\mathbf{\Lambda}^p\right] \\ &= \frac{K}{M}, \end{aligned} \tag{B.2}$$

where the last equation is due to the fact that only K elements in $\mathbf{\Lambda}$ are nonzero since $K \le M$. Therefore, $m_p^{\Gamma'} = 1/\alpha$, $\forall\, p \in \mathcal{N}$.

Applying the same argument as in Lemma 3, we obtain

$$m_p^{\Gamma} = \begin{cases} 1, & \text{if } K \le M, \\ \alpha, & \text{if } K \ge M, \end{cases} \quad \forall\, p \in \mathcal{N}, \tag{B.3}$$

which is equivalent to

$$\begin{aligned} k_1^{\Gamma} &= \begin{cases} 1, & \text{if } K \le M, \\ \alpha, & \text{if } K \ge M, \end{cases} \\ k_2^{\Gamma} &= \begin{cases} 0, & \text{if } K \le M, \\ \alpha - \alpha^2, & \text{if } K \ge M. \end{cases} \end{aligned} \tag{B.4}$$

Define $\mathbf{\Omega}_r' = \beta\mathbf{F}^H(\mathbf{I}+\beta\mathbf{F}\mathbf{F}^H)^{-1}\mathbf{F}\mathbf{H}\mathbf{H}^H$. Due to the law of the matrix product in Lemma 1, the free cumulants of $\mathbf{\Omega}_r'$ are given by

$$\begin{aligned} k_1^{\mathbf{\Omega}_r'} &= k_1^{\Gamma}k_1^{HH^H}, \\ k_2^{\mathbf{\Omega}_r'} &= k_2^{\Gamma}\left(k_1^{HH^H}\right)^2 + k_2^{HH^H}\left(k_1^{\Gamma}\right)^2. \end{aligned} \tag{B.5}$$

Then, combining (B.5), $k_1^{HH^H} = \alpha$ and $k_2^{HH^H} = \alpha$, we obtain

$$\begin{aligned} k_1^{\mathbf{\Omega}_r'} &= \begin{cases} \alpha^2, & \text{if } \alpha \le 1, \\ \alpha, & \text{if } \alpha \ge 1, \end{cases} \\ k_2^{\mathbf{\Omega}_r'} &= \begin{cases} 2\alpha^3 - \alpha^4, & \text{if } \alpha \le 1, \\ \alpha, & \text{if } \alpha \ge 1. \end{cases} \end{aligned} \tag{B.6}$$

which imply

$$\begin{aligned} m_1^{\mathbf{\Omega}_r'} &= \begin{cases} \alpha^2, & \text{if } \alpha \le 1, \\ \alpha, & \text{if } \alpha \ge 1, \end{cases} \\ m_2^{\mathbf{\Omega}_r'} &= \begin{cases} 2\alpha^3, & \text{if } \alpha \le 1, \\ \alpha + \alpha^2, & \text{if } \alpha \ge 1. \end{cases} \end{aligned} \tag{B.7}$$

Applying the same argument as in Lemma 3, we obtain

$$\begin{aligned} m_1^{\Omega_r} &= \begin{cases} \gamma_2\alpha, & \text{if } \alpha \le 1, \\ \gamma_2, & \text{if } \alpha \ge 1, \end{cases} \\ m_2^{\Omega_r} &= \begin{cases} 2\gamma_2^2\alpha^2, & \text{if } \alpha \le 1, \\ \gamma_2^2(1+\alpha), & \text{if } \alpha \ge 1. \end{cases} \end{aligned} \tag{B.8}$$

which is equivalent to

$$\begin{aligned} k_1^{\Omega_r} &= \begin{cases} \gamma_2\alpha, & \text{if } \alpha \le 1, \\ \gamma_2, & \text{if } \alpha \ge 1, \end{cases} \\ k_2^{\Omega_r} &= \begin{cases} \gamma_2^2\alpha^2, & \text{if } \alpha \le 1, \\ \gamma_2^2\alpha, & \text{if } \alpha \ge 1. \end{cases} \end{aligned} \tag{B.9}$$

The conclusion follows from the facts that $\forall\, p \in \mathcal{N}$, $k_p^{\Omega_s} = \gamma_1^p$ and $k_p^{\Omega} = k_p^{\Omega_s} + k_p^{\Omega_r}$. □

C. PROOF OF PROPOSITION 5

Proof. When $\beta \to 0$, we have (recall $D = \gamma_2\beta$)

$$\begin{aligned} \Omega &= \gamma_1\mathbf{G}^H\mathbf{G} + D\mathbf{H}^H\mathbf{F}^H\mathbf{F}\mathbf{H}, \\ k_1^{F^HF} &= 1, \\ k_2^{F^HF} &= \frac{1}{\alpha}, \\ k_1^{HH^H} &= \alpha, \\ k_2^{HH^H} &= \alpha. \end{aligned} \tag{C.1}$$

Then, applying (B.5), we obtain

$$\begin{aligned} k_1^{F^HFHH^H} &= \alpha, \\ k_2^{F^HFHH^H} &= 2\alpha, \end{aligned} \tag{C.2}$$

which is equivalent to

$$\begin{aligned} m_1^{F^HFHH^H} &= \alpha, \\ m_2^{F^HFHH^H} &= \alpha^2 + 2\alpha. \end{aligned} \tag{C.3}$$

Then, for matrix $\mathbf{H}^H\mathbf{F}^H\mathbf{F}\mathbf{H}$, we have

$$\begin{aligned} m_1^{H^HF^HFH} &= 1, \\ m_2^{H^HF^HFH} &= \alpha + 2, \end{aligned} \tag{C.4}$$

which results in

$$\begin{aligned} k_1^{H^HF^HFH} &= 1, \\ k_2^{H^HF^HFH} &= \alpha + 1. \end{aligned} \tag{C.5}$$

The remaining part of the proof is the same as the proof of Proposition 4 in Appendix B. □

D. PROOF OF PROPOSITION 6

We first prove the following lemma which provides the impact of perturbation on m_1^{Γ} and m_2^{Γ}. We use $\widetilde{X}$ to represent the perturbed version of the quantity X.

Lemma 8. *Suppose the first and second moments of the matrix* $\mathbf{\Gamma}$ *are perturbed by small* δ_1 *and* δ_2*, respectively, where* δ_1 *and* δ_2 *are of the same order* $O(\delta)$*, namely,*

$$\begin{aligned} \widetilde{m}_1^{\Gamma} &= m_1^{\Gamma} + \delta_1, \\ \widetilde{m}_2^{\Gamma} &= m_2^{\Gamma} + \delta_2. \end{aligned} \tag{D.1}$$

Then, we have

$$\widetilde{m}_1^{\Omega} = m_1^{\Omega} + \gamma_2\delta_1,$$

$$\widetilde{m}_2^{\Omega} = m_2^{\Omega} + \alpha\gamma_2^2\delta_2 + 2\gamma_2\left(k_1^{\mathbf{\Omega}_r'}\gamma_2 - m_1^{\Omega_r} + k_1^{\Omega} + (1-\alpha)k_1^{\Gamma}\gamma_2\right)\delta_1 + o(\delta), \tag{D.2}$$

where

$$\mathbf{\Omega}_r' = \beta\mathbf{F}^H\left(\mathbf{I} + \beta\mathbf{F}\mathbf{F}^H\right)^{-1}\mathbf{F}\mathbf{H}\mathbf{H}^H. \tag{D.3}$$

Proof. We begin from $\widetilde{k}_1^{\Gamma}$ and $\widetilde{k}_2^{\Gamma}$. Suppose small perturbations ϵ_1 and ϵ_2, which are both of order $O(\epsilon)$, are placed on k_1^{Γ} and k_2^{Γ}, namely,

$$\begin{aligned} \widetilde{k}_1^{\Gamma} &= k_1^{\Gamma} + \epsilon_1, \\ \widetilde{k}_2^{\Gamma} &= k_2^{\Gamma} + \epsilon_2. \end{aligned} \tag{D.4}$$

We have

$$\begin{aligned} \widetilde{k}_1^{\mathbf{\Omega}_r'} &= k_1^{\mathbf{\Omega}_r'} + \alpha\epsilon_1, \\ \widetilde{k}_2^{\mathbf{\Omega}_r'} &= k_2^{\mathbf{\Omega}_r'} + \alpha^2\epsilon_2 + 2\alpha k_1^{\Gamma}\epsilon_1 + o(\epsilon), \end{aligned} \tag{D.5}$$

which implies

$$\begin{aligned} \widetilde{m}_1^{\mathbf{\Omega}_r'} &= m_1^{\mathbf{\Omega}_r'} + \alpha\epsilon_1, \\ \widetilde{m}_2^{\mathbf{\Omega}_r'} &= m_2^{\mathbf{\Omega}_r'} + \alpha^2\epsilon_2 + 2\alpha\left(k_1^{\Gamma} + k_1^{\mathbf{\Omega}_r'}\right)\epsilon_1 + o(\epsilon). \end{aligned} \tag{D.6}$$

For $\mathbf{\Omega}_r = \gamma_2\beta\mathbf{H}^H\mathbf{F}^H\left(\mathbf{I} + \beta\mathbf{F}\mathbf{F}^H\right)^{-1}\mathbf{F}\mathbf{H}$, we have

$$\begin{aligned} \widetilde{m}_1^{\Omega_r} &= m_1^{\Omega_r} + \gamma_2\epsilon_1, \\ \widetilde{m}_2^{\Omega_r} &= m_2^{\Omega_r} + \alpha\gamma_2^2\epsilon_2 + 2\gamma_2^2\left(k_1^{\Gamma} + k_1^{\mathbf{\Omega}_r'}\right)\epsilon_1 + o(\epsilon), \end{aligned} \tag{D.7}$$

which implies that we have

$$\begin{aligned} \widetilde{k}_1^{\Omega_r} &= k_1^{\Omega_r} + \gamma_2\epsilon_1, \\ \widetilde{k}_2^{\Omega_r} &= k_2^{\Omega_r} + \alpha\gamma_2^2\epsilon_2 + 2\gamma_2\left(k_1^{\Gamma}\gamma_2 + k_1^{\mathbf{\Omega}_r'}\gamma_2 - m_1^{\Omega_r}\right)\epsilon_1 + o(\epsilon). \end{aligned} \tag{D.8}$$

Then, for $\mathbf{\Omega}$, we have

$$\begin{aligned} \widetilde{k}_1^{\Omega} &= k_1^{\Omega} + \gamma_2\epsilon_1, \\ \widetilde{k}_2^{\Omega} &= k_2^{\Omega} + \alpha\gamma_2^2\epsilon_2 + 2\gamma_2\left(k_1^{\Gamma}\gamma_2 + k_1^{\mathbf{\Omega}_r'}\gamma_2 - m_1^{\Omega_r}\right)\epsilon_1 + o(\epsilon), \end{aligned} \tag{D.9}$$

which implies

$$\begin{aligned} \widetilde{m}_1^{\Omega} &= m_1^{\Omega} + \gamma_2\epsilon_1, \\ \widetilde{m}_2^{\Omega} &= m_2^{\Omega} + \alpha\gamma_2^2\epsilon_2 + 2\gamma_2\left(k_1^{\Gamma}\gamma_2 + k_1^{\mathbf{\Omega}_r'}\gamma_2 - m_1^{\Omega_r} + k_1^{\Omega}\right)\epsilon_1 + o(\epsilon). \end{aligned} \tag{D.10}$$

Now, we compute ϵ_1 and ϵ_2. Equation (D.1) implies

$$\begin{aligned} \widetilde{k}_1^{\Gamma} &= k_1^{\Gamma} + \delta_1, \\ \widetilde{k}_2^{\Gamma} &= k_2^{\Gamma} + \delta_2 - 2m_1^{\Gamma}\delta_1 + o(\delta), \end{aligned} \tag{D.11}$$

which is equivalent to

$$\begin{aligned} \epsilon_1 &= \delta_1, \\ \epsilon_2 &= \delta_2 - 2m_1^{\Gamma}\delta_1. \end{aligned} \tag{D.12}$$

Combining (D.10) and (D.12), we obtain (D.2). □

Based on Lemma 8, we can obtain the following lemma, where δ_1 and δ_2 are defined the same as in Lemma 8. The proof is straightforward by applying the intermediate results in the proofs of Propositions 4 and 5.

Lemma 9. *For sufficiently high* β*,* (D.2) *is equivalent to*

$$\widetilde{m}_1^{\Omega} = m_1^{\Omega} + \gamma_2\delta_1,$$

$$\widetilde{m}_2^{\Omega} = m_2^{\Omega} + \alpha\gamma_2^2\delta_2 + 2\gamma_2\left(\alpha\gamma_2 + \gamma_1\right)\delta_1 + o(\delta), \quad \textit{when } \alpha \le 1, \tag{D.13}$$

or

$$\widetilde{m}_1^{\Omega} = m_1^{\Omega} + \gamma_2\delta_1,$$

$$\widetilde{m}_2^{\Omega} = m_2^{\Omega} + \alpha\gamma_2^2\delta_2 + 2\gamma_2\left(\gamma_1 + \gamma_2\right)\delta_1 + o(\delta), \quad \textit{when } \alpha \ge 1. \tag{D.14}$$

For sufficiently small β*, we have*

$$\begin{aligned} \widetilde{m}_1^{\Omega} &= m_1^{\Omega} + \gamma_2\delta_1, \\ \widetilde{m}_2^{\Omega} &= m_2^{\Omega} + \alpha\gamma_2^2\delta_2 + 2\gamma_2\left(\gamma_1 + \beta\gamma_2\right)\delta_1 + o(\delta). \end{aligned} \tag{D.15}$$

Now, we can prove the proposition by computing explicit expressions of δ_1 and δ_2.

Proof. We note that

$$E[\lambda^{\Gamma}] = \alpha E\left[\frac{\beta\lambda^{FF^H}}{1+\beta\lambda^{FF^H}}\right], \tag{D.16}$$

which has been addressed in (37).

When β is sufficiently small, we have

$$E\left[\frac{\beta\lambda^{FF^H}}{1+\beta\lambda^{FF^H}}\right] = \beta E[\lambda^{FF^H}(1-\beta\lambda^{FF^H}) + o(\beta)] = \beta\left(\frac{1-\beta}{\alpha} - \frac{\beta}{\alpha^2}\right) + o(\beta), \tag{D.17}$$

where we have applied the facts that $E[\lambda^{FF^H}] = 1/\alpha$ and $E[(\lambda^{FF^H})^2] = 1/\alpha + 1/\alpha^2$. This implies

$$\delta_1 = -\beta^2\left(1+\frac{1}{\alpha}\right) + o(\beta). \tag{D.18}$$

Now, we consider the case of large β, for which we have

$$E\left[\frac{\beta\lambda^{FF^H}}{1+\beta\lambda^{FF^H}}\right] = E\left[\frac{1}{1/\beta\lambda^{FF^H}} \mid \lambda^{FF^H} > 0\right]$$
$$= 1 - E\left[\frac{1}{\beta\lambda^{FF^H}} \mid \lambda^{FF^H} > 0\right] + o\left(\frac{1}{\beta}\right). \tag{D.19}$$

Therefore, we have

$$\delta_1 = -\alpha E\left[\frac{1}{\beta\lambda^{FF^H}} \mid \lambda^{FF^H} > 0\right] + o\left(\frac{1}{\beta}\right). \tag{D.20}$$

Then, we need to compute $E[1/\beta\lambda^{FF^H} \mid \lambda^{FF^H} > 0]$. An existing result for an $m \times n$ $(m > n)$ large random matrix $\mathbf{X}$ having independent elements and unit-norm columns is [26]

$$E\left[\frac{1}{\lambda^{X^HX}}\right] = \frac{1}{1-n/m}. \tag{D.21}$$

We apply (D.21) to (D.20). When $\alpha < 1$ $(M \le K)$, all $\lambda^{FF^H} > 0$ almost surely. Therefore

$$E\left[\frac{1}{\beta\lambda^{FF^H}} \mid \lambda^{FF^H} > 0\right]$$
$$= E\left[\frac{1}{\beta\lambda^{FF^H}}\right] = E\left[\frac{\alpha}{\beta\lambda^{\hat{F}^H\hat{F}}}\right] = \frac{\alpha}{\beta\alpha(1-\alpha)}, \tag{D.22}$$

where $\hat{\mathbf{F}} \triangleq \sqrt{\alpha}\mathbf{F}^H$ is a $K \times M$ matrix and $\mathbf{F}\mathbf{F}^H = (1/\alpha)\hat{\mathbf{F}}^H\hat{\mathbf{F}}$. This is equivalent to

$$\delta_1 = -\frac{\alpha^2}{\beta(1-\alpha)} + o\left(\frac{1}{\beta}\right). \tag{D.23}$$

When $\alpha > 1$ $(M > K)$, we have

$$P(\lambda^{FF^H} > 0) = \frac{1}{\alpha}. \tag{D.24}$$

Note that $\mathbf{F}^H\mathbf{F}$ is of full rank when $\alpha > 1$. Then we have

$$E\left[\frac{1}{\beta\lambda^{FF^H}} \mid \lambda^{FF^H} > 0\right]$$
$$= \frac{1}{\alpha}E\left[\frac{1}{\beta\lambda^{F^HF}}\right] = \frac{1}{\alpha\beta}\frac{1}{1-1/\alpha} = \frac{1}{\beta(\alpha-1)}, \tag{D.25}$$

which implies

$$\delta_1 = -\frac{\alpha}{\beta(\alpha-1)} + o\left(\frac{1}{\beta}\right). \tag{D.26}$$

It is easy to verify that $\delta_2 = o(\beta^2)$ for small β and $\delta_2 = o(1/\beta)$ for large β. This concludes the proof. □

ACKNOWLEDGMENT

This research was supported by the U.S. National Science Foundation under Grants ANI-03-38807, CNS-06-25637, and CCF-07-28208.

REFERENCES

[1] A. Sendonaris, E. Erkip, and B. Aazhang, "User cooperation diversity. Part I. System description," *IEEE Transactions on Communications*, vol. 51, no. 11, pp. 1927–1938, 2003.

[2] J. N. Laneman and G. W. Wornell, "Distributed space-time-coded protocols for exploiting cooperative diversity in wireless networks," *IEEE Transactions on Information Theory*, vol. 49, no. 10, pp. 2415–2425, 2003.

[3] J. N. Laneman, D. N. C. Tse, and G. W. Wornell, "Cooperative diversity in wireless networks: efficient protocols and outage behavior," *IEEE Transactions on Information Theory*, vol. 50, no. 12, pp. 3062–3080, 2004.

[4] A. Host-Madsen, "A new achievable rate for cooperative diversity based on generalized writing on dirty paper," in *Proceedings of the IEEE International Symposium on Information Theory (ISIT '03)*, p. 317, Yokohama, Japan, June-July 2003.

[5] Y. Zhao, R. Adve, and T. J. Lim, "Improving amplify-and-forward relay networks: optimal power allocation versus selection," in *Proceedings of the IEEE International Symposium on Information Theory (ISIT '06)*, pp. 1234–1238, Seattle, Wash, USA, July 2006.

[6] I. Maric and R. D. Yates, "Cooperative multihop broadcast for wireless networks," *IEEE Journal on Selected Areas in Communications*, vol. 22, no. 6, pp. 1080–1088, 2004.

[7] A. Bletsas and A. Lippman, "Efficient collaborative (viral) communication in OFDM based WLANs," in *Proceedings of the International Symposium on Advanced Radio Technologies (ISART '03)*, Institute of Standards and Technology, Boulder, Colo, USA, March 2003.

[8] J. Luo, R. S. Blum, L. J. Greenstein, L. J. Cimini, and A. M. Haimovich, "New approaches for cooperative use of multiple antennas in ad hoc wireless networks," in *Proceedings of the 60th IEEE Vehicular Technology Conference Fall (VTC '04)*, vol. 4, pp. 2769–2773, Los Angeles, Calif, USA, September 2004.

[9] A. Bletsas, A. Lippman, and D. P. Reed, "A simple distributed method for relay selection in cooperative diversity wireless networks, based on reciprocity and channel measurements," in *Proceedings of the 61st IEEE Vehicular Technology Conference (VTC '05)*, vol. 3, pp. 1484–1488, Stockholm, Sweden, May-June 2005.

[10] T. Himsoon, W. P. Siriwongpairat, Z. Han, and K. J. R. Liu, "Lifetime maximization via cooperative nodes and relay deployment in wireless networks," *IEEE Journal on Selected Areas in Communications*, vol. 25, no. 2, pp. 306–316, 2007.

[11] Z. Han and H. V. Poor, "Lifetime improvement in wireless sensor networks via collaborative beamforming and cooperative transmission," *IET Microwaves, Antennas & Propagation*, vol. 1, no. 6, pp. 1103–1110, 2007.

[12] Z. Han, X. Zhang, and H. V. Poor, "Cooperative transmission protocols with high spectral efficiency and high diversity order using multiuser detection and network coding," to appear in *IEEE Transactions on Wireless Communications*.

[13] B. Wang, Z. Han, and K. J. R. Liu, "Distributed relay selection and power control for multiuser cooperative communication networks using buyer/ seller game," in *Proceedings of the 26th*

IEEE International Conference on Computer Communications (INFOCOM '07), pp. 544–552, Anchorage, Alaska, USA, May 2007.

[14] Z. Han and H. V. Poor, "Coalition games with cooperative transmission: A cure for the curse of boundary nodes in selfish packet-forwarding wireless networks," in *Proceedings of the 5th International Symposium on Modeling and Optimization in Mobile, Ad Hoc, and Wireless Networks (WiOpt07)*, Limassol, Cyprus, April 2007.

[15] J. Huang, Z. Han, M. Chiang, and H. V. Poor, "Auction-based distributed resource allocation for cooperation transmission in wireless networks," in *Proceedings of the 50th Annual IEEE Global Telecommunications Conference (GLOBECOM '07)*, pp. 4807–4812, Washington, DC, USA, November 2007.

[16] V. I. Morgenshten and H. Bölcskei, "Random matrix analysis of large relay networks," in *Proceedings of the 44th Annual Allerton Conference on Communication, Control, and Computing*, pp. 106–112, Monticello, Ill, USA, September 2006.

[17] A. J. Grant, "Performance analysis of transmit beamforming," *IEEE Transactions on Communications*, vol. 52, no. 4, pp. 738–744, 2005.

[18] V. L. Girko, *Theory of Random Determinants*, Kluwer Academic Publishers, Boston, Mass, USA, 1990.

[19] A. M. Tulino and S. Verdú, *Random Matrix Theory and Wireless Communications*, Foundations and Trends in Communications and Information, Now Publisher, Amsterdam, The Netherlands, 2004.

[20] L. Li, A. M. Tulino, and S. Verdú, "Asymptotic eigenvalue moments for linear multiuser detection," *Communications in Information and Systems*, vol. 1, pp. 273–304, 2001.

[21] R. Speicher, "Multiplicative functions on the lattice of non-crossing partitions and free convolution," *Mathematische Annalen*, vol. 298, no. 1, pp. 611–628, 1994.

[22] R. Speicher, "Free probability theory and non-crossing partitions," in unpublished lecture notes, at 39e Seminare Lotharingien de Combinatoire, http://citeseer.ist.psu.edu/speicher97free.html.

[23] D. Voiculescu, "Limit laws for random matrices and free products," *Inventiones Mathematicae*, vol. 104, no. 1, pp. 201–220, 1991.

[24] J. W. Silverstein and Z. D. Bai, "On the empirical distribution of eigenvalues of a class of large dimensional random matrices," *Journal of Multivariate Analysis*, vol. 54, no. 2, pp. 175–192, 1995.

[25] A. Nica and R. Speicher, "A "Fourier transform" for multiplicative functions on non-crossing partitions," *Journal of Algebraic Combinatorics*, vol. 6, no. 2, pp. 141–160, 1997.

[26] S. Verdú, *Multiuser Detection*, Cambridge University Press, Cambridge, UK, 1998.

Hindawi Publishing Corporation
EURASIP Journal on Advances in Signal Processing
Volume 2008, Article ID 683105, 11 pages
doi:10.1155/2008/683105

Research Article

Bandwidth-Efficient Cooperative Relaying Schemes with Multiantenna Relay

Khuong Ho-Van and Tho Le-Ngoc

Department of Electrical and Computer Engineering, McGill University, Montreal, QC, Canada H3A 2A7

Correspondence should be addressed to Tho Le-Ngoc, tho@ece.mcgill.ca

Received 1 November 2007; Revised 12 February 2008; Accepted 17 March 2008

Recommended by Hyundong Shin

We propose coded cooperative relaying schemes in which all successfully decoded signals from multiple sources are forwarded simultaneously by a multiantenna relay to a common multiantenna destination to increase bandwidth efficiency. These schemes facilitate various retransmission strategies at relay and single-user and multiuser iterative decoding techniques at destination, suitable for trade-offs between performance, latency, and complexity. Simulation results show that the proposed schemes significantly outperform direct transmission under the same transmit power and bandwidth efficiency.

1. INTRODUCTION

Cooperative relaying has attracted a great deal of attention recently due to its capability of improving performance, increasing system capacity, extending coverage, and so forth [1, 2]. Different signal processing techniques for retransmission and detection at relays and destination for cooperative relaying have been presented. In [3–6], the relays receive signals from sources in one phase and simply amplify or demodulate source signals before forwarding processed signals to the destination in another phase. The destination can use maximum ratio combining in both phases to recover the original information. In [7–11], coded cooperative relaying schemes were proposed, in which the relays decode the source signals and re-encode the decoded information in a different manner as compared to the sources (e.g., the decoded information is interleaved before being re-encoded [8]) so that the destination can use code combining techniques such as iterative decoding to recover the original information. Coded cooperative relaying schemes are not only better than those based on repetition coding under various channel conditions [1], but also provide a great degree of flexibility to adapt channel conditions by allowing different code rates and partitions, for example, relayed signal can include just new parity bits [9] or with a fraction of repeated information bits [10].

The cooperative relaying schemes in [2–11] only consider a simple scenario with a source, a relay, and a destination; all are equipped with a single antenna. To increase spatial diversity order as well as cooperation probability between the source and the relay, several multiantenna relays were investigated using the diversity combining schemes in [12]. In general, all schemes in [2–12] reduce dramatically bandwidth efficiency as extended to a scenario with multiple sources. This comes from the fact that at least one additional phase is required to relay the signal for each source.

Different from those in [2–12], the coded cooperative relaying scheme in [13, 14] illustrates another scenario in which a relay assists the information transmission of two sources. This scheme can be extended to the case of multiple sources. However, it suffers the same disadvantage of low bandwidth efficiency as those in [2–12]. It is noted that, in order to achieve high bandwidth efficiency, a *single-antenna* relay can detect *multiple* source signals and retransmit them in only one time slot as a *multiplexed* signal using a much higher modulation level than that of the sources at the expense of increased complexity and transmit power. In [15], a cooperative relaying scheme is proposed, where a *multiantenna relay* helps multiple single-antenna sources in their information transmission to a common multiantenna destination. By relaying each source signal on each antenna of the relay, this scheme exploits the multiplexing gain of

multi-input multi-output (MIMO) systems, thus improving bandwidth efficiency. Theoretical analysis in terms of outage probability shows its superiority to direct transmission. However, the choice of channel codes that can approach the theoretical limit on outage probability is not addressed. In addition, the cooperative relaying scheme under consideration is based on repetition coding and, hence, is not comparable with coded cooperative relaying schemes.

In this paper, we propose *coded* cooperative relaying schemes using *multiantenna* relay to achieve high bandwidth efficiency and high cooperation probability between the sources and the relay (due to receive diversity), which is essential to provide spatial diversity at the destination. In addition, instead of demodulate-and-forward and zero-forcing detection as in [15], we explore the proposed *colocated multiantenna* relaying and code combining structures to develop different efficient retransmission schemes at the relay and single-user and multiuser iterative decoding techniques at the destination in order to improve the system performance. As an example of channel coding, we consider a convolutional code and investigate the performance of the proposed scheme in terms of bit error rate (BER) instead of the outage probability as in [15].

The rest of this paper is organized as follows. In Section 2, we present the system model under consideration. The proposed signal processing techniques at the relay and destination are discussed in Sections 3 and 4, respectively. Simulation results are presented in Section 5 for performance evaluation of the proposed schemes and comparison. Finally, the paper is concluded in Section 6.

2. SYSTEM MODEL

Figure 1 shows the cooperative relaying system under consideration with T single-antenna sources, a T-antenna destination, and a T-antenna relay to assist the communication between the sources and destination. For simplicity, we consider the number of sources equal to that of antennas at the destination and the relay. However, it is straightforward to extend to the general case with F single-antenna sources, a destination with U antennas, and a relay with K antennas where $U, K \geq F$ as in [15]. In addition, we do not consider the cooperation between sources (i.e., similar to [15]), although this cooperation can improve the system performance.

All terminals operate in a half-duplex mode as follows.

Each source S_t, $t \in \{1, \ldots, T\}$ takes turn to transmit its signal in its assigned time slot as shown in Table 1. Throughout this paper, equal-length time slots are assumed. Its information bit segment $\mathbf{I}_t$ is first encoded and then mapped into modulation signaling elements $\mathbf{s}_t$ (e.g., M-PSK, M-QAM) to be transmitted, that is,

$$\mathbf{s}_t = \left(s_t[1], \ldots, s_t[l], \ldots, s_t[L_t]\right) = \varphi\{\Phi\{\mathbf{I}_t\}\}, \quad (1)$$

where $\varphi\{\cdot\}$ and $\Phi\{\cdot\}$ represent the modulation and encoding functions, respectively; $s_t[l]$ is a complex symbol transmitted from the source S_t at the time instant l ($l = 1, \ldots, L_t$); L_t is the number of modulated symbols in the time slot t. If all sources use the same modulation channel coding schemes, $L_t = L$ for any $t \in \{1, \ldots, T\}$.

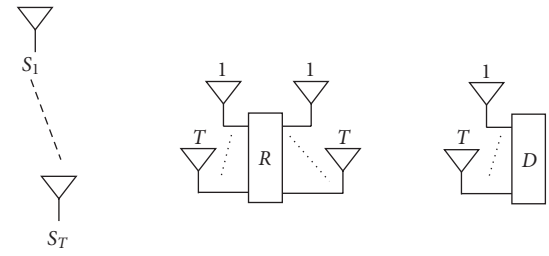


Figure 1: System model.

Table 1: Half-duplex transmission mode.

Time slot	1	⋯	T	$T+1$
Terminal	S_1	⋯	S_T	R

During the first T time slots , the relay decodes the signals received from T sources. Subsequently, the relay processes only the successfully decoded signals (e.g., indicated by the cyclic redundancy check (CRC)) and forwards the processed signals to the destination in the time slot $(T + 1)$ as shown in Table 1. The destination uses both the signals directly received from the sources and the signal from the relay to perform the signal detection.

With only one additional time slot $(T + 1)$ required to relay all decoded signals of T sources, the bandwidth efficiency of the proposed schemes is reduced by a factor of $T/(T + 1)$ as compared to 1/2 for the conventional schemes in [2–14]. For large T, $T/(T + 1)$ approaches 1, that is, the bandwidth loss for relaying is negligible. In a synchronized system with T-antenna relay and destination, simultaneous transmission from T *single-antenna* sources in one time slot is possible for further improved bandwidth efficiency at the expense of receiver complexity and possible performance degradation at relay and destination, and beyond the scope of this paper.

We assume all channels experience independent block frequency-flat fading, that is, frequency-flat fade is fixed during a time slot but independently changed from one time slot to another. Furthermore, channel state information is available only at the receivers, not at the transmitters.

3. PROPOSED COOPERATIVE RELAYING SCHEMES

In this section, we will discuss the signal processing at the relay for detection and retransmission.

3.1. Signal detection at relay

Figure 2 shows the simplified receiver structure at the relay. The baseband-equivalent, discrete-time received signal vector $\mathbf{r}_t[l]$ at the relay can be expressed as

$$\mathbf{r}_t[l] = \mathbf{a}_t s_t[l] + \mathbf{n}_t[l], \quad (2)$$

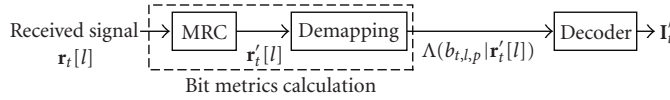


Figure 2: Decoding the signal of the source S_t at the relay.

where $\mathbf{a}_t$ is the $T \times 1$ channel vector from the transmit antenna of the source S_t to the T receive antennas of the relay (each element of $\mathbf{a}_t$ is modeled as circularly symmetric zero-mean complex Gaussian random variable), and $\mathbf{n}_t[l]$ is the $T \times 1$ noise vector with the covariance matrix $N_0\mathbf{I}_{T\times T}$ (i.e., the elements of $\mathbf{n}_t[l]$ are modeled as circularly symmetric zero-mean complex Gaussian random variables with variance $N_0/2$ per dimension). Here, $\mathbf{I}_{T\times T}$ is the unity matrix of the size $T \times T$.

To produce $\mathbf{I}'_t$, at first maximum ratio combining is applied to the elements of $\mathbf{r}_t[l]$ as

$$r'_t[l] = \frac{\mathbf{a}_t^H \mathbf{r}_t[l]}{\sqrt{\mathbf{a}_t^H \mathbf{a}_t}} = a'_t s_t[l] + n'_t[l], \tag{3}$$

where $a'_t = \sqrt{\mathbf{a}_t^H \mathbf{a}_t}$, $n'_t[l]$ is the noise variable with variance N_0, and $(\cdot)^H$ is the complex conjugate transpose.

The resulting signals $r'_t[l]$ are then soft demapped to produce the log-likelihood ratios (LLRs) for all the coded bits, that is, the bit metrics, as follows

$$\Lambda(b_{t,l,p} \mid r'_t[l]) = \log\left(\frac{\sum_{s_x\in\chi_{1,p}} \exp\left(-\left|r'_t[l] - a'_t s_x\right|^2/N_0\right)}{\sum_{s_x\in\chi_{0,p}} \exp\left(-\left|r'_t[l] - a'_t s_x\right|^2/N_0\right)}\right), \tag{4}$$

where $p \in \{1, 2, \ldots, m = \log_2 M\}$, $b_{t,l,p}$ is the pth coded bit in a group of $m = \log_2 M$ bits carried by $s_t[l]$, and M is the constellation size. The subsets $\chi_{1,p}$ and $\chi_{0,p}$ contain the signal points in the M-ary constellation whose pth labeling bits are "1" and "0," respectively.

Finally, the bit metrics are applied to decoding $\mathbf{I}'_t$ (e.g., [16]) and error detection (e.g., using CRC) is performed.

3.2. Signal retransmission at relay

For unsuccessful error detection, the corresponding $\mathbf{I}'_t$ is disregarded. The successfully recovered $\mathbf{I}'_t$ is first interleaved by a random interleaver Π and then processed for retransmission.

For low implementation complexity, the relay applies the same channel coding and modulation schemes used by the sources.

We propose two following retransmission techniques.

3.2.1. Parallel transmission (PT)

For parallel transmission (PT), the N ($\leq T$) successfully recovered information segments, $\mathbf{I}'_t$, $t \in \{1, \ldots, T\}$ are processed separately and retransmitted on different antennas as shown in Figure 3. The relay randomly chooses N among T transmit antennas (e.g., the first N out of T antennas as in the simulations). With channel knowledge at relay *transmitter*, an optimum choice of N antennas for retransmission can be derived. For notational simplicity, we assume $T = N$ in the sequel. Obviously, by simply changing the sizes of vectors and matrices in equations, we easily obtain equations for the case of $T \geq N$.

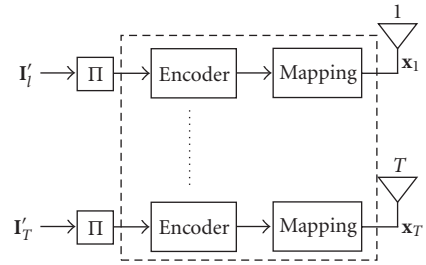


Figure 3: Parallel transmission.

The signal $\mathbf{x}_t$ transmitted on the antenna t can be represented as

$$\mathbf{x}_t = (x_t[1], \ldots, x_t[l], \ldots, x_t[L]) = \varphi\{\Phi\{\Pi\{\mathbf{I}'_t\}\}\}, \tag{5}$$

where $\Pi\{\cdot\}$ represents the interleaving function, and $x_t[l]$ is the modulated symbol transmitted on the antenna t at the time instant l.

3.2.2. Multiplexing transmission (MT)

Figure 4 shows the block diagram of the proposed multiplexing transmission (MT) technique. The interleaved information segments $\Pi\{\mathbf{I}'_t\}$ are first bit-level multiplexed as in [17], that is, the information bits of $\Pi\{\mathbf{I}'_1\}, \ldots, \Pi\{\mathbf{I}'_T\}$ are alternately selected. Therefore, the correlation between $\mathbf{I}'_t$ is introduced to facilitate the high-performance *multiuser joint iterative decoding* (MUJID) to be done at the destination. While multiplexing increases the volumes (in bits), it also makes longer parity segments, and hence stronger codes. Then, the multiplexed segment $\mathbf{J} = \Omega\{\Pi\{\mathbf{I}'_1\}, \ldots, \Pi\{\mathbf{I}'_T\}\}$ is encoded, where $\Omega\{\cdot, \cdot\}$ represents the multiplexing function. Finally, the resulting coded bits $\Phi\{\mathbf{J}\}$ are subsequently split into T parallel streams; each is modulated and transmitted on one antenna.

4. SIGNAL PROCESSING AT DESTINATION

The destination processes the signals from T sources received in the first T time slots to produce their corresponding bit metrics in a similar manner as the relay. Hence, we use the same notations as in Section 3.1 to avoid the duplication. For example, $\Lambda(b_{t,l,p} \mid r'_t[l])$ is the LLR of the pth coded bit in a group of m bits carried by $s_t[l]$, which is computed based on the signal at the destination received from S_t.

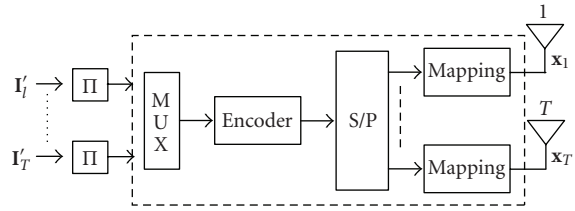


Figure 4: Multiplexing transmission.

In the last $(T+1)$th time slot, the destination receives the signal from the relay. The baseband-equivalent, discrete-time received signal vector $\mathbf{y}[l]$ at the time instant l in the time slot $(T+1)$ at the destination can be modeled as

$$\mathbf{y}[l] = \mathbf{H}\mathbf{x}[l] + \mathbf{n}[l], \quad l = 1, \ldots, L, \tag{6}$$

where $\mathbf{y}[l]$ is the $T \times 1$ received signal vector on the T receive antennas of the destination, $\mathbf{H}$ is the $T \times T$ channel matrix from the T transmit antennas of the relay to the T receive antennas of the destination (the elements of $\mathbf{H}$ are modeled as circularly symmetric zero-mean complex Gaussian random variables), $\mathbf{x}[l] = (x_1[l], x_2[l], \ldots, x_T[l])^T$ is the $T \times 1$ symbol vector transmitted from the relay at the time instant l, and $\mathbf{n}[l]$ is the $T \times 1$ noise vector with the covariance matrix $N_0\mathbf{I}_{T\times T}$. Here $(\cdot)^T$ is the transpose operator.

In the following subsections, we will discuss the proposed bit metric calculations and iterative decoding structures.

4.1. Bit metric calculations in time slot $(T+1)$

The destination also needs to calculate the bit metrics for all coded bits (retransmitted by the relay) in order to perform the iterative decoding for all T source signals. We consider three calculation techniques based on maximum likelihood (ML), zero-forcing (ZF), and QR decomposition.

4.1.1. ML-based bit metric calculation (MLC)

The LLRs for all coded bits transmitted from the relay are computed as

$$\Lambda(b_{r,t,l,p} \mid \mathbf{y}[l]) = \log\left(\frac{\sum_{\mathbf{x}\in\chi_{1,t,p}} \exp\left(-\|\mathbf{y}[l] - \mathbf{H}\mathbf{x}\|^2/N_0\right)}{\sum_{\mathbf{x}\in\chi_{0,t,p}} \exp\left(-\|\mathbf{y}[l] - \mathbf{H}\mathbf{x}\|^2/N_0\right)}\right), \tag{7}$$

where $p \in \{1, 2, \ldots, m\}$, $b_{r,t,l,p}$ is the pth coded bit in a group of m bits carried by $x_t[l]$. The subsets $\chi_{1,t,p}$ and $\chi_{0,t,p}$ contain the symbol vectors $\mathbf{x} = (x_1, x_2, \ldots, x_T)^T$ so that the signal points x_t in the M-ary constellation whose pth labeling bits are "1" and "0," respectively.

The ML-based bit metric calculationis optimum in the sense of minimum bit error probability. However, to calculate $\Lambda(b_{r,t,l,p} \mid \mathbf{y}[l])$ in (7), we need to sum over 2^{mT-1} possible symbol vectors in the set $\chi_{1,t,p}$. So, the complexity of the ML-based bit metrics calculation can be prohibitive for large M and T. This problem can be remedied by applying the list slab-sphere detection method in [18], but the searching range of this method depends on the received signals, thus making the complexity still high. In this paper, we propose two low-complexity methods: *ZF-based bit metric calculation (ZFC)* and *QR -based bit metric calculation (QRC)*.

4.1.2. ZF-based bit metric calculation (ZFC)

The received vector $\mathbf{y}[l]$ is first multiplied by $\mathbf{W} = (\mathbf{H}^H\mathbf{H})^{-1}\mathbf{H}^H$ to suppress the interference between transmitted symbols on different transmit antennas:

$$\mathbf{z}[l] = \mathbf{W}\mathbf{y}[l] = \mathbf{x}[l] + \boldsymbol{\eta}[l], \tag{8}$$

where $\mathbf{z}[l] = (z_1[l], \ldots, z_T[l])^T$ and $\boldsymbol{\eta}[l] = \mathbf{W}\mathbf{n}[l] = (\eta_1[l], \ldots, \eta_T[l])^T$ with $\eta_t[l]$ being a circularly symmetric zero-mean complex Gaussian random variable with variance $\sigma_t[l] = \mathbf{W}(t,:)\mathbf{W}(t,:)^H N_0$. $\mathbf{W}(t,:)$ denotes the tth row of the matrix $\mathbf{W}$.

Explicitly, (8) can be rewritten as

$$z_t[l] = x_t[l] + \eta_t[l]. \tag{9}$$

Therefore, we apply (4) to compute the LLRs for all coded bits from the relay as

$$\Lambda(b_{r,t,l,p} \mid z_t[l]) = \log\left(\frac{\sum_{s_x\in\chi_{1,p}} \exp\left(-|z_t[l] - s_x|^2/\sigma_t[l]\right)}{\sum_{s_x\in\chi_{0,p}} \exp\left(-|z_t[l] - s_x|^2/\sigma_t[l]\right)}\right). \tag{10}$$

Although the ZF-based bit metrics calculation is much simpler than the ML-based bit metrics calculation (i.e., to calculate $\Lambda(b_{r,t,l,p} \mid z_t[l])$ in (10), we only need to sum over 2^{m-1} possible symbols in the set $\chi_{1,p}$), multiplying $\mathbf{y}[l]$ by $\mathbf{W}$ causes the noise enhancement with a factor of $\mathbf{W}(t,:)\mathbf{W}(t,:)^H$ and therefore, leading to the performance degradation.

4.1.3. QR-based bit-metric calculation (QRC)

Using QR decomposition [19], that is, $\mathbf{H} = \mathbf{Q}\mathbf{R}$ where $\mathbf{Q}$ is a unitary matrix and $\mathbf{R} = [r_{i,j}]$ is an upper triangular matrix (i.e., $r_{i,j} = 0$ if $i > j$), (6) can be rewritten as

$$\mathbf{k}[l] = \mathbf{Q}^H\mathbf{y}[l] = \mathbf{R}\mathbf{x}[l] + \boldsymbol{\nu}[l], \tag{11}$$

where $\mathbf{k}[l] = (k_1[l], \ldots, k_T[l])^T$ and $\nu[l] = \mathbf{Q}^H\,\mathbf{n}[l] = (\nu_1[l], \ldots, \nu_T[l])^T$ has the same probability distribution of $\mathbf{n}[l]$ since $\mathbf{Q}$ is a unitary matrix. The elements of $\mathbf{k}[l]$ can be expressed as

$$k_T[l] = r_{T,T}x_T[l] + \nu_T[l], \tag{12}$$

$$k_t[l] = r_{t,t}x_t[l] + \sum_{j=t+1}^{T} r_{t,j}x_j[l] + \nu_t[l], \quad t = T-1, \ldots, 1. \tag{13}$$

The above expressions, (12)-(13), indicate that the signal element $x_T[l]$ does not contain any interference from the

other elements, and the element $x_t[l]$ contains interference from only the elements $x_{t+j}[l]$, where $j = 1,\ldots,(T-t)$ and $t = T-1,\ldots,1$. Consequently, we propose the bit metrics calculation in accompany with the successive soft interference cancellation (e.g., [20, 21]) as follows.

Based on (12), and similar to (4), the LLRs for the coded bits transmitted on the antenna T of the relay can be first computed as

$$\Lambda(b_{r,T,l,p} \mid k_T[l]) = \log\left(\frac{\sum_{s_x\in\chi_{1,p}} \exp\left(-\left|k_T[l]-r_{T,T}[l]s_x\right|^2/N_0\right)}{\sum_{s_x\in\chi_{0,p}} \exp\left(-\left|k_T[l]-r_{T,T}[l]s_x\right|^2/N_0\right)}\right). \tag{14}$$

Then, $\Lambda(b_{r,T,l,p} \mid k_T[l])$'s are used to compute the soft symbols, $m_T[l]$'s, corresponding to $x_T[l]$'s for the transmit antenna T and the variances, $\lambda_T[l]$'s, of these soft symbols as

$$\begin{aligned} m_T[l] &= \mathrm{E}\{x_T[l]\} = \sum_{c=1}^{M} x_c \Pr\{x_T[l]=x_c\}, \\ \lambda_T[l] &= \mathrm{E}\{\left|x_c - m_T[l]\right|^2\} \\ &= \sum_{c=1}^{M} \left|x_c - m_T[l]\right|^2 \Pr\{x_T[l]=x_c\}, \end{aligned} \tag{15}$$

where x_c for $c = 1,\ldots,M = 2^m$ are the M possible values of $x_T[l]$, $\mathrm{E}\{\cdot\}$ is the expectation, and the probability of each possible value of $x_T[l]$ is given by

$$\Pr\{x_T[l]=x_c\} = \prod_{p=1}^{m} \Pr\{b_{r,T,l,p}\}. \tag{16}$$

In (16), we assume the statistical independence of each bit $b_{r,T,l,p}$ carried by the symbol $x_T[l]$ and the probability of $b_{r,T,l,p}$ is

$$\Pr\{b_{r,T,l,p}\} = \frac{1}{1+\exp\{(-1)^{b_{r,T,l,p}}\Lambda(b_{r,T,l,p} \mid k_T[l])\}}. \tag{17}$$

Finally, we calculate the LLRs for the coded bits on the remaining transmit antennas in the order $t = T-1,\ldots,1$ in two steps. In the first step, all interferences from the symbols $x_j[l]$'s, on other transmit antennas j, $j = t+1,\ldots,T$ on the symbol $x_t[l]$, on the considered transmit antenna t (see (13)), are softly cancelled out from $k_t[l]$ as

$$\begin{aligned} k'_t[l] &= k_t[l] - \sum_{j=t+1}^{T} r_{t,j} m_j[l] \\ &= r_{t,t}x_t[l] + \underbrace{\sum_{j=t+1}^{T} r_{t,j}\left(x_j[l]-m_j[l]\right) + \nu_t[l]}_{\nu'_t[l]}. \end{aligned} \tag{18}$$

Based on (18) and the Gaussian assumption on the residual interference (same as [20]), the $\nu'_t[l]$ in (18) is the circularly symmetric zero-mean complex Gaussian random variable with variance $\sigma'_t[l]$

$$\sigma'_t[l] = \sum_{j=t+1}^{T} \left|r_{t,j}\right|^2 \lambda_j[l] + N_0. \tag{19}$$

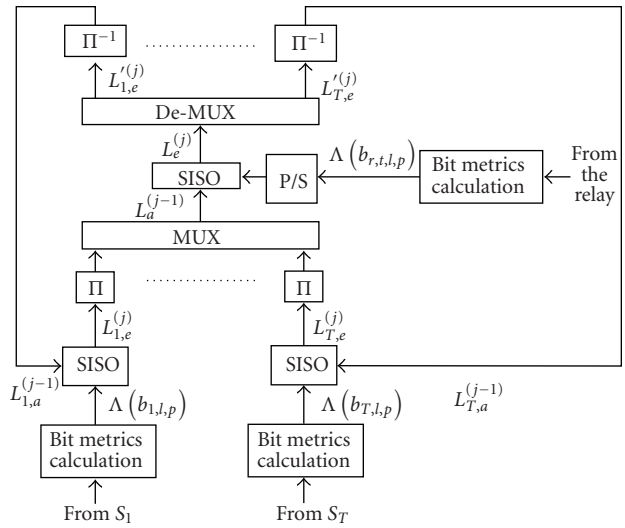


Figure 5: Multiuser joint iterative decoding for multiplexing transmission at the relay.

In (18) and (19), $m_j[l]$ and $\lambda_j[l]$ are given by (15), respectively, with T being substituted by j.

In the second step, we compute the LLRs for the coded bits transmitted on the transmit antenna t of the relay as

$$\Lambda(b_{r,t,l,p} \mid k'_t[l]) = \log\left(\frac{\sum_{s_x\in\chi_{1,p}} \exp\left(-\left|k'_t[l]-r_{t,t}[l]s_x\right|^2/\sigma'_t[l]\right)}{\sum_{s_x\in\chi_{0,p}} \exp\left(-\left|k'_t[l]-r_{t,t}[l]s_x\right|^2/\sigma'_t[l]\right)}\right). \tag{20}$$

From (14) and (20), we realize that to calculate the LLRs for the coded bits we only need to sum over 2^{m-1} possible symbols in the set $\chi_{1,p}$. Therefore, the searching range of QRC and ZFC is the same. However, QRC can avoid the noise enhancement of ZFC (see (18)).

4.2. Iterative decoding

Depending on the transmission techniques at the relay (parallel or multiplexing), we apply the corresponding iterative decoding techniques. For notational convenience, we simplify $\Lambda(b_{t,l,p} \mid r'_t[l])$ in (4) as $\Lambda(b_{t,l,p})$, and unify $\Lambda(b_{r,t,l,p} \mid \mathbf{y}[l])$ in (7), $\Lambda(b_{r,t,l,p} \mid z_t[l])$ in (10), $\Lambda(b_{r,T,l,p} \mid k_T[l])$ in (14), and $\Lambda(b_{r,t,l,p} \mid k'_t[l])$ in (20) as $\Lambda(b_{r,t,l,p})$.

4.2.1. Multiuser joint iterative decoding (MUJID)

Figure 5 shows the decoding diagram for the multiplexing transmission at the relay. Owing to multiplexing the information bit segments of T sources, the MUJID is exploited. The decoder considers a sequence of $(T+1)$ LLR segments, $\Lambda(b_{t,l,p})$, $\Psi\{\Lambda(b_{r,1,l,p}), \Lambda(b_{r,2,l,p}), \ldots, \Lambda(b_{r,T,l,p})\}$ where $\Psi\{\cdot,\cdot\}$ represents the parallel-to-serial converting function, for $t \in \{1,2,\ldots,T\}$ and uses a component soft-in soft-out (SISO) decoder in [16] to recover T information bit segments $\mathbf{I}_t$'s, from T sources within J iterations.

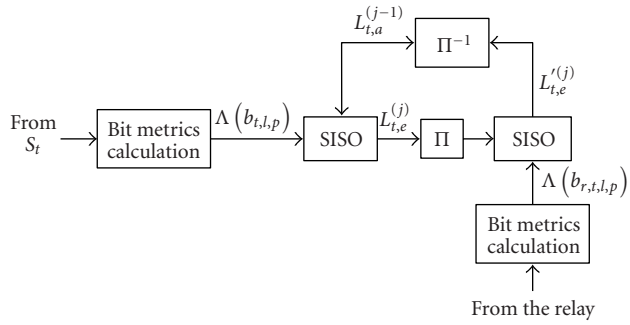


Figure 6: Single-user iterative decoding for source S_t.

Table 2: Proposed cooperative relaying schemes.

Scheme	Description
PT_ZF	PT, ZFC, SUID
PT_QR	PT, QRC, SUID
PT_ML	PT, MLC, SUID
MT_ZF	MT, ZFC, MUJID
MT_QR	MT, QRC, MUJID
MT_ML	MT, MLC, MUJID

In each iteration j for $j \in \{1, 2, \ldots, J\}$, based on the LLR segments, $\Lambda(b_{t,l,p})$, and the *intrinsic* segments, ${L_{t,a}}^{(j-1)}$, the SISO decoder computes the *extrinsic* segments, ${L_{t,e}}^{(j)}$, corresponding to the information bit segments, $\mathbf{I}_t$'s, where ${L_{t,a}}^{(0)} = \mathbf{0}$, since no prior information about the coded bits is available in the first iteration. Then, the *extrinsic* segments, ${L_{t,e}}^{(j)}$, are interleaved and multiplexed into the *intrinsic* segment, ${L_a}^{(j-1)} = \Omega\{\Pi\{{L_{1,e}}^{(j)}\}, \ldots, \Pi\{{L_{T,e}}^{(j)}\}\}$, corresponding to the information bit segment, $\Omega\{\Pi\{\mathbf{I}_1\}, \ldots, \Pi\{\mathbf{I}_T\}\}$. Sequentially, the SISO decoder computes the *extrinsic* segment, ${L_e}^{(j)}$, based on the LLR segment, $\Psi\{\Lambda(b_{r,1,l,p}), \Lambda(b_{r,2,l,p}), \ldots, \Lambda(b_{r,T,l,p})\}$, and the *intrinsic* segment, ${L_a}^{(j-1)}$. Finally, ${L_e}^{(j)}$ is demultiplexed into T *extrinsic* segments, ${L'_{t,e}}^{(j)}$.

At the end of each iteration j, the SISO decoder will produce a sequence of T *extrinsic* segments, ${L'_{t,e}}^{(j)}$, which are the *soft* outputs corresponding to T *information* segments of the T sources, $\mathbf{I}_t$'s. They are stored to be used as inputs of the SISO decoder in the next iteration $(j + 1)$. After a sufficient number of iterations, T *extrinsic* segments, ${L'_{t,e}}^{(j)}$, can be used to make a decision on the transmitted information bit segments.

4.2.2. Single-user iterative decoding (SUID)

As the parallel transmission does not introduce any correlation among the T source signals, the SUID can be used to recover the information bit segment of the source t as shown in Figure 6. This iterative decoding is akin to the standard Turbo decoding and, hence, will not be described further in detail for briefness.

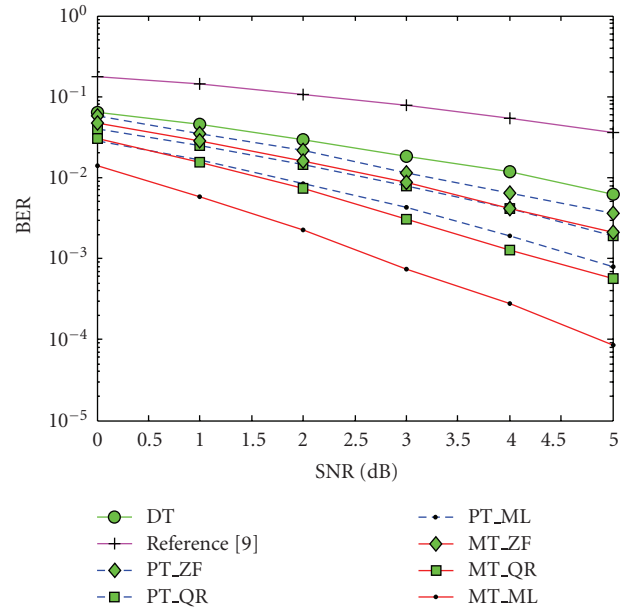


Figure 7: BER versus SNR ($\text{SNR}_{\text{in}} = \text{SNR} + 10$ dB, $\text{SNR}_{\text{rd}} = \text{SNR} + 5$ dB).

5. SIMULATION RESULTS

Simulation is used to evaluate and compare the performance of the proposed schemes and others in an independent frequency-flat block Rayleigh fading environment under various conditions.

5.1. Simulation setup

Table 2 summarizes the 6 proposed schemes under consideration by simulation, as the results of 2 relay retransmission techniques are PT and MT, and 3 bit metric calculations: MLC, ZFC, and QRC.

As reference, we consider the direct transmission (i.e., without the relay) using the 4-state, rate 1/2 recursive systematic convolutional code (RSCC) of generator polynomial [1, 5/7] in octal form, and the cooperative relaying scheme in [9] where T single-antenna relays help T single-antenna sources in the pairwise manner. All considered schemes use the same encoder.

Obviously, the difference in the system model between our proposed schemes and the scheme in [9] is the way to deploy T relay antennas: T colocated antennas as in our system model or T distributed antennas as in [9]. Using T colocated antennas as in our system model benefits from the high cooperation probability between the sources and the relay which is essential to provide spatial diversity at the destination and high bandwidth efficiency (reduced by a factor of $T/(T + 1)$ compared to 1/2 for [9]). On the other hand, the proposed schemes suffer the symbol interference in the time slot $(T + 1)$ while that in [9] does not. However, the low bandwidth efficiency of the scheme in [9] requires an increase in modulation level, thus degrading

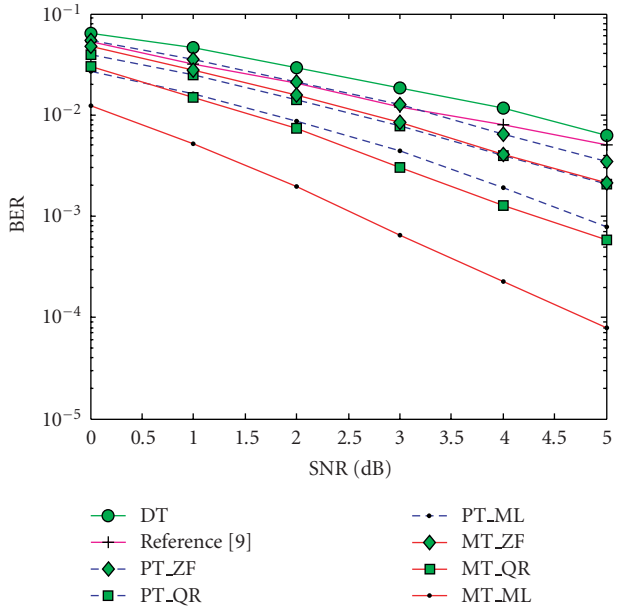


FIGURE 8: BER versus SNR (SNR_{in} = SNR + 20 dB, SNR_{rd} = SNR + 5 dB).

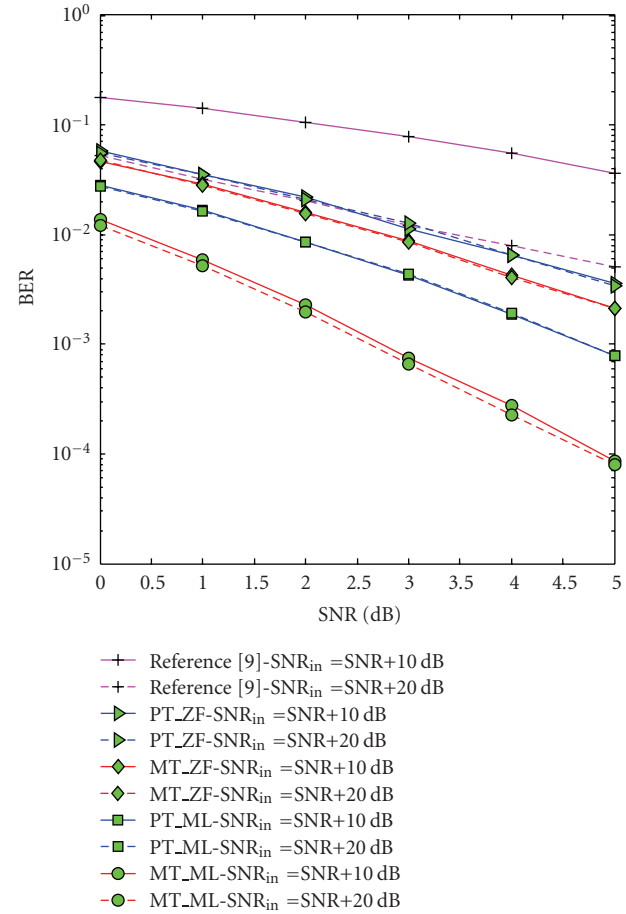


FIGURE 9: BER versus SNR with different interuser channel qualities and SNR_{rd} = SNR + 5 dB.

the performance, which cannot be compensated by the interference-free advantage if the cooperation probability between the source and the relay is low (i.e., interuser channel is bad). These aspects will be demonstrated by the following simulation results.

For the purpose of illustration, we investigate the case of $T = 3$. For a fair comparison in terms of bandwidth efficiency, the direct transmission, the proposed schemes, and that in [9] use 8-PSK, 16-QAM, and 64-QAM, respectively. We also assume equal transmitted power for all terminals and for the relay antennas (i.e., the total relay transmitted power is equally shared by its antennas, $\mathrm{E}\{|x_t[l]|^2\} = \mathrm{E}\{|s_t[l]|^2\}/N$).

We assume identically and independently distributed (iid) frequency-flat fading over any source-relay (or destination) or relay-destination channel. For the scheme in [9], we assume that the relay t corresponds to the antenna t of the relay in our model. We denote the average signal-to-noise ratio of the channel between the source and the receive antenna of the relay as SNR_{in}, between the source and the receive antenna of the destination as SNR, and between the transmit antenna of the relay to the receive antenna of the destination as SNR_{rd}.

The information bit segment is of 180-bit length and the CRC-16-CCITT code is used to check if the recovered source's information segment is error free. In addition, we examine $J = 5$ iterations.

Due to the above iid fading assumption, all sources in the schemes PT_ZF, PT_ML, MT_ZF, and MT_ML have identical performance. However, PT_QR and MT_QR offer different performances for different sources due to the nature of the soft interference cancellation. For this, the performance curves for PT_QR and MT_QR in the following results represent the BER averaged over all sources (i.e., sum of BERs of all sources divided by the number of sources).

5.2. Simulation results

Figure 7 shows the performance curves of the investigated schemes with SNR_{in} = SNR + 10 dB and SNR_{rd} = SNR + 5 dB. We observe that all the proposed schemes significantly outperform the others. Among the proposed schemes, those with MUJID (i.e., MT_ML/MT_QR/MT_ZF) are considerably better than those with SUID (i.e., PT_ML/PT_QR/PT_ZF) due to the longer codeword generated from the multiplexing operation. However, the longer codeword also makes longer decoding latency for the MUJID. Therefore, *performance delay trade-off* can be made for different requirements. In addition, among those with MUJID (or SUID), MT_ML, MT_QR, and MT_ZF (or PT_ML, PT_QR, and PT_ZF) perform in the descending order but their complexities are in the reversed order. This is consistent with the previous discussions. Consequently, another *trade-off between performance and complexity* is also an option for different requirements. Moreover, the scheme in [9] performs even worse than the direct transmission. This comes from the fact that the former (due to the nature of the two time slot cooperative relaying) must use a higher modulation level than that of the latter for the same bandwidth efficiency, while the interuser channel is of low

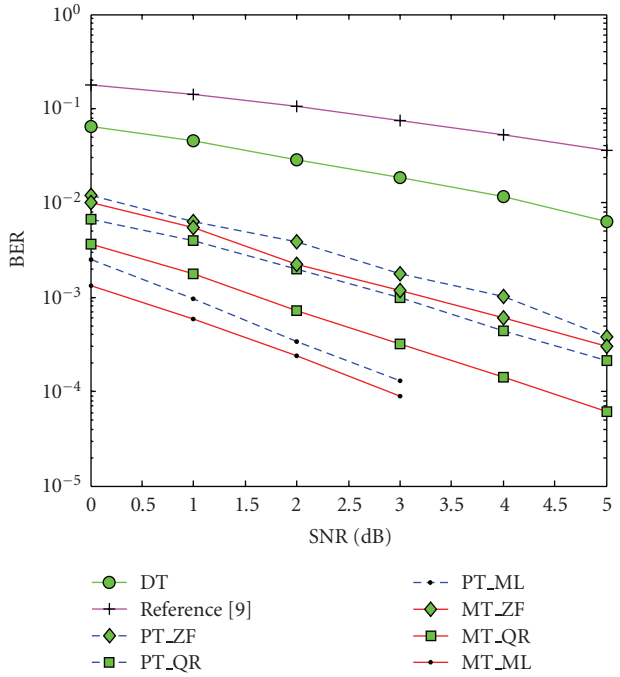


Figure 10: BER versus SNR (SNR_{in} = SNR + 10 dB, SNR_{rd} = SNR + 15 dB).

quality, making the cooperation between the source and the relay take place less frequently. Therefore, the scheme in [9] is almost in the direct transmission mode (i.e., the direct transmission with 64-QAM in [9] is obviously worse than that with 8-PSK).

Figure 8 shows the performance curves of the investigated schemes with better quality interuser channels, SNR_{in} = SNR + 20 dB. Since the source-destination channel qualities are unchanged, the direct transmission has the same performance as previously shown in Figure 7, while the performance of the scheme in [9] is drastically improved with the interuser channel quality. This is because with the improved interuser channel, the cooperation probability between the source and the relay increases, thus enhancing the spatial diversity at the destination. However, it is still worse than any proposed scheme.

The simulation results in Figures 7 and 8 are combined in Figure 9 to see the impact of the interuser channel on the BER performance. It is seen that the proposed schemes are relatively insensitive to the change of the individual interuser channel, while the scheme in [9] is greatly affected. This is obvious since multiple colocated antennas at the relay increase the spatial diversity of the received signals, providing an overall highly reliable transmission over the source-relay channel. As a result, improving an individual source-relay SNR does not contribute significantly to the performance of signal detection at the relay. In contrast, the single-input, single-output source-relay channel in the scheme [9] makes the transmission reliability over this channel heavily dependent on its channel quality (or SNR).

Figure 10 illustrates the performance of various schemes with SNR_{rd} = SNR + 15 dB and SNR_{in} = SNR + 10 dB. The

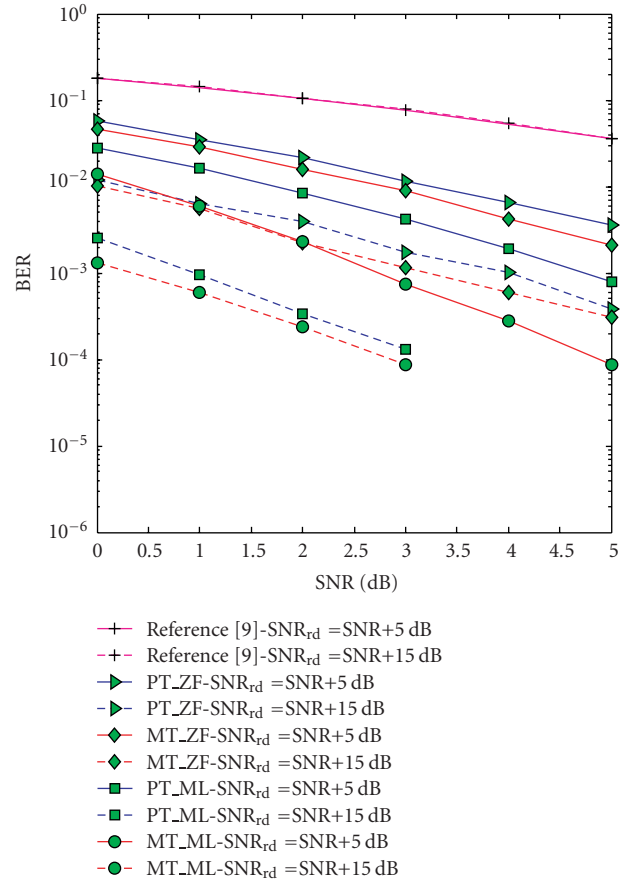


Figure 11: BER versus SNR with different relay destination channel qualities, SNR_{in} = SNR + 10 dB.

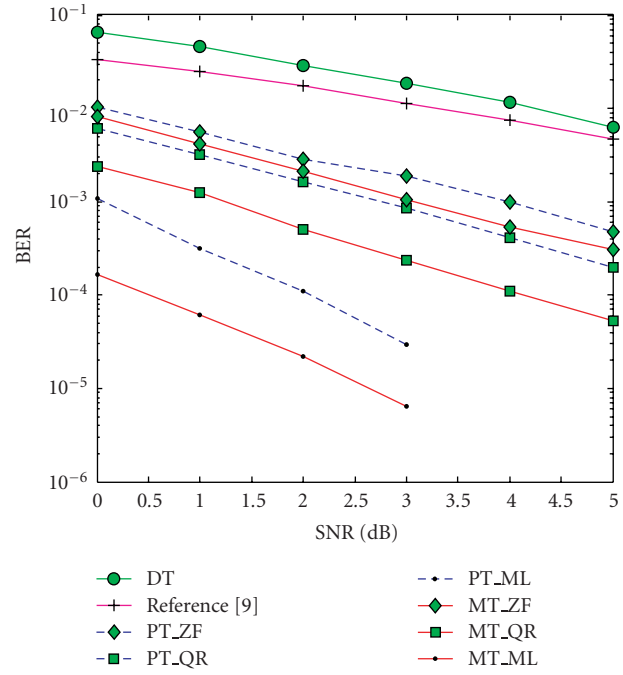


Figure 12: BER versus SNR (SNR_{in} = SNR + 20 dB, SNR_{rd} = SNR + 15 dB).

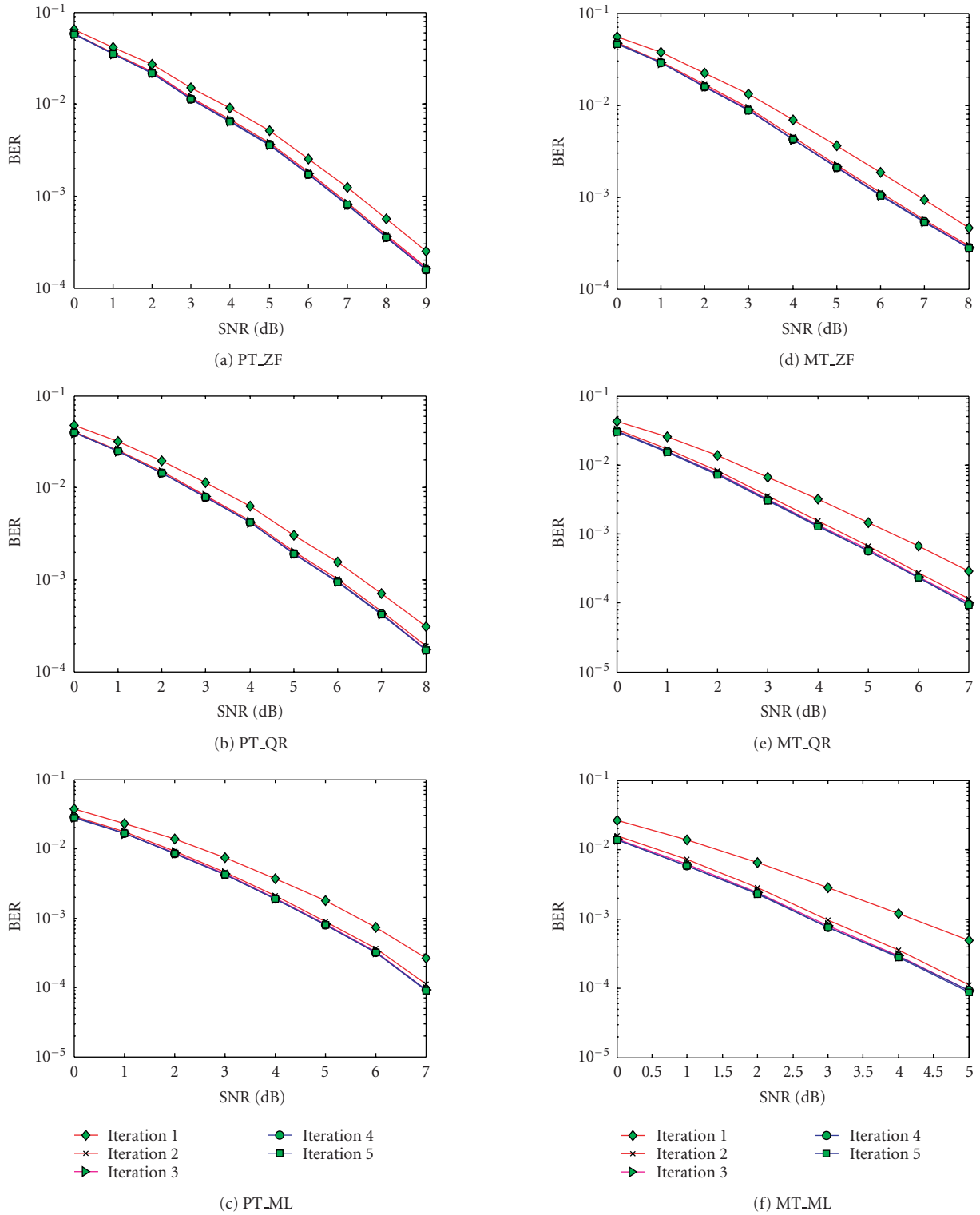


FIGURE 13: BER versus SNR for different iterations (SNR_{in} = SNR + 10 dB and SNR_{rd} = SNR + 5 dB).

performance of the direct transmission is the same as shown in Figure 7 due to the unchanged source-destination channel qualities. With the improved relay-destination channel, the relay forwards the processed information of the sources more reliably, thus enhancing the spatial diversity at the destination. For the scheme in [9], its performance is not improved much, since the cooperation between the relays and the sources are rare (due to unchanged SNR_{in} = SNR + 10 dB as for Figure 7), and as a consequence the better relay-destination channel does not contribute much

to its performance improvement. For easy comparison, we combine the results in Figures 7 and 10 into Figure 11. Figure 11 indicates that the proposed schemes perform drastically better with improved relay-destination channel quality as compared to the others. Figure 11 also shows that MUJID is significantly better than SUID, but their performance difference is reduced with the increased SNR_{rd}. For example, at the target BER of 10^{-3}, the improvement offered by MT_ML as compared to PT_ML is around 2 dB for $SNR_{rd} = SNR + 5$ dB and reduced to only 0.75 dB for $SNR_{rd} = SNR + 15$ dB.

To see the effect of both the source-relay channels and the relay-destination channels on the performance of the investigated schemes, we consider the case where the source-relay channels are improved (e.g., $SNR_{in} = SNR + 20$ dB), while the relay-destination channels are similar to those in Figure 10, that is, $SNR_{rd} = SNR + 15$ dB. The simulation results are illustrated in Figure 12. Since the source-destination channel qualities are unchanged, the direct transmission has the same performance as shown in Figure 7, while the performance of the proposed schemes and that in [9] are substantially improved. In addition, the performance gap between the proposed scheme and that in [9] is dramatically increased with the improvement of the source-relay channels and the relay-destination channels (by comparing Figures 7 and 12).

Figure 13 indicates the BER performance of the 6 proposed schemes for different iterations where $SNR_{in} = SNR + 10$ dB and $SNR_{rd} = SNR + 5$ dB. We see that all the proposed schemes converge after 3 iterations.

6. CONCLUSIONS

We proposed the coded cooperative relaying schemes using a multiantenna relay to assist the information retransmission of multiple sources. These schemes achieve high bandwidth efficiency as well as high performance due to different transmission techniques at the relay and the diversified iterative decoding at the destination. In addition, different from the conventional cooperative relaying schemes (e.g., [9]) whose performance heavily depends on the individual source-relay channel quality, the proposed schemes are almost insensitive to the individual source-relay channel due to the diversity provided by multiple receive antennas. Therefore, the relay can help the sources to improve their performances in a large range of SNR.

In the proposed schemes, we do not consider the cooperation between sources. This cooperation is expected to improve further performance but also makes the cooperative schemes more complicated. It could be an interesting topic for further research.

For a fixed relay as considered in this paper, the channel from the relay and the destination is less time variant. Consequently, the channel state information can be available at the relay so that some techniques such as precoding and power allocation at the relay can be exploited to enhance the information transmission reliability over the relay-destination channel, thus improving the overall system performance.

ACKNOWLEDGMENT

This work was partially supported by the Prompt/NSERC/CRD Grants with InterDigital Canada.

REFERENCES

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

[2] R. Pabst, B. H. Walke, D. C. Schultz, et al., "Relay-based deployment concepts for wireless and mobile broadband radio," *IEEE Communications Magazine*, vol. 42, no. 9, pp. 80–89, 2004.

[3] J. N. Laneman, D. N. C. Tse, and G. W. Wornell, "Cooperative diversity in wireless networks: efficient protocols and outage behavior," *IEEE Transactions on Information Theory*, vol. 50, no. 12, pp. 3062–3080, 2004.

[4] A. Bletsas, A. Khisti, D. P. Reed, and A. Lippman, "A simple cooperative diversity method based on network path selection," *IEEE Journal on Selected Areas in Communications*, vol. 24, no. 3, pp. 659–672, 2006.

[5] A. S. Ibrahim, A. K. Sadek, W. Su, and K. J. R. Liu, "Cooperative communications with partial channel state information: when to cooperate?" in *Proceedings of the IEEE Global Telecommunications Conference (GLOBECOM '05)*, vol. 5, pp. 3068–3072, St. Louis, Mo, USA, November-December 2005.

[6] K. G. Vardhe, D. Reynolds, and M. C. Valenti, "Outage probability of a multi-user cooperation protocol in an asynchronous CDMA cellular uplink," in *Proceedings of the 41st Annual Conference on Information Sciences and Systems (CISS '07)*, pp. 179–184, Baltimore, Md, USA, March 2007.

[7] M. C. Valenti and B. Zhao, "Distributed turbo codes: towards the capacity of the relay channel," in *Proceedings of the 58th IEEE Vehicular Technology Conference (VTC '03)*, vol. 1, pp. 322–326, Orlando, Fla, USA, October 2003.

[8] T. E. Hunter, S. Sanayei, and A. Nosratinia, "Outage analysis of coded cooperation," *IEEE Transactions on Information Theory*, vol. 52, no. 2, pp. 375–391, 2006.

[9] M. Janani, A. Hedayat, T. E. Hunter, and A. Nosratinia, "Coded cooperation in wireless communications: space-time transmission and iterative decoding," *IEEE Transactions on Signal Processing*, vol. 52, no. 2, pp. 362–371, 2004.

[10] H. T. Nguyen, H. H. Nguyen, and T. Le-Ngoc, "A bandwidth-efficient coded cooperative communications system," in *Proceedings of the 64th IEEE Vehicular Technology Conference (VTC '06)*, pp. 1–5, Montreal, QC, Canada, September 2006.

[11] H. Ochiai, P. Mitran, and V. Tarokh, "Variable-rate two-phase collaborative communication protocols for wireless networks," *IEEE Transactions on Information Theory*, vol. 52, no. 9, pp. 4299–4313, 2006.

[12] A. Adinoyi and H. Yanikomeroglu, "Cooperative relaying in multi-antenna fixed relay networks," *IEEE Transactions on Wireless Communications*, vol. 6, no. 2, pp. 533–544, 2007.

[13] C. Hausl and P. Dupraz, "Joint network-channel coding for the multiple-access relay channel," in *Proceedings of the International Workshop on Wireless Ad Hoc and Sensor Networks (IWWAN '06)*, New York, NY, USA, June 2006.

[14] H. T. Nguyen, H. H. Nguyen, and T. Le-Ngoc, "A joint network-channel coding scheme for relay-based communications," in *Proceedings of the Canadian Conference on Electrical and Computer Engineering (CCECE '07)*, pp. 904–907, Vancouver, BC, Canada, April 2007.

[15] F. A. Onat, H. Yanikomeroglu, and S. Periyalwar, "Relay-assisted spatial multiplexing in wireless fixed relay networks," in *Proceedings of the IEEE Global Telecommunications Conference (GLOBECOM '06)*, San Francisco, Calif, USA, November 2006.

[16] S. Benedetto, D. Divsalar, G. Montorsi, and F. Pollara, "A soft-input soft-output APP module for iterative decoding of concatenated codes," *IEEE Communications Letters*, vol. 1, no. 1, pp. 22–24, 1997.

[17] H.-V. Khuong and T. Le-Ngoc, "A bandwidth-efficient coded user-cooperation scheme for flat block fading channels," in *Proceedings of the 4th International Symposium on Wireless Communication Systems (ISWCS '07)*, pp. 421–425, Trondheim, Norway, October 2007.

[18] K.-K. Wong, A. Paulraj, and R. D. Murch, "Efficient high-performance decoding for overload MIMO antenna system," *IEEE Transactions on Wireless Communications*, vol. 6, no. 5, pp. 1833–1843, 2007.

[19] R. Horn and C. Johnson, *Matrix Analysis*, Cambridge University Press, Cambridge, UK, 1985.

[20] W.-J. Choi, K.-W. Cheong, and J. M. Cioffi, "Iterative soft interference cancellation for multiple antenna systems," in *Proceedings of the IEEE Wireless Communications and Networking Conference (WCNC '00)*, vol. 1, pp. 304–309, Chicago, Ill, USA, September 2000.

[21] Y. Li, X.-G. Xia, and G. Wang, "Simple iterative methods to exploit the signal-space diversity," *IEEE Transactions on Communications*, vol. 53, no. 1, pp. 32–38, 2005.

Hindawi Publishing Corporation
EURASIP Journal on Advances in Signal Processing
Volume 2008, Article ID 580368, 7 pages
doi:10.1155/2008/580368

Research Article

Performance of Multiple-Relay Cooperative Diversity Systems with Best Relay Selection over Rayleigh Fading Channels

Salama S. Ikki and Mohamed H. Ahmed

Faculty of Engineering and Applied Science, Memorial University, St. John's, Newfoundland and Labrador, Canada A1B3X5

Correspondence should be addressed to Salama S. Ikki, ikki@engr.mun.ca

Received 16 November 2007; Accepted 17 March 2008

Recommended by Andrea Conti

We consider an amplify-and-forward (AF) cooperative diversity system where a source node communicates with a destination node directly and indirectly (through multiple relays). In regular multiple-relay cooperative diversity systems, all relay nodes relay the source signal using orthogonal channels (time slots, carriers, or codes) to avoid cochannel interference. Hence, for a regular cooperative diversity network with M relays, we need $M+1$ channels (one for the direct link and M for the M indirect links). This means that the number of required channels increases linearly with the number of relays. In this paper, we investigate the performance of the *best-relay selection* scheme where the "best" relay only participates in the relaying. Therefore, two channels only are needed in this case (one for the direct link and the other one for the best indirect link) regardless of the number of relays (M). The best relay is selected as the relay node that can achieve the highest signal-to-noise ratio (SNR) at the destination node. We show that the *best-relay selection* not only reduces the amount of required resources but also maintains a full diversity order (which is achievable by the regular multiple-relay cooperative diversity system but with much more amount of resources). We derive closed form expressions for tight lower bounds of the symbol error probability and outage probability. Since it is hard to find a closed-form expression for the probability density function (PDF) of SNR of the relayed signal at the destination node, we use an approximate value instead. Then, we find a closed-form expression for the moment generating function (MGF) of the total SNR at the destination. This MGF is used to derive the closed-form expressions of the performance metrics such as the average symbol error probability, the outage probability, the average SNR, the amount of fading, and the SNR moments. Furthermore, we derive the asymptotic behavior of the symbol error probability. From this asymptotic behavior, the diversity order and its dependence on the number of relays (M) can be explicitly determined. Simulation results are also given to verify the analytical results.

1. INTRODUCTION

Ever increasing demand for higher data rates in wireless systems has imposed serious challenges on system design and link budget planning. In many scenarios, the desired ubiquitous high rate coverage cannot be achieved by the direct transmission. Multihop relaying has emerged as an intuitive approach to this challenge. The idea is to split the distance between a source node and a destination node into several hops; the nonlinear relation between propagation loss and distance helps in reducing the end-to-end attenuation and thus in relaxing link budget. While such conventional relaying has long been known for some applications as microwave links and satellite relays, it was only until recently that this concept has received interest for wireless and mobile networks [1–5].

Cooperative diversity goes one step further, by considering the participation of several relay nodes (in addition to the source node) in delivering the signal to the destination node, to achieve diversity gain. The cooperative diversity concept is based on the following two features. First, the broadcasting nature of the wireless medium: a signal transmitted by a node propagates not only to the intended final destination, but also to other neighbor nodes. Second, viewing the individual nodes of relaying systems as distributed antennas leads to regarding cooperative diversity networks as a generalization of multiple-antenna systems. In this sense, cooperative diversity brings together

the worlds of conventional relaying and multiple-antenna systems.

The advantages of the cooperative diversity protocols come at the expense of a reduction in the spectral efficiency since the relays must transmit on orthogonal channels in order to avoid interfering with the source node and with each other as well [6]. Hence in cooperative diversity networks with M relaying nodes, $M + 1$ channels are employed, which incurs a bandwidth penalty.

This problem of the inefficient use of the channel resources can be eliminated with the use of the *best-relay selection* scheme. In such a scheme, the "best" relay node only is selected to retransmit to the destination [7]. Hence, two channels only are required in this case (regardless of the number of relays). However, it will be shown (in Section 6) that a full diversity order (which is achievable by the regular cooperative diversity network) can still be achieved with the *best-relay selection*. Therefore, the efficient resource utilization by the *best-relay selection* scheme does not sacrifice the signal quality as will be shown later.

The *best-relay selection* scheme for cooperative networks has been introduced in [7], and the authors showed that this scheme has the same diversity order as the regular cooperative diversity in terms of the capacity outage. However, this important result was given using semianalytical asymptotic analysis at high SNR (without deriving a closed-form expression for the capacity outage). In [8], the authors presented an asymptotic analysis (at high SNR values) only of the symbol error probability of amplify-and-forward *best-relay selection* scheme, and compared it with the regular cooperative systems. The authors showed that *best-relay selection* scheme maintains full diversity order in terms of the symbol error probability. In [9], the authors analyzed the capacity outage probability of the *best-relay selection* scheme with decode-and-forward, and they showed that it outperforms distributed spacetime codes for network with more than three relaying nodes. This gain is due to the efficient use of power by the *best-relay selection* scheme networks. However, to the best of our knowledge, no one has derived closed-form expressions for the symbol error probability and capacity outage of the cooperative diversity network using the *best-relay selection* scheme at any SNR (not only high SNR values).

In this paper, we focus on nonregenerative (amplify-and-forward) dual-hop cooperative diversity network to study their end-to-end performance using the *best-relay selection* scheme over independent nonidentical Rayleigh fading channels. The main contribution of this paper is the derived novel closed-form expressions for the probability density function (PDF), the cumulative distribution function (CDF), and the moment generating function (MGF) of a tight lower bound value of SNR of the relayed signal at the destination. Moreover, the average symbol error probability (SEP) and the capacity outage (C_{out}) are determined using closed-form expressions.

The remaining of this paper is organized as follows. Section 2 presents the system model. The analytical closed-form expressions of the symbol error probability, outage capacity, and the asymptotic symbol error probability are derived in Sections 3, 4, and 5, respectively. Numerical results are discussed in Section 6. Finally, the conclusions are given in Section 7.

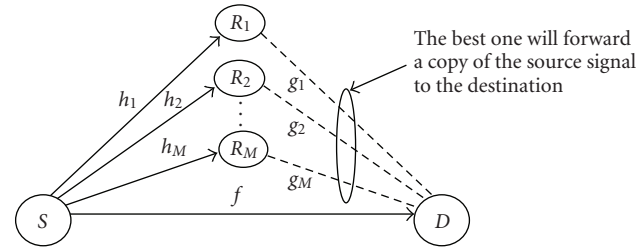


Figure 1: Illustration of the cooperative diversity network with the *best-relay selection* scheme.

2. SYSTEM MODEL

As shown in Figure 1, a source node (S) communicates with the destination (D) through the direct link and the indirect link (through the best relay). This *best-relay selection* scheme allows the destination to get two copies of the source signal. The first one is from the source (direct link), while the second one is from the best relay as shown in Figure 1. The channel coefficients between the source S and the ith relay $R_i(h_i)$, between R_i and $D(g_i)$ and between S and $D(f)$ are flat Rayleigh fading coefficients. In addition, h_i, g_j, and f are mutually-independent and nonidentical for all i and j. We also assume here, without any loss of generality that additive white Gaussian noise (AWGN) terms of all links have zero mean and equal variance N_0.

Assuming that the relaying gain equals $\sqrt{1/(E_s h_i^2 + N_0)}$ [1] (to keep the relay power within its constraints, especially when the fading coefficient (h_i) is low), where E_s is the transmitted signal energy of the source, it is straightforward to show that the end-to-end SNR of the indirect link $S \to R_i \to D$ can be written as [1–10]

$$\gamma_{S \to R_i \to D} = \frac{\gamma_{h_i} \gamma_{g_i}}{\gamma_{h_i} + \gamma_{g_i} + 1}, \tag{1}$$

where $\gamma_{h_i} = h_i^2 E_s / N_0$ is the instantaneous SNR of the source signal at R_i and $\gamma_{g_i} = g_i^2 E_i / N_0$ is the instantaneous SNR of the relayed signal (by R_i) at D, where E_i is the signal transmitted energy of the relay. The best relay will be selected as the one that achieves the highest end-to-end SNR of the indirect link. Then assuming that maximum ratio combining (MRC) technique is employed at the destination node, the total SNR at the destination node can be written as

$$\gamma_{\text{tot}} = \gamma_f + \max_i \left(\frac{\gamma_{h_i} \gamma_{g_i}}{\gamma_{h_i} + \gamma_{g_i} + 1} \right), \tag{2}$$

where $\gamma_f = f^2 E_s / N_0$ is the instantaneous SNR between S and D. In order to use the total SNR in the outage and error performance calculations, (1) should be expressed in a more mathematically tractable form. To achieve it, we proposed in [10] a tight upper bound for $\gamma_{S \to R_i \to D}$, given by

$$\gamma_{S \to R_i \to D} \leq \gamma_i = \min \left(\gamma_{h_i} \gamma_{g_i} \right). \tag{3}$$

The PDF of γ_i can be expressed in terms of the average SNR $\overline{\gamma}_{h_i} = \mathbf{E}(h_i^2)E_s/N_0$ and $\overline{\gamma}_{g_i} = \mathbf{E}(g_i^2)E_s/N_0$ (where $\mathbf{E}(\bullet)$ is the statistical average operator) as $f_{\gamma i}(\gamma) = (1/\overline{\gamma}_i)e^{-\gamma/\overline{\gamma}_i}$, where $\overline{\gamma}_i = \overline{\gamma}_{h_i}\overline{\gamma}_{g_i}/(\overline{\gamma}_{h_i} + \overline{\gamma}_{g_i})$. Using the value of γ_i, we can rewrite the total SNR in (2) as

$$\gamma_{\text{tot}} \leq \gamma_f + \gamma_b, \tag{4}$$

where $\gamma_b = \max_i(\gamma_i) = \max_i(\min(\gamma_{h_i}, \gamma_{g_i}))$. This approximation of the end-to-end SNR in (4) is analytically more tractable than the exact value in (2); and as a result, this facilitates the derivation of the SNR statistics (CDF, PDF, and MGF).This approximation is also adopted in many recent papers (e.g., [7, 11]) and it is shown to be accurate enough, especially at medium and high SNR values as will be discussed in Section 6.

3. ERROR PERFORMANCE ANALYSIS

Since we assume that the MRC technique is employed at the destination, the symbol error probability (SEP) is evaluated for coherent reception only. When multichannel coherent reception is used, we can calculate SEP by averaging the multichannel conditional SEP $P_{\text{se}}(\gamma_f, \gamma_b) = A \times \text{erfc}\left(\sqrt{B(\gamma_f + \gamma_b)}\right)$, (where $\text{erfc}(\cdot)$ is the complementary error function [12, Equation (8.250.4)] given by $\text{erfc}(x) = (2/\pi^{1/2})\int_x^\infty \exp(-t^2)dt$) over the joint random variables representing the SNR values of the direct and indirect links (γ_f, γ_b). Since the random variables (γ_f, γ_b) are assumed to be independent, the joint PDF $f_{\gamma_f,\gamma_b}(\gamma_f, \gamma_b)$ can be given by $f_{\gamma_f}(\gamma_f)f_{\gamma_b}(\gamma_b)$. Therefore, SEP can be determined as follows:

$$\overline{P}_{\text{se}} = \int_0^\infty\int P_{\text{se}}(\gamma_f, \gamma_b) f_{\gamma_f}(\gamma_f) f_{\gamma_b}(\gamma_b) d\gamma_f\, d\gamma_b. \tag{5}$$

Using the alternative definition of the $\text{erfc}(\cdot)$ function as [13]

$$\text{erfc}(x) = \frac{2}{\pi}\int_0^{\pi/2} \exp\left(-\frac{x^2}{\sin^2\theta}\right)d\theta, \tag{6}$$

and by substituting (6) into (5), we obtain

$$\begin{aligned}\overline{P}_{\text{se}} = \int_0^\infty\int \frac{2}{\pi}\int_0^{\pi/2} \exp\left(-\frac{B\gamma_f}{\sin^2\theta}\right)\exp\left(-\frac{B\gamma_b}{\sin^2\theta}\right)\\ \times f_{\gamma_f}(\gamma_f) f_{\gamma_b}(\gamma_b) d\theta\, d\gamma_f\, d\gamma_b.\end{aligned} \tag{7}$$

Since the order of integration can be interchanged [13], we obtain

$$\overline{P}_{\text{se}} = \frac{2}{\pi}\int_0^{\pi/2} M_{\gamma_f}\left(\frac{B}{\sin^2\theta}\right)M_{\gamma_b}\left(\frac{B}{\sin^2\theta}\right)d\theta, \tag{8}$$

where $M_{\gamma_f}(s) = \int_0^\infty f_{\gamma_f}(\gamma_f)\exp(-s\gamma_f)d\gamma_f$ and $M_{\gamma_b}(s) = \int_0^\infty f_{\gamma_b}(\gamma_b)\exp(-s\gamma_b)d\gamma_b$ are the MGF of γ_f and γ_b, respectively.

In order to find $\overline{P}_{\text{se}}$, we need to find the PDF (and then the MGF) of γ_f and γ_b. Since f is Rayleigh distributed random variable, the PDF of γ_f has an exponential distribution with a mean $\overline{\gamma}_f = \mathbf{E}(f^2)E_s/N_0$; hence the MGF of γ_f can be easily found as

$$M_{\gamma_f}(s) = \frac{1}{1 + s\overline{\gamma}_f}. \tag{9}$$

The PDF of γ_b, $f_{\gamma_b}(\gamma)$, can be found as follows. The CDF of γ_b can be written as $F_{\gamma_b}(\gamma) = P(\gamma_b \leq \gamma)$, which can be obtained as

$$F_{\gamma_b}(\gamma) = \prod_{i=1}^{M}(1 - e^{-\gamma/\overline{\gamma}_i}). \tag{10}$$

Then the PDF can be found by taking the derivative of (10) with respect to γ, and after doing some manipulations, $f_{\gamma_b}(\gamma)$, can be written as

$$\begin{aligned}f_{\gamma_b}(\gamma) = \sum_{n=1}^{M}(-1)^{n+1}\sum_{k_1=1}^{M-n+1}\sum_{k_2=k_1+1}^{M-n+2}\cdots\sum_{k_n=k_{n-1}+1}^{M}\\ \times\prod_{j=1}^{n}(e^{-\gamma/\overline{\gamma}_{k_j}})\sum_{j=1}^{n}\frac{1}{\overline{\gamma}_{k_j}}.\end{aligned} \tag{11}$$

By using the PDF in (11), the MGF can be written as

$$\begin{aligned}M_{\gamma_b}(s) = \int_0^\infty e^{-s\gamma}\sum_{n=1}^{M}(-1)^{n+1}\sum_{k_1=1}^{M-n+1}\sum_{k_2=k_1+1}^{M-n+2}\cdots\sum_{k_n=k_{n-1}+1}^{M}\\ \times\prod_{j=1}^{n}(e^{-\gamma/\overline{\gamma}_{k_j}})\sum_{j=1}^{n}\frac{1}{\overline{\gamma}_{k_j}}d\gamma,\end{aligned} \tag{12}$$

and this integral can be evaluated in a closed form as

$$M_{\gamma_b}(s) = \sum_{n=1}^{M}(-1)^{n+1}\sum_{k_1=1}^{M-n+1}\sum_{k_2=k_1+1}^{M-n+2}\cdots\sum_{k_n=k_{n-1}+1}^{M}\frac{\lambda}{s+\lambda}, \tag{13}$$

where $\lambda = \sum_{j=1}^{n}(1/\overline{\gamma}_{k_j})$. Substituting (13) and (9) in (8) and evaluating the integration with the help of [14, Chapter 5], $\overline{P}_{\text{se}}$ can be written in a closed form as

$$\begin{aligned}\overline{P}_{\text{se}} = A\sum_{n=1}^{M}(-1)^{n+1}\sum_{k_1=1}^{M-n+1}\sum_{k_2=k_1+1}^{M-n+2}\cdots\sum_{k_n=k_{n-1}+1}^{M}\\ \times\left(1 - \frac{1/\lambda}{1/\lambda - \overline{\gamma}_f}\sqrt{\frac{B/\lambda}{1+1/\lambda}} + \frac{\overline{\gamma}_f}{1/\lambda - \overline{\gamma}_f}\sqrt{\frac{B\overline{\gamma}_f}{1+\overline{\gamma}_f}}\right).\end{aligned} \tag{14}$$

4. CAPACITY OUTAGE ANALYSIS

The CDF of the total end-to-end SNR using the *best-relay selection* cooperative diversity can be found as follows [13]:

$$F_{\gamma_{\text{tot}}}(\gamma) = \Im^{-1}\left(M_{\gamma_f}(s)M_{\gamma_b}(s)/s\right)\big|_{s=\gamma}, \tag{15}$$

where $\Im^{-1}(\bullet)$ denotes the inverse Laplace transform. This inverse Laplace transform can be performed analytically, and the CDF of the total SNR can be expressed as (by doing

the multiplication first and then using the partial fraction method)

$$F_{\gamma_{\text{tot}}}(\gamma) = \sum_{n=1}^{M}(-1)^{n+1}\sum_{k_1=1}^{M-n+1}\sum_{k_2=k_1+1}^{M-n+2}\cdots\sum_{k_n=k_{n-1}+1}^{M} \times\left(1-\frac{1/\lambda}{1/\lambda-\overline{\gamma}_f}e^{-\lambda\gamma}+\frac{\overline{\gamma}_f}{1/\lambda-\overline{\gamma}_f}e^{-\gamma/\overline{\gamma}_f}\right). \tag{16}$$

The capacity outage (C_{out}) is defined as the probability that the channel average mutual information (I_{Sel}) falls below the required rate R. C_{out} is a very important characterization of any cooperation protocol [1]. For the *best-relay selection* cooperative diversity networks, C_{out} can be written as

$$\begin{aligned} C_{\text{out}} &= \Pr(I_{\text{sel}} \le R) \\ &= \Pr\left(\frac{1}{2}\log_2(1+\gamma_f+\gamma_b) \le R\right) \\ &= \Pr(\gamma_f+\gamma_b \le 2^{2R}-1). \end{aligned} \tag{17}$$

Hence, C_{out} is actually the CDF of γ_{tot} evaluated at $2^{2R}-1$; therefore, $C_{\text{out}} = F_{\gamma_{\text{tot}}}(2^{2R}-1)$. For a regular dual-hop cooperative diversity network (without the *best-relay selection* scheme), it can easily be shown that C_{out} is given by $C_{\text{out}} = F_{\gamma_{\text{tot}}}(2^{(M+1)R}-1)$, which is clearly greater than that of the *best-relay selection* scheme for $M > 1$.

In order to find the other statistics of the total SNR, we have to find the PDF of γ_{tot}, which can be found directly by finding the derivative of the CDF, $F_{\gamma_{\text{tot}}}(\gamma)$, given in (16) with respect to γ yielding

$$f_{\gamma_{\text{tot}}}(\gamma) = \sum_{n=1}^{M}(-1)^{n+1}\sum_{k_1=1}^{M-n+1}\sum_{k_2=k_1+1}^{M-n+2}\cdots\sum_{k_n=k_{n-1}+1}^{M} \times\left(\frac{1}{1/\lambda-\overline{\gamma}_f}e^{-\lambda\gamma}-\frac{1}{1/\lambda-\overline{\gamma}_f}e^{-\gamma/\overline{\gamma}_f}\right). \tag{18}$$

The lth moments of $\gamma_{\text{tot}}(\mu_l = \mathbf{E}(\gamma_{\text{tot}}^l))$ can be found using (18) in a closed form as

$$\mu_l = \Gamma(1+l)\sum_{n=1}^{M}(-1)^{n+1}\sum_{k_1=1}^{M-n+1}\sum_{k_2=k_1+1}^{M-n+2}\cdots\sum_{k_n=k_{n-1}+1}^{M} \times\left(\frac{\overline{\gamma}_f^{l+1}}{1/\lambda-\overline{\gamma}_f}-\frac{(1/\lambda)^{l+1}}{1/\lambda-\overline{\gamma}_f}\right), \tag{19}$$

where $\Gamma(\bullet)$ is the gamma function [12, Equation (8.310.1)]. By setting $l = 1$ in (19), the average total SNR ($\overline{\gamma}_{\text{tot}}$) can be obtained. Furthermore, the first two moments of γ_{tot} can be used in order to evaluate the amount of fading (AF) at the destination [13, Chapter 1]. The AF is defined as the ratio of the variance to the square mean of γ_{tot} ($\text{AF} = \mu_2/\overline{\gamma}_{\text{tot}}^2 - 1$).

5. ASYMPTOTIC ANALYSIS OF THE SYMBOL ERROR PROBABILITY

Although the expression for $\overline{P}_{\text{se}}$ in (14) enables numerical evaluation of the system performance and may not be computationally intensive, this expression does not offer insight into the effect of the different parameters (e.g., the number of relays M) that influence the system performance. In this section, we aim at expressing $\overline{P}_{\text{se}}$ in a simpler form in such a way we can see the effect of the different parameters as $\overline{\gamma}_{h_i}$, $\overline{\gamma}_{g_i}$ and $\overline{\gamma}_f \to \infty$.

The advantage of our accurate approximate solution obtained in the previous sections for the total SNR that we have a closed-form expression for the PDF. For this obtained PDF, the technique developed in [15] can be used to find asymptotic behavior of $\overline{P}_{\text{se}}$ at high SNR. If the approximate PDF of γ_f and γ_b can be written as $f_{\gamma_f}(\gamma) = a_f\gamma^{t_f} + o(\gamma)$ and $f_{\gamma_b}(\gamma) = a_b\gamma^{t_b} + o(\gamma)$, respectively, where, t_f and t_b are positive integers, a_f and a_b are constants, and $o(\gamma)$ is a polynomial function of γ. For γ_f, the value of a_f is $a_f = (1/\overline{\gamma}_f)$ and $t_f = 0$ [15], for γ_b, the value of a_b and t_b can be found as follows. Using the series expansion, the CDF in (10) can be easily rewritten and approximated as

$$\begin{aligned} F_{\gamma_b}(\gamma) &= \prod_{i=1}^{M}\left(1-\left(1-\frac{\gamma}{\overline{\gamma}_i}+\frac{\gamma^2}{2\overline{\gamma}_i^2}-\frac{\gamma^3}{6\overline{\gamma}_i^3}+\cdots\right)\right) \\ &\approx \gamma\prod_{i=1}^{M}\frac{1}{\overline{\gamma}_i}+o(\gamma). \end{aligned} \tag{20}$$

From (20) the values of a_b and t_b are as follows:

$$a_b = \prod_{i=1}^{M}\frac{1}{\overline{\gamma}_i}, \qquad t_b = 0. \tag{21}$$

Then, the approximate PDF of γ_{tot} can be written as [15, 16]

$$f_{\gamma_{\text{tot-approx.}}}(\gamma) \approx a_f a_b \gamma + o(\gamma). \tag{22}$$

Notice that the asymptotic SEP can be given through $\overline{P}_{\text{se}} \to A\int_0^\infty \text{erfc}\sqrt{B\gamma}\, f_{\gamma_{\text{tot-approx.}}}(\gamma)d\gamma$, and after doing the integration, the asymptotic $\overline{P}_{\text{se}}$ can be written as

$$\overline{P}_{\text{se}} \longrightarrow \frac{3A}{8B^2}\frac{1}{\overline{\gamma}_f}\prod_{i=1}^{M}\frac{1}{\overline{\gamma}_i}. \tag{23}$$

In order to see the effect of increasing number of branches explicitly, we assume a special case where all the channels are identical ($\overline{\gamma}_1 = \overline{\gamma}_2 = \cdots = \overline{\gamma}_M = \overline{\gamma}_f = \overline{\gamma}$), then (23) can be written as $\overline{P}_{\text{se}} \to 3A/(8B^2\overline{\gamma}^{M+1})$. It can clearly be seen that the diversity order is equal to $M+1$. This means that the diversity order increases linearly with the number of relays although we use one relay only.

6. NUMERICAL RESULTS

In this section, we show numerical results of the analytical bit error rate (BER) for binary phase shift keying (BPSK) modulation and capacity outage (C_{out}). We plot the performance curves in terms of BER and C_{out} versus the SNR of the transmitted signal (E_s/N_0 dB), where E_s is the transmit energy signal. We also show the results

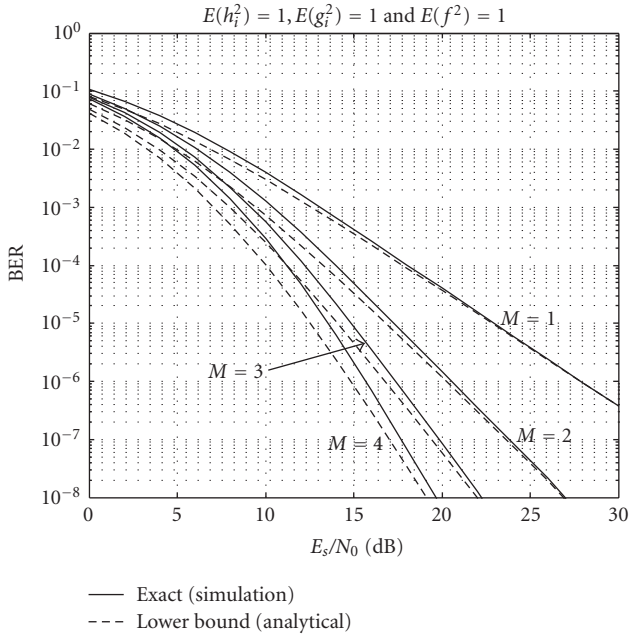


Figure 2: Error performance for the *best-relay selection* scheme over Rayleigh fading channels.

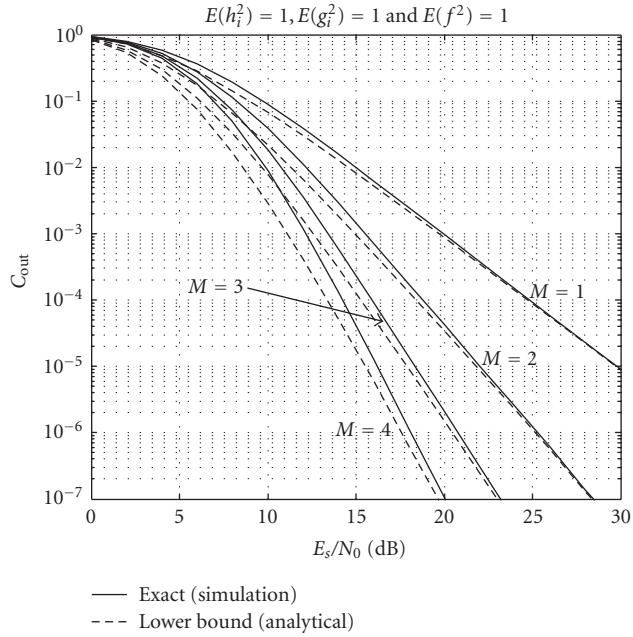


Figure 3: Outage performance for path the *best-relay selection* scheme over Rayleigh fading channels.

of the computer simulations of the *best-relay selection* scheme. We used MATLAB to build Mont-Carlo link-level simulation of the exact model shown in Figure 1 (without any approximations) to compare its results with those found from the analytical approximate model developed in the previous sections.

Figure 2 shows the BER performance of the *best-relay selection* scheme for different values of the number of relays (M). It can be noticed form Figure 2 that the derived lower bound is tight enough, especially at medium and high SNR values. For example, the exact BER (from simulation) for $M = 3$ at $E_s/N_0 = 15$ dB equals 7×10^{-6}, while the analytical BER is 5×10^{-6}. This trend (the tightness of our bound) is valid at different values of M as shown in Figure 2. We can also notice that the BER decreases significantly with the increase in the number of relays (M) since the diversity gain and the virtual antenna gain are monotonically increasing functions of M.

Figure 3 shows the capacity outage (C_{out}) performance for $R = 1$ bit/sec/Hz. Again, it is obvious that the derived lower bound and the simulation results are in excellent agreement. It should be noted that for Figures 2 and 3, the tightness of the derived lower bounds (for BER and C_{out}) improves as E_s/N_0 increases; however, both bounds (for BER and C_{out}) slightly lose their tightness at low E_s/N_0 values, particularly when M increases. This is due to the fact that the accuracy of total SNR approximation (in (4)) improves as E_s/N_0 increases. From Figures 2 and 3, it is evident that the diversity order is equal to $M+1$, which verifies the asymptotic analysis.

Figures 4 and 5 compare the performance of the *best-relay selection* scheme and the regular cooperative diversity in terms of the BER and C_{out} for different values of M. To make the comparisons fair, the transmitted power of the $M + 1$ transmitting nodes (the source node plus the M relays) in the regular cooperative system is set to $E_s = E_i = 1/(M + 1)$. For the *best-relay selection* scheme, we have only two node (the source and the best relay), so $E_s = E_1 = 1/2$. From Figure 4, we can see an interesting result that the *best-relay selection* cooperative diversity scheme outperforms the regular cooperative diversity in terms of the BER. Also, we can see that as M increases this improvement also increases. This behavior is due to the efficient use of power by the *best-relay selection* scheme.

Figure 5 depicts the outage capacity for $R = 1$ bit/sec/Hz. Figure 5 shows the dramatic improvement of the *best-relay selection* cooperative diversity over the regular one in terms of capacity outage. In this figure, as M increases, the capacity outage of the regular cooperative diversity does not necessarily improve. Actually, at low and medium SNR values the capacity outage increases. This is due to the fact that with regular cooperative diversity networks, when the number of relays increases, more channels are needed for relaying; hence it becomes more difficult to achieve the required rate (R). This behavior is completely avoided in the *best-relay selection* scheme because we need only two orthogonal channels for transmissions regardless of the number of relays. Hence increasing the number of relays in the *best-relay selection* scheme always improves the capacity outage without any additional channel resources. This improvement does not depend on the value of E_s/N_0, unlike regular cooperative networks, where the value of E_s/N_0 determines whether increasing the number of relays will decrease the capacity outage or not. For instance, increasing the number of relays (M), from 1 to 2, will reduce C_{out} regardless of the value of E_s/N_0, if the *best-relay selection* scheme is used. However, if

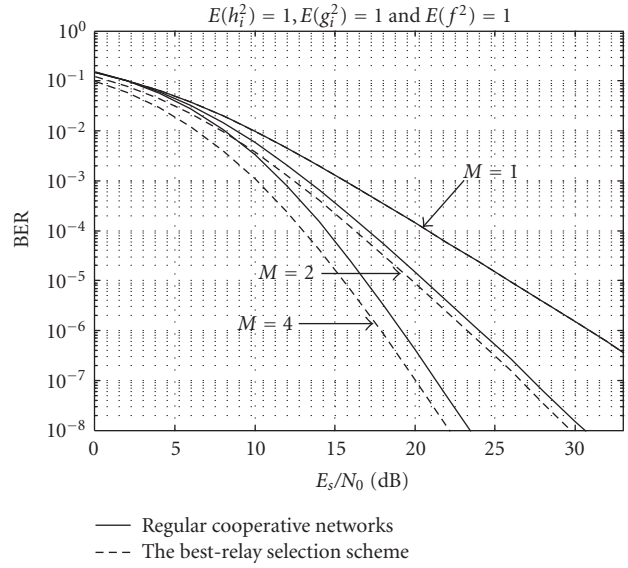


Figure 4: Comparison between the regular cooperative diversity and the *best-relay selection* scheme over Rayleigh fading channels. (Note that for $M = 1$, the regular cooperative diversity and *best-relay selection* scheme are the same.)

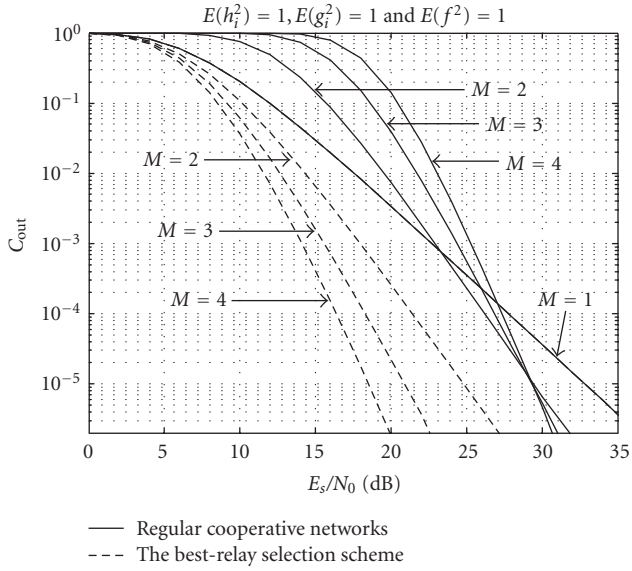


Figure 5: Comparison between the regular cooperative diversity and the *best-relay selection* scheme over Rayleigh fading channels. (Note that for $M = 1$, the regular cooperative diversity and *best-relay selection* scheme are the same.)

regular cooperative diversity is employed, increasing M from 1 to 2 reduces C_{out} only if $E_s/N_0 > 23$ dB.

7. CONCLUSION

We have analyzed the performance of the *best-relay selection* scheme for cooperative diversity networks operating over independent but not necessarily identically distributed Rayleigh fading channels. Novel closed-form expressions for the average SNR, amount of fading, symbol error probability, and capacity outage were derived for any range of SNR (not only high SNR values). Computer simulation results verified the accuracy and the correctness of the derived expressions. We can conclude that *best-relay selection* scheme offers full diversity order.

It should be emphasized that the *best-relay selection* scheme has a strong advantage in saving the channel resources compared to regular cooperative diversity networks. Since the former the total capacity is reduced by 50% only while the latter reduced the channel by $1/M$. This means that in path-selection system increasing M will increase the diversity order without decreasing the channel capacity.

The only disadvantage of this system is the need for a mechanism to find the best relay; however, the implementation of *best-relay selection* scheme can be achieved with minimal signaling overhead and minor additional complexity as shown in [7]. As a future work, our analysis will be extended to the decode-and-forward relaying technique.

REFERENCES

[1] M. O. Hasna and M.-S. Alouini, "End-to-end performance of transmission systems with relays over Rayleigh-fading channels," *IEEE Transactions on Wireless Communications*, vol. 2, no. 6, pp. 1126–1131, 2003.

[2] M. O. Hasna and M.-S. Alouini, "Harmonic mean and end-to-end performance of transmission systems with relays," *IEEE Transactions on Communications*, vol. 52, no. 1, pp. 130–135, 2004.

[3] M. O. Hasna and M.-S. Alouini, "Outage probability of multihop transmission over Nakagami fading channels," *IEEE Communications Letters*, vol. 7, no. 5, pp. 216–218, 2003.

[4] G. K. Karagiannidis, "Performance bounds of multihop wireless communications with blind relays over generalized fading channels," *IEEE Transactions on Wireless Communications*, vol. 5, no. 2, pp. 498–503, 2006.

[5] G. K. Karagiannidis, T. A. Tsiftsis, and R. K. Mallik, "Bounds for multihop relayed communications in Nakagami-*m* fading," *IEEE Transactions on Communications*, vol. 54, no. 1, pp. 18–22, 2006.

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

[7] A. Bletsas, A. Khisti, D. P. Reed, and A. Lippman, "A simple cooperative diversity method based on network path selection," *IEEE Journal on Selected Areas in Communications*, vol. 24, no. 3, pp. 659–672, 2006.

[8] Y. Zhao, R. Adve, and T. J. Lim, "Symbol error rate of selection amplify-and-forward relay systems," *IEEE Communications Letters*, vol. 10, no. 11, pp. 757–759, 2006.

[9] E. Beres and R. S. Adve, "Selection Cooperation in Multi-Source Cooperative Networks," *IEEE Transactions on Wireless Communications*, vol. 7, no. 1, pp. 118–127, 2008.

[10] S. S. Ikki and M. H. Ahmed, "Performance analysis of cooperative diversity wireless networks over Nakagami-*m* fading channel," *IEEE Communications Letters*, vol. 11, no. 4, pp. 334–336, 2007.

[11] T. Wang, A. Cano, G. B. Giannakis, and J. N. Laneman, "High-performance cooperative demodulation with decode-and-forward relays," *IEEE Transactions on Communications*, vol. 55, no. 7, pp. 1427–1438, 2007.

[12] I. S. Gradshteyn and I. M. Ryzhik, *Table of Integrals, Series and Products*, Academic Press, San Diego, Calif, USA, 5th edition, 1994.

[13] M. K. Simon and M.-S. Alouini, *Digital Communication over Fading Channels*, John Wiley & Sons, New York, NY, USA, 2000.

[14] A. Papoulis, *Probability, Random Variables, and Stochastic Processes*, McGraw-Hill, New York, NY, USA, 1991.

[15] Z. Wang and G. B. Giannakis, "A simple and general parameterization quantifying performance in fading channels," *IEEE Transactions on Communications*, vol. 51, no. 8, pp. 1389–1398, 2003.

[16] J. Proakis, *Digital Communications*, McGraw-Hill, New York, NY, USA, 4th edition, 2001.

Hindawi Publishing Corporation
EURASIP Journal on Advances in Signal Processing
Volume 2008, Article ID 531786, 11 pages
doi:10.1155/2008/531786

Research Article

Cooperative Communications over Flat Fading Channels with Carrier Offsets: A Double-Differential Modulation Approach

Manav R. Bhatnagar, Are Hjørungnes, and Lingyang Song

UniK - University Graduate Center, University of Oslo, P.O. Box 70, 25 Instituttveien, 2027 Kjeller, Norway

Correspondence should be addressed to Manav R. Bhatnagar, manav@unik.no

Received 27 October 2007; Revised 14 February 2008; Accepted 1 April 2008

Recommended by Hyundong Shin

We propose double-differential (DD) modulation for the amplify-and-forward protocol over Nakagami-*m* fading channels with carrier offsets. We propose an emulated maximum ratio combining (EMRC) decoder, which could be used by the double-differential receiver in the absence of exact channel knowledge. Approximate bit error rate (BER) analysis is performed for the proposed double-differential modulation-based cooperative communication system. The proposed double-differential system is immune to random carrier offsets, whereas the conventional single-differential modulation-based cooperative system breaks down. In addition, the proposed scheme is able to perform better than the same rate training-based cooperative system which utilizes training data for finding estimates of carrier offsets and channel gains.

1. INTRODUCTION

Cooperative communications has several promising features to become a main technology in future wireless communications systems. It has been shown in the literature [1, 2] that the cooperative communication can avoid the difficulties of implementing actual antenna arrays and convert the single-input single-output (SISO) system into a virtual multiple-input multiple-output (MIMO) system. In this way, cooperation between the users allows them to exploit the diversity gain and other advantages of MIMO system in an SISO wireless network. Most of the existing works within cooperative communications assume that there is no carrier offset present over any link [1–3]. However, this assumption is not justified as cooperative communications are proposed for wireless mobile system, where the mobile users are *moving*. Moreover, the transmit and receive oscillators can never achieve perfect matching. Another practically infeasible assumption is that all nodes in the cooperative network have perfect knowledge about the channel coefficients of all the links in the network. Several single-differential strategies for cooperative communications have been proposed so far to avoid the estimation of the channel coefficients at the receiver side [4–7]. All of these conventional differential schemes assume that the channel is constant over at least *two* consecutive time intervals. However, in the presence of carrier offsets, the flat fading wireless channel does *not* remain constant over two consecutive time intervals and these differential schemes experience substantial performance loss.

Double-differential (DD) modulation [8–11] is a key differential technique to remove the effect of carrier offset from the communication system. It differs from single-differential modulation in a sense that the decoder uses *three* consecutively received data samples for decoding the current symbol. Two levels of single-differential modulation are employed at the transmitter as shown in Figure 2(a) and a simple heuristic decoder [9, Equation (15)] is used at the receiver to find the estimate of the transmitted data as shown in Figure 2(b). It has been shown in [9, Section III] that the heuristic decoder coincides with the maximum likelihood decoder (MLD) under the assumption that the product of two zero-mean white circularly symmetric Gaussian noise terms in the decision variable is also zero-mean white circularly symmetric Gaussian. Symbol error rate (SER) expressions for the double-differentially modulated data over Rayleigh and Ricean fading SISO channels with carrier offsets are provided in [11]. In [12], a double-differential orthogonal space-time block code for time-selective MIMO channels

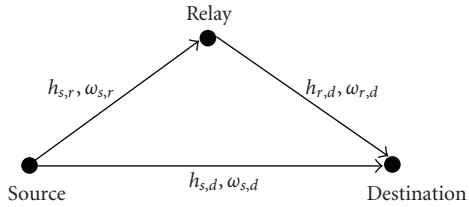


Figure 1: Cooperative communication system with carrier offsets.

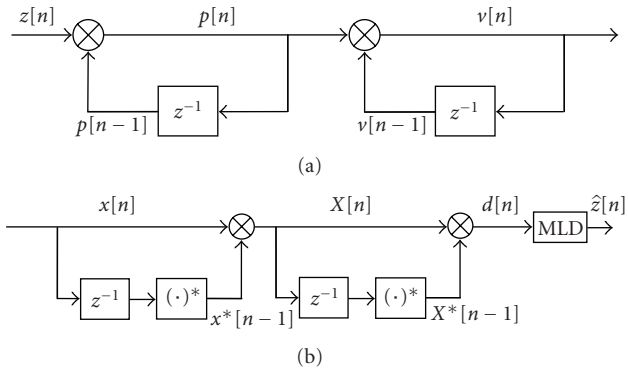


Figure 2: Double-differential (a) encoder and (b) decoder.

is proposed. A distributed double-differential modulation based on [12] with *regenerative* relays over cooperative Rayleigh channels is proposed and an upper bound of the pairwise error performance (PEP) is obtained in [13]. It is assumed in [13] that the flat-fading part of the time-varying channel remains constant over 10^4 time intervals [13, Sections IV and VI].

In this paper, we consider DD modulation for cooperative communications with *nonregenerative* relays over flat-fading Nakagami-m channels with random carrier offsets. The Nakagami-m family of distributions [14], also known as "m-distribution," contains Rayleigh fading ($m = 1$) as a special case; along with cases of fading that are more severe than Rayleigh ($1/2 \le m < 1$) as well as cases less severe than Rayleigh ($m > 1$). In contrast to [13], the proposed scheme works with the basic assumption of DD modulation for SISO system [8–11] that the flat-fading part of the time-varying channel is constant over at least *three* time intervals.

The main contributions of this paper are the following. (1) A double-differential modulation-based amplify-and-forward (AAF) cooperative wireless communication is proposed to improve the performance over flat-fading Nakagami-m channels with random carrier offsets. (2) Approximate analytical bit error rate (BER) of the DD modulation with AAF protocol (DDAAF) is obtained. (3) Based on the approximate BER analysis, we determine the numerical power allocation for the DD cooperative systems.

The rest of this paper is organized as follows. In Section 2, the system model, channel model, and DD modulation for an SISO link are discussed. Section 3 implements DD modulation in the AAF cooperative communications. The approximate BER performance expressions for DD modulation with AAF protocol are found in Section 4. Training-based cooperative communication is discussed in Section 5. In Section 6, the analytical and simulation results are discussed and details of numerical power allocation for DDAAF cooperative system is provided. Section 7 contains some conclusions. The article contains two appendices, which provide details of the derivations.

2. SYSTEM MODEL

We consider a basic cooperative communication system, which consists of one source, one relay, and one destination terminal as shown in Figure 1. Each of them can either transmit or receive a signal at a time. The transmission of the data from the source to the destination terminal is furnished in two phases. In the first phase, the source broadcasts data to the destination and the relay. The relay amplifies the received data and retransmits it to the destination, in the second phase. To avoid the interference, source and relay use orthogonal channels for transmission [3]. For ease of presentation, we assume that in both phases, the source and relay transmit stream of data through time-division multiplexing (TDMA). In the TDMA scheme, the source has to remain silent in the second phase in order to maintain the orthogonality between the transmissions. However, in the frequency-division multiplexing (FDMA) or the code-division multiplexing (CDMA) schemes, the source and the relay can transmit simultaneously.

2.1. Channel model

All links are assumed to be Nakagami-m distributed with the following probability density function (pdf) [15, Equation (2.21)]:

$$f_{\gamma_{p,q}}(\gamma_{p,q}) = \frac{m_{p,q}^{m_{p,q}} \gamma_{p,q}^{m_{p,q}-1}}{\overline{\gamma}_{p,q}^{m_{p,q}} \Gamma(m_{p,q})} \exp\left(-\frac{m_{p,q}\gamma_{p,q}}{\overline{\gamma}_{p,q}}\right), \quad \gamma_{p,q} \ge 0, \tag{1}$$

where $\Gamma(\cdot)$ is the gamma function [16, Equation (8.310.1)], $m_{p,q} \ge 1/2$ is the Nakagami-m fading parameter, $\gamma_{p,q} = P_p|h_{p,q}|^2/\sigma^2$ is the instantaneous signal to noise ratio (SNR), $P_p \in \{P_s, P_r\}$ is the power transmitted by source or relay with $P_s = P_1$ and $P_r = P_2$, $h_{p,q}$ is a *zero-mean* Nakagami-m channel coefficient, σ^2 is the variance of the additive white Gaussian noise (AWGN), and $\overline{\gamma}_{p,q} = P_p\sigma_{p,q}^2/\sigma^2$ is the average SNR over the link between p and q terminals in the cooperative system, where $\sigma_{p,q}^2 = E\{|h_{p,q}|^2\}$ is channel variance and $E\{\cdot\}$ represents expectation. If we represent the source by s, the relay by r, and the destination by d, then $(p,q) \in \{(s,d),(s,r),(r,d)\}$. The channel of each link is assumed to be a block fading channel, which remains constant for *at least three* consecutive time intervals and all the channel coefficients are assumed to be independent of each other. It is assumed that all three links are perturbed by different carrier offsets $\omega_{p,q} = 2\pi f_{p,q} T_s$ [17], where $f_{p,q}$ is the physical carrier frequency offset (CFO) in Hertz, $f_{p,q} \in [-1/(2T_s), 1/(2T_s))$, and T_s is the sampling period in seconds. Apparently, $\omega_{p,q} \in [-\pi, \pi)$ and the maximum value of $\omega_{p,q}$ corresponds to the offset of 50% of the carrier.

We assume that the carrier offsets $\omega_{p,q}$ are random and uniformly distributed over $[-\pi, \pi)$, however, in general, there is no restriction over the probability distribution of the carrier offsets and they could have any probability distribution. We have assumed that these carrier offsets remain fixed for *at least three* consecutive time intervals. The presence of carrier offsets makes all three block-fading channels behave as *time-varying channels*, which do *not* remain stationary over two consecutive time intervals. Since the phase term $e^{\omega_{p,q}n}$ is multiplied with the channel coefficient $h_{p,q}$, the effective channel is time-varying even though $\omega_{p,q}$ and $h_{p,q}$ stay constant for the same three consecutive time instants.

2.2. Double-differential modulation [8–11]

Let $z[n]$ denote the symbols belonging to the unit-norm M-PSK constellation $\mathcal{A}$ to be transmitted at the time n. In a DD modulation-based system [8–11], the transmitted signal $v[n]$ is obtained from $z[n]$ as shown in Figure 2(a):

$$\begin{aligned} p[n] &= p[n-1]z[n], \\ v[n] &= v[n-1]p[n], \quad n = 2, 3, \ldots, \end{aligned} \tag{2}$$

with $v[0] = v[1] = p[1] = 1$. As $|z[n]| = 1$ for the unit-norm M-PSK symbols, it follows from (2) that $|v[n]| = |p[n]| = 1$. We consider a flat fading SISO channel with carrier offset described by

$$x[n] = \sqrt{\rho} h e^{\omega n} v[n] + e[n], \quad n = 0, 1, \ldots, \tag{3}$$

where $x[n]$ is the received signal, ρ is the transmitted signal power, h is the channel gain, $e[n]$ is complex-valued AWGN noise, and $\omega \in [-\pi, \pi)$ is the unknown frequency offset. The receiver makes a decision variable, $d[n] = X[n]X^*[n-1]$, where $X[n] = x[n]x^*[n-1]$ as shown in Figure 2(b). The decoding of $z[n]$ is performed in the following way [9, Equation (15)]:

$$\hat{z}[n] = \arg \max_{z \in \mathcal{A}} \mathrm{Re}\{X[n]X^*[n-1]z^*\}. \tag{4}$$

The decoding of (4) corresponds to maximum-likelihood decoding (MLD) in [9] under the assumption that the product of two zero-mean white circularly symmetric Gaussian noise terms in the decision variable is also zero-mean white circularly symmetric Gaussian.

3. DOUBLE-DIFFERENTIAL MODULATION FOR AAF COOPERATIVE COMMUNICATION SYSTEM

If we use DD modulation in the cooperative communication system, the data received during the *first phase* at the destination is

$$x_{s,d}[n] = \sqrt{P_1} h_{s,d} e^{\omega_{s,d}n} v[n] + e_{s,d}[n], \quad n = 0, 1, \ldots, \tag{5}$$

and at the relay is

$$x_{s,r}[n] = \sqrt{P_1} h_{s,r} e^{\omega_{s,r}n} v[n] + e_{s,r}[n], \quad n = 0, 1, \ldots, \tag{6}$$

where P_1 is the power transmitted by the source, $h_{s,d}$ and $h_{s,r}$ are the channel gains, and $\omega_{s,d}$ and $\omega_{s,r}$ are the carrier offsets between source and destination, and source and relay, respectively, and $e_{s,d}[n]$ and $e_{s,r}[n]$ are complex-valued circular symmetric AWGN noise on the two links. During the *second phase*, the relay amplifies the received data of (6) and retransmits such that the received signal by the destination in the second phase is

$$x_{r,d}[l] = \sqrt{\hat{P}_2} h_{r,d} e^{\omega_{r,d}l} x_{s,r}[l] + e_{r,d}[l], \quad l = 0, 1, \ldots, \tag{7}$$

where l is the time index which is used in the place of n to show the difference in time of first and second phases, $h_{r,d}$ is the channel gain, $\omega_{r,d}$ is the carrier offset between relay and destination, $e_{r,d}[l]$ is the AWGN noise, and $\hat{P}_2$ is the amplification factor which ensures constant average transmission power during the second phase. It can be seen from (6) that the average power of $x_{s,r}[l]$ is $P_1\sigma_{s,r}^2 + \sigma^2$, where $\sigma_{s,r}^2$ is variance of $h_{s,r}$, hence, $\hat{P}_2$ is given by

$$\hat{P}_2 = \frac{P_2}{P_1\sigma_{s,r}^2 + \sigma^2}, \tag{8}$$

where P_2 is the average power transmitted by the relay. It is also assumed that $P_1 + P_2 = P$, where P is the total transmitted power.

Next, we propose the following maximal ratio combining (MRC) [18] based scheme for a DDAAF receiver

$$\begin{aligned} d[k] = {} & \alpha_1 \left(x_{s,d}[n]x_{s,d}^*[n-1]\right)\left(x_{s,d}[n-1]x_{s,d}^*[n-2]\right)^* \\ & + \alpha_2 \left(x_{r,d}[l]x_{r,d}^*[l-1]\right)\left(x_{r,d}[l-1]x_{r,d}^*[l-2]\right)^*, \end{aligned} \tag{9}$$

where $k = n = l$, that is, the data received by the destination during the same time interval with respect to the beginning of each phase is combined, and α_1 and α_2 are given by

$$\begin{aligned} \alpha_1 &= \frac{1}{\left(2P_1 |h_{s,d}|^2 + \sigma^2\right)\sigma^2}, \\ \alpha_2 &= \frac{\left(P_1\sigma_{s,r}^2 + \sigma^2\right)^2}{\kappa}, \end{aligned} \tag{10}$$

where

$$\begin{aligned} \kappa = {} & 2P_1P_2^2 |h_{r,d}|^4 |h_{s,r}|^2\sigma^2 + 2P_1P_2 \\ & \times \left(P_1\sigma_{s,r}^2 + \sigma^2\right) |h_{r,d}|^2 |h_{s,r}|^2\sigma^2 + P_2^2 |h_{r,d}|^4\sigma^4 + 2P_2 \\ & \times \left(P_1\sigma_{s,r}^2 + \sigma^2\right) |h_{r,d}|^2\sigma^4 + \left(P_1\sigma_{s,r}^2 + \sigma^2\right)^2\sigma^4. \end{aligned} \tag{11}$$

The normalization factors can be found as $\alpha_1 = 1/E_1$ and $\alpha_2 = 1/E_2$, where E_1 and E_2 are the average noise powers of $X_{s,d}[n] = x_{s,d}[n]x_{s,d}^*[n-1]$ and $X_{r,d}[n] = x_{r,d}[n]x_{r,d}^*[n-1]$, respectively. However, as we intend to use DD modulation, the destination and relay are not expected to have knowledge of the exact channel coefficients, therefore, we can use emulated maximum ratio combining (EMRC) by replacing the channel coefficients $|h_{s,r}|^2$, $|h_{r,d}|^2$, and $|h_{s,d}|^2$ by their variances $\sigma_{s,r}^2$, $\sigma_{r,d}^2$, and $\sigma_{s,d}^2$, respectively, in (10). Then, the data is decoded as

$$\hat{z}[n] = \arg \max_{z \in \mathcal{A}} \mathrm{Re}\{d[k]z^*\}, \tag{12}$$

where $n = k$. It is shown by simulation in Section 6.2 that the EMRC scheme performs very close to the MRC scheme from moderate to high SNR. Hence, the approximation of using the variances in the place of instantaneous channel values is reasonable from moderate to high SNR.

4. BER PERFORMANCE ANALYSIS

The EMRC obtained by replacing channel gains by their variances in (9) will perform worse than the ideal MRC scheme given by (9) and (10) [18]. For simplicity, we assume that the instantaneous signal-to-noise ratio (SNR) of the EMRC scheme is

$$\gamma = \gamma_{s,d} + \gamma_{s,r,d}, \tag{13}$$

where $\gamma_{s,d}$ and $\gamma_{s,r,d}$ are the instantaneous SNRs of the direct link between source and destination (s,d), and cooperative link between source and destination through relay (s,r,d), respectively. This assumption is justified by the simulation results in Section 6.2 as the EMRC scheme performs close to the ideal MRC scheme.

4.1. Analogy between double-differential and single-differential modulations

In single-differential modulation, $p[n]$ is obtained from $z[n]$ as shown in the first line of (2) with $p[0] = 1$. The received data when $p[n]$ was sent over a channel h without carrier offset is

$$x[n] = \sqrt{\rho} h p[n] + e[n], \quad n = 0, 1, \ldots. \tag{14}$$

The ML decoding of $z[n]$ is performed as follows [10]:

$$\hat{z}[n] = \arg\max_{z \in \mathcal{A}} \mathrm{Re}\{x[n]x^*[n-1]z^*\}. \tag{15}$$

It can be seen by comparing (4) and (15) that the decoding of *double-differentially* modulated signal depends upon $X[n]$ in the similar manner as the decoding of *single-differentially* modulated signal depends upon $x[n]$. Therefore, we can approximate the performance of DDMPSK by the BER expressions of DMPSK with the SNR of $X[n]$ under the assumption that the product of two zero-mean white circularly symmetric Gaussian noise terms in $X[n]$ is also zero-mean white circularly symmetric Gaussian. This connection is shown in more detail in [8, 10, 11]. It can be shown using (3) that

$$\begin{aligned} X[n] = {} & \rho e^{\omega} |h|^2 p[n] + \sqrt{\rho} h e^{\omega n} v[n] e^*[n-1] \\ & + \sqrt{\rho} h^* e^{-\omega n} v^*[n-1] e[n] + e[n] e^*[n-1]. \end{aligned} \tag{16}$$

From (16), we can find the SNR of $X[n]$ as

$$\frac{E_s}{E_N} = \frac{\gamma'}{2 + (\gamma')^{-1}}, \tag{17}$$

where E_s is the signal power, E_N is the total noise power, and γ' is SNR of $x[n]$ in (12). We may further take the following high SNR approximation to maintain the mathematical feasibility of the analysis:

$$\frac{\gamma'}{2 + (\gamma')^{-1}} \approx \frac{\gamma'}{2} - \frac{1}{4}. \tag{18}$$

As a cross-check, we have compared the exact and approximate SNRs in Figure 3, and it is satisfying to see that the approximate SNR follows closely the exact one for $(\gamma' \geq 5)$ dB, which is the region of γ' values of most practical interest.

4.2. Average BER of DDAAF system

From (6), (7), and (8), it can be shown that the SNR of the cooperative link between source and destination through relay (s,r,d) under double-differential modulation is given by

$$\gamma_{s,r,d} = \frac{P_1^2 P_2^2 |h_{s,r}|^4 |h_{r,d}|^4}{\kappa}, \tag{19}$$

where κ is defined in (11). After some manipulations it can be shown that

$$\gamma_{s,r,d} = \frac{\gamma'_{s,r,d}}{2 + (\gamma'_{s,r,d})^{-1}} \approx \frac{\gamma'_{s,r,d}}{2} - \frac{1}{4}, \tag{20}$$

where

$$\gamma'_{s,r,d} = \frac{P_1 P_2 |h_{s,r}|^2 |h_{r,d}|^2}{(P_1 \sigma_{s,r}^2 + P_2 |h_{r,d}|^2 + \sigma^2)\sigma^2}. \tag{21}$$

It can be seen from [19, Equation (6)] that $\gamma'_{s,r,d}$ is the instantaneous SNR of a dual-hop fixed gain relay transmission scheme. Let

$$\begin{aligned} \gamma_{s,d} &= \frac{P_1 |h_{s,d}|^2}{\sigma^2}, \qquad \gamma_{s,r} = \frac{P_1 |h_{s,r}|^2}{\sigma^2}, \\ \gamma_{r,d} &= \frac{P_2 |h_{r,d}|^2}{\sigma^2}, \qquad \overline{\gamma}_{s,d} = \frac{P_1 \sigma_{s,d}^2}{\sigma^2}, \\ \overline{\gamma}_{s,r} &= \frac{P_1 \sigma_{s,r}^2}{\sigma^2}, \qquad \overline{\gamma}_{r,d} = \frac{P_2 \sigma_{r,d}^2}{\sigma^2}. \end{aligned} \tag{22}$$

It can be seen from (1) that for Nakagami-m independent fading channels, $|h_{s,d}|^2$, $|h_{s,r}|^2$, and $|h_{r,d}|^2$ are independent gamma random variables [20] with parameters $(m_{s,d}, 1/\sigma_{s,d}^2)$, $(m_{s,r}, 1/\sigma_{s,r}^2)$, and $(m_{r,d}, 1/\sigma_{r,d}^2)$, respectively.

Theorem 1. *The pdf of $\gamma'_{s,r,d}$ can be written as*

$$f_{\gamma'_{s,r,d}}(\gamma) = \frac{2\gamma^{m_{s,r}-1}}{\Gamma(m_{s,r})\Gamma(m_{r,d})}\left(\frac{m_{s,r}\sigma^2}{P_1\sigma_{s,r}^2}\right)^{m_{s,r}} \times \left(\frac{m_{r,d}\sigma^2}{P_2\sigma_{r,d}^2}\right)^{m_{r,d}} \exp\left(-\frac{m_{s,r}\sigma^2\gamma}{P_1\sigma_{s,r}^2}\right) \times \sum_{k=0}^{\infty}\left\{(-1)^k\frac{[-m_{s,r}]_k}{k!}\left(\frac{P_1\sigma_{s,r}^2+\sigma^2}{\sigma^2}\right)^k \times \left(\frac{m_{s,r}(P_1\sigma_{s,r}^2+\sigma^2)P_2\sigma_{r,d}^2\gamma}{m_{r,d}P_1\sigma_{s,r}^2\sigma^2}\right)^{(m_{r,d}-k)/2} \times K_{m_{r,d}-k}\left(2\sqrt{\frac{m_{s,r}m_{r,d}(P_1\sigma_{s,r}^2+\sigma^2)\sigma^2\gamma}{P_1P_2\sigma_{s,r}^2\sigma_{r,d}^2}}\right)\right\}, \quad (23)$$

where $K_\zeta(\cdot)$ denotes ζth order modified Bessel function of second kind [21, Equation (9.6.2)], $[\cdot]_k$ is Pochhammer's symbol [21, Equation (6.1.22)], and the values of $m_{s,r}$ and $m_{r,d}$ can be noninteger with $m_{s,r}, m_{r,d} \geq 1/2$.

Proof. Theorem 1 can be proved with the help of results given in [22, Section III] and Gauss Hypergeometric series [21, Equation (15.1.1)]. □

It can be seen from (17) and (18) that the SNR of the direct link from source to destination under DD modulation can be expressed as

$$\gamma_{s,d} \approx \frac{\gamma'_{s,d}}{2} - \frac{1}{4}, \quad (24)$$

where $\gamma'_{s,d}$ is the SNR of the link under single-differential modulation.

From the analogy between double- and single-differential modulations in Section 4.1, it follows that the approximate BER expressions of DD modulation can be obtained by replacing the SNR of the single-differential system by the SNR of the DD system. For a single-differential BPSK using two independent (but not identically) distributed channels, the BER conditioned on $\gamma = \gamma_{s,d} + \gamma_{s,r,d}$ is given by [23, Equation (12.1.13)] as

$$P_b(\gamma) = \frac{1}{8}(4+\gamma)\exp(-\gamma). \quad (25)$$

Substituting the values of γ, $\gamma_{s,r,d}$, and $\gamma_{s,d}$ from (13), (20), and (24), respectively, in (25) we can have the BER for DDAAF system as

$$P_b(h_{s,d}, h_{s,r}, h_{r,d}) = \frac{1}{8}\left(\frac{7}{2} + \gamma'_{s,d} + \gamma'_{s,r,d}\right) e^{1/2}\ e^{-(\gamma'_{s,d}+\gamma'_{s,r,d})}. \quad (26)$$

It can be noticed from Sections 4.1 and 4.2 that for obtaining (26) we have taken the following approximations:

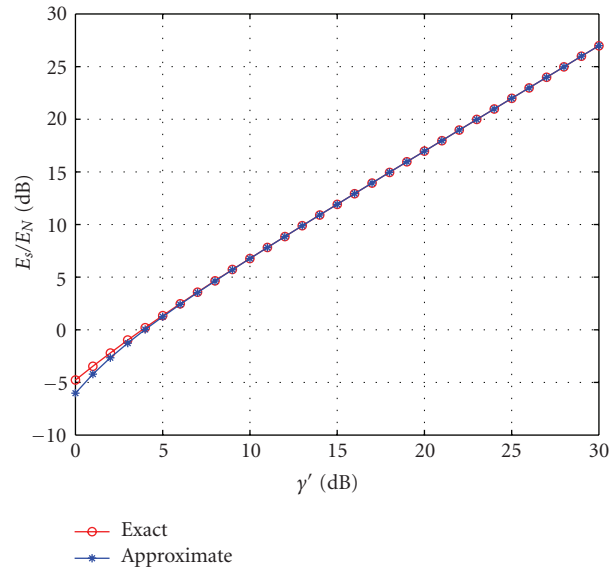


FIGURE 3: Comparison of approximate and exact SNRs. SNR of DD system is represented by E_s/E_N and shown over y-axis and SNR of single-differential system is represented by γ' and shown over x-axis.

the instantaneous SNR of EMRC scheme is assumed to be equal to the instantaneous SNR of MRC scheme, the product of two noise terms in $X_{s,d}[n]$ and $X_{r,d}[n]$ is assumed to be zero-mean white circularly symmetric Gaussian, and the high SNR approximations for $\gamma_{s,r,d}$ and $\gamma_{s,d}$ given in (20) and (24), respectively, are assumed.

Theorem 2. *The approximate BER of the DDAAF system with BPSK modulation averaged over all channels can be written as*

$$P_b = \frac{\exp(1/2)}{16}\{7P_{b1}P_{b2} + P_{b2}P_{b3} + P_{b1}P_{b4}\}, \quad (27)$$

where

$$P_{b1} = \frac{1}{\Gamma(m_{r,d})}\left(\frac{m_{s,r}\sigma^2}{P_1\sigma_{s,r}^2}\right)^{m_{s,r}-1/2}\left(\frac{m_{r,d}\sigma^2}{P_2\sigma_{r,d}^2}\right)^{m_{r,d}-1/2} \times \exp\left(\frac{m_{s,r}m_{r,d}(P_1\sigma_{s,r}^2+\sigma^2)\sigma^2}{P_2\sigma_{r,d}^2(2m_{s,r}\sigma^2+P_1\sigma_{s,r}^2)}\right) \times \sum_{k=0}^{\infty}\left\{(-1)^k\frac{[-m_{s,r}]_k}{k!}\left(\frac{m_{s,r}(\sigma^2+P_1\sigma_{s,r}^2)P_2\sigma_{r,d}^2}{m_{r,d}P_1\sigma_{s,r}^2\sigma^2}\right)^{m_{r,d}-k/2} \times \left(\frac{2m_{s,r}\sigma^2+P_1\sigma_{s,r}^2}{2P_1\sigma_{s,r}^2}\right)^{(k+1-m_{r,d}-2m_{s,r})/2} \times \Gamma(m_{s,r}+m_{r,d}-k)\left(\frac{\sigma^2+P_1\sigma_{s,r}^2}{\sigma^2}\right)^{k-1/2} \times W_{(1+k-2m_{s,r}-m_{r,d})/2,m_{r,d}-k/2}\left(\frac{2m_{s,r}m_{r,d}(P_1\sigma_{s,r}^2+\sigma^2)\sigma^2}{P_2\sigma_{r,d}^2(2m_{s,r}\sigma^2+P_1\sigma_{s,r}^2)}\right)\right\},$$

$$P_{b2} = \left(\frac{2m_{s,d}\sigma^2}{2m_{s,d}\sigma^2 + P_1\sigma_{s,d}^2}\right)^{m_{s,d}},$$

$$P_{b3} = \frac{\Gamma(m_{s,r}+1)}{\Gamma(m_{s,r})\Gamma(m_{r,d})}\left(\frac{m_{s,r}\sigma^2}{P_1\sigma_{s,r}^2}\right)^{m_{s,r}-1/2}\left(\frac{m_{r,d}\sigma^2}{P_2\sigma_{r,d}^2}\right)^{m_{r,d}-1/2} \times \exp\left(\frac{m_{s,r}m_{r,d}(P_1\sigma_{s,r}^2+\sigma^2)\sigma^2}{P_2\sigma_{r,d}^2(2m_{s,r}\sigma^2+P_1\sigma_{s,r}^2)}\right) \times \sum_{k=0}^{\infty}\left\{(-1)^k\frac{[-m_{s,r}]_k}{k!} \times \left(\frac{m_{s,r}(\sigma^2+P_1\sigma_{s,r}^2)P_2\sigma_{r,d}^2}{m_{r,d}P_1\sigma_{s,r}^2\sigma^2}\right)^{(m_{r,d}-k)/2} \times \left(\frac{2m_{s,r}\sigma^2+P_1\sigma_{s,r}^2}{2P_1\sigma_{s,r}^2}\right)^{(k-m_{r,d}-2m_{s,r}-1)/2} \times \Gamma(m_{s,r}+m_{r,d}-k+1)\left(\frac{\sigma^2+P_1\sigma_{s,r}^2}{\sigma^2}\right)^{k-1/2} \times W_{(k-2m_{s,r}-m_{r,d}-1)/2,m_{r,d}-k/2}\left(\frac{2m_{s,r}m_{r,d}(P_1\sigma_{s,r}^2+\sigma^2)\sigma^2}{P_2\sigma_{r,d}^2(2m_{s,r}\sigma^2+P_1\sigma_{s,r}^2)}\right)\right\},$$

$$P_{b4} = \frac{\Gamma(m_{s,d}+1)}{\Gamma(m_{s,d})}\frac{2(2m_{s,d}\sigma^2)^{m_{s,d}}P_1\sigma_{s,d}^2}{(2m_{s,d}\sigma^2+P_1\sigma_{s,d}^2)^{m_{s,d}+1}}, \tag{28}$$

where $W_{\mu,\nu}(\cdot)$ represents Whittaker function [16, Equation (9.220.4)].

Proof. Equation (26) can be averaged over all the channels as follows:

$$P_b = \frac{e^{1/2}}{16}\left\{7\int_0^\infty e^{-\alpha/2}f_{\gamma_{s,r,d}}(\alpha)d\alpha\int_0^\infty e^{-\beta/2}f_{\gamma_{s,d}}(\beta)d\beta + \int_0^\infty \alpha e^{-\alpha/2}f_{\gamma_{s,r,d}}(\alpha)d\alpha\int_0^\infty e^{-\beta/2}f_{\gamma_{s,d}}(\beta)d\beta + \int_0^\infty e^{-\alpha/2}f_{\gamma_{s,r,d}}(\alpha)d\alpha\int_0^\infty \beta e^{-\beta/2}f_{\gamma_{s,d}}(\beta)d\beta\right\}. \tag{29}$$

Each integral in (29) can be represented by P_{bi}, $i \in \{1,2,3,4\}$, as shown by (27). These integrals can be solved by introducing another integration variable and [16, Equations (3.478.1) and (6.631.3)]. □

Corollary 1. *The approximate BER of binary cooperative system with double-differential modulationover Rayleigh channels is given as*

$$P_b = \frac{\exp(\beta+0.5)}{16} \times \left\{\frac{56\sigma^6+36P_1\sigma_{s,d}^2\sigma^4}{(2\sigma^2+P_1\sigma_{s,d}^2)^2(2\sigma^2+P_1\sigma_{s,r}^2)}W_{-1,0.5}(2\beta) + \frac{\sqrt{2}(\sigma^2+P_1\sigma_{s,r}^2)^{1/2}(28\sigma^5+18P_1\sigma_{s,d}^2\sigma^3)}{\sqrt{P_2}\sigma_{r,d}(2\sigma^2+P_1\sigma_{s,d}^2)^2(2\sigma^2+P_1\sigma_{s,r}^2)^{1/2}} \times W_{-0.5,0}(2\beta) + \frac{16P_1\sigma_{s,r}^2\sigma^4}{(2\sigma^2+P_1\sigma_{s,r}^2)^2(2\sigma^2+P_1\sigma_{s,d}^2)}W_{-2,0.5}(2\beta) + \frac{4\sqrt{2}P_1\sigma_{s,r}^2\sigma^4(\sigma^2+P_1\sigma_{s,r}^2)^{1/2}}{\sqrt{P_2}\sigma_{r,d}(2\sigma^2+P_1\sigma_{s,r}^2)^{3/2}(2\sigma^3+P_1\sigma_{s,d}^2\sigma)} \times W_{-1.5,0}(2\beta)\right\}, \tag{30}$$

where

$$\beta = \frac{(\sigma^2+P_1\sigma_{s,r}^2)\sigma^2}{P_2\sigma_{r,d}^2(2\sigma^2+P_1\sigma_{s,r}^2)}. \tag{31}$$

Proof. Substituting $m_{s,d} = m_{s,r} = m_{r,d} = 1$ in (27) and after some manipulations we can obtain (30). □

Let us consider a symmetric case when $P_1 = P_2 = P$, $\sigma_{s,r} = \sigma_{r,d} = \sigma_{s,d} = \sigma_s$, and the SNR of each link is $\gamma_s = P\sigma_s^2/\sigma^2$. From [21, Equation (13.1.33)], the Whittaker function can be expressed as

$$W_{\lambda,\mu}(z) = e^{-z}z^{\mu+0.5}U(0.5+\mu-\lambda, 1+2\mu, z), \tag{32}$$

where $U(\cdot,\cdot,\cdot)$ is the confluent hypergeometric function of second kind. If $\gamma_s \to \infty$, (30) can be written using (32) as

$$P_b \approx \frac{\exp(1/2)}{16}[36\gamma_s^{-3}U(2,2,2\gamma_s^{-1}) + 18\sqrt{2}\gamma_s^{-2}U(1,1,2\gamma_s^{-1}) + 16\gamma_s^{-3}U(3,2,2\gamma_s^{-1}) + 4\sqrt{2}\gamma_s^{-2}U(2,1,2\gamma_s^{-1})]. \tag{33}$$

At high SNR, the probability of error of the DDAAF system can be further approximated by using [21, Equations (13.5.7) and (13.5.9)] as

$$P_b \approx \epsilon\gamma_s^{-2}, \tag{34}$$

where ϵ is a positive constant which is independent of γ_s. It can be seen from (34) that $\lim_{\gamma_s\to\infty}\epsilon\gamma_s^{-2} = 0$, therefore, the DDAAF system achieves diversity of the order of two over the Rayleigh channel.

The approximate BER of DDAAF with M-PSK, $M > 2$, can also be found using the analogy we have developed in Section 4.1. With the help of [23, Equation (B.21)], valid for

single-differential modulation, the BER of DDAAF with M-PSK conditioned on the channels can be written as

$$P_b(\gamma) = \frac{1}{16\pi}\int_{-\pi}^{\pi} u(\psi)\exp\big(-v(\psi)\gamma\big)d\psi, \tag{35}$$

where

$$u(\psi) = \frac{(1-\delta^2)\big[3+\cos(2\psi)-(\delta+1/\delta)\sin\psi\big]}{1+2\delta\sin(\psi)+\delta^2}, \tag{36}$$

$$v(\psi) = \frac{r^2(1+2\delta\sin(\psi)+\delta^2)}{2}, \tag{37}$$

and $\delta = q/r$. For QPSK constellation, $q = \sqrt{2-\sqrt{2}}$ and $r = \sqrt{2+\sqrt{2}}$. For other M-PSK constellations, the values of q and r can be obtained using the results in [23, Appendix B]. The approximate BER of the DDAAF system with M-PSK can be obtained by substituting the value of γ from (13) into (35) and then averaging (35) over the three channels as

$$P_b = \frac{1}{16\pi}\int_{-\pi}^{\pi} u(\psi)M_{\gamma_{s,d}}(v(\psi))M_{\gamma_{s,r,d}}(v(\psi))d\psi, \tag{38}$$

where $M_\gamma(\cdot)$ denotes the moment which is generating function (MGF) of γ.

Theorem 3. *The MGF of $\gamma_{s,r,d}$ is given as*

$$\begin{aligned}
&M_{\gamma_{s,r,d}}(v(\psi)) \\
&= \frac{1}{\Gamma(m_{r,d})}\left(\frac{m_{s,r}\sigma^2}{P_1\sigma_{s,r}^2}\right)^{m_{s,r}-1/2}\left(\frac{m_{r,d}\sigma^2}{P_2\sigma_{r,d}^2}\right)^{m_{r,d}-1/2}\exp(\tau) \\
&\quad\times\sum_{k=0}^{\infty}\Bigg\{(-1)^k\frac{[-m_{s,r}]_k}{k!} \\
&\qquad\times\left(\frac{m_{s,r}(\sigma^2+P_1\sigma_{s,r}^2)P_2\sigma_{r,d}^2}{m_{r,d}P_1\sigma_{s,r}^2\sigma^2}\right)^{(m_{r,d}-k)/2} \\
&\qquad\times\left(\frac{2m_{s,r}\sigma^2+P_1\sigma_{s,r}^2 v(\psi)}{2P_1\sigma_{s,r}^2}\right)^{(k+1-m_{r,d}-2m_{s,r})/2} \\
&\qquad\times\Gamma(m_{s,r}+m_{r,d}-k)\left(\frac{\sigma^2+P_1\sigma_{s,r}^2}{\sigma^2}\right)^{k-1/2} \\
&\qquad\times W_{(1+k-2m_{s,r}-m_{r,d})/2,m_{r,d}-k/2}(2\tau)\Bigg\},
\end{aligned} \tag{39}$$

where

$$\begin{aligned}
\tau &= \frac{P_1\sigma_{s,r}^2(4m_{s,r}m_{r,d}+P_2\sigma_{s,r}^2 v^2(\psi))}{4P_2\sigma_{r,d}^2(2m_{s,r}\sigma^2+P_1\sigma_{s,r}^2 v(\psi))} \\
&\quad+\frac{2m_{s,r}\sigma^2(2m_{r,d}\sigma^2+P_2\sigma_{r,d}^2 v(\psi))}{4P_2\sigma_{r,d}^2(2m_{s,r}\sigma^2+P_1\sigma_{s,r}^2 v(\psi))}.
\end{aligned} \tag{40}$$

Proof. From the analogy between double- and single-differential systems in Section 4.1, it is clear that the MGF of $\gamma_{s,r,d}$ can be obtained by using the formulations of single-differential modulation. Hence, using (20) and (23), the MGF of $\gamma_{s,r,d}$ can be written as

$$M_{\gamma_{s,r,d}}(v(\psi)) = \int_0^\infty \exp\left(-\left(\frac{\gamma}{2}-\frac{1}{4}\right)v(\psi)\right)f_{\gamma'_{s,r,d}}(\gamma)d\gamma. \tag{41}$$

The integral of (41) can be solved by introducing another integration variable and [16, Equations (3.478.1) and (6.631.3)]. □

The MGF of $\gamma_{s,d}$ can be expressed as

$$M_{\gamma_{s,d}}(v(\psi)) = \left(\frac{2m_{s,d}\sigma^2\exp\left(v(\psi)/4m_{s,d}\right)}{2m_{s,d}\sigma^2+P_1\sigma_{s,d}^2}\right)^{m_{s,d}}. \tag{42}$$

A derivation of (42) is provided in Appendix A. It is very difficult to solve the integral in (38) and find a closed-form solution. However, we can obtain an upper bound of the BER of the DDAAF system with M-PSK constellation.

Corollary 2. *The approximate BER of the DDAAF system using M-PSK constellation can be upper bounded as*

$$P_b \le \frac{1}{8}\widetilde{u}M_{\gamma_{s,d}}(\widetilde{v})M_{\gamma_{s,r,d}}(\widetilde{v}), \tag{43}$$

where $\widetilde{u} = (1+\delta)^3/[\delta(1-\delta)]$ and $\widetilde{v} = r^2(1-\delta)^2/2$.

See Appendix B for a proof of Corollary 2.

Let us summarize the analytical results found for DDAAF system in this section. Theorem 1 provides the pdf of the cooperative link from the source to the destination through the relay under general (integer and noninteger values of m) Nakagami-m fading. Theorem 2 suggests approximate analytical BER of DDAAF system with BPSK constellation under Nakagami-m fading. For finding BER expressions, we have assumed that higher-order noise terms in $X_{s,d}[n]$ and $X_{r,d}[n]$ are zero-mean white circularly symmetric Gaussian and made high SNR approximations for $\gamma_{s,r,d}$ and $\gamma_{s,d}$ given in (20) and (24), respectively. The BER of DDAAF system under Rayleigh fading with BPSK constellation is given by Corollary 1. The MGF of the cooperative link under Nakagami-m fading with M-PSK constellation is given by Theorem 3. Corollary 2 provides an upper bound of the BER of the DDAAF system under Nakagami-m fading with M-PSK signal constellation.

5. IMPLEMENTATION OF TRAINING-BASED COOPERATIVE SYSTEM

In this section, we will show how to implement a trained amplification-based cooperative system for comparison with the proposed DDAAF system. Let us assume that the trained decoder at the destination utilizes the *two* initialization symbols as training data, and estimates the carrier offsets and channels using the following maximum likelihood estimators [17, Equations (9.7.27) and (9.7.28)]:

$$\widehat{\omega} = \arg\{x[1]x^*[0]\}, \qquad \widehat{h} = \frac{1}{2}(x[0]+\exp(-2\pi\widehat{\omega})x[1]), \tag{44}$$

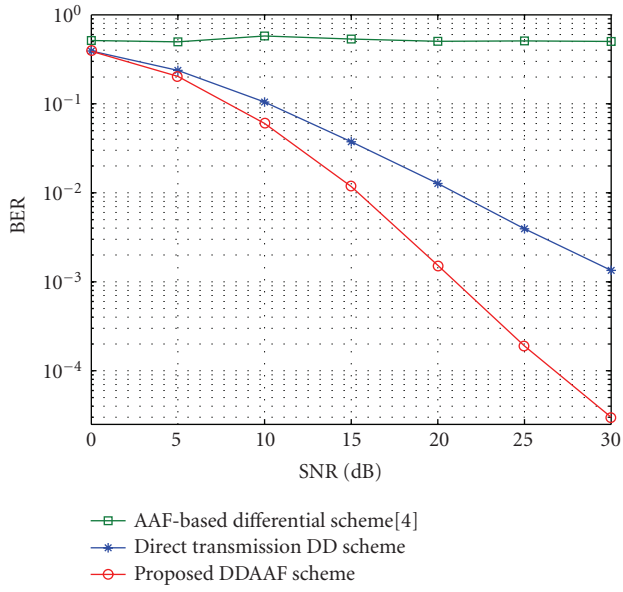


Figure 4: Comparison of BER versus SNR performance of DDAAF cooperative system with conventional schemes.

where $\arg\{\cdot\}$ provides angle of the complex scalar and $(\cdot)^*$ stands for the complex conjugate. The estimators of [17, Equations (9.7.27) and (9.7.28)] are proposed for an $n_t \times N$ space-time block code (STBC) in an MIMO system, where n_t is the number of transmit antennas and N is the time dimension. However, we are working with a cooperative system containing SISO links. Therefore, we use $n_t = N = 1$ in [17, Equations (9.7.27) and (9.7.28)] for obtaining (44). In the trained system, the symbols $z[n]$ are directly transmitted in the space without any differential encoding. Therefore, the received data equations for such a system can be obtained by replacing $v[n]$ by $z[n]$ in (5), (6), and (7). Let us also assume that $z[0] = z[1] = 1$. The receiver at the destination makes the following MRC-based decision variable [23]:

$$d[k] = \frac{\hat{h}^*_{s,d}}{\sigma^2} \exp\left(-2\pi\hat{\omega}_{s,d}\right) y_{s,d}[n] + \frac{\hat{h}^*_{s,r,d}}{E_N} \exp\left(-2\pi\hat{\omega}_{s,r,d}\right) y_{r,d}[m], \quad k = n = m, \tag{45}$$

where $h_{s,r,d}$ is the effective channel over the cooperative link (s, r, d), $\omega_{s,r,d}$ is the effective carrier offset introduced by the cooperative link, and $y_{s,d}[n]$ and $y_{r,d}[m]$ are the data received due to the direct transmission and relayed transmission, respectively, and E_N is the total noise power in $y_{r,d}[m]$, which is given by

$$E_N = \frac{\left(P_1\sigma^2_{s,r} + P_2\left|h_{r,d}\right|^2 + \sigma^2\right)\sigma^2}{P_1\sigma^2_{s,r} + \sigma^2}. \tag{46}$$

From (46), it can be seen that E_N contains $|h_{r,d}|^2$. However, it is difficult to estimate $h_{r,d}$ separately as it can be seen from (7) that the relay transmits an *amplified* version of the *received* signal corresponding to the *training data* transmitted by the source. As the channel statistics vary far more slowly than the channel coefficients, we can assume that the destination has a perfect knowledge of $\sigma^2_{r,d}$. Therefore, the trained decoder can obtain the decision variable by replacing $|h_{r,d}|^2$ by $\sigma^2_{r,d}$ in (46). In addition, it can also be assumed that relay and destination has perfect knowledge about $\sigma^2_{s,r}$, σ^2, P_1, and P_2.

In [24], channel estimation over a single cooperative link between the source and the destination through the relay using amplify-and-forward protocol is studied. It is assumed in [24] that there is no direct link between the source and the destination. However, the proposed training-based cooperative system is more general than [24], since we also consider a direct link between the source and the destination.

6. ANALYTICAL AND EXPERIMENTAL PERFORMANCE EVALUATIONS OF THE DDAAF SYSTEM

All the simulations are achieved by 10^6 channel realizations.

6.1. Comparisons of direct transmission and conventional differential cooperative system [4]

Figure 4 shows the comparison of the performance of the proposed DDAAF-based cooperative scheme, the DD direct transmission, and previously proposed *single differential* amplify-and-forward cooperative scheme [4] for Rayleigh fading channels, that is, $m_{s,d} = m_{s,r} = m_{r,d} = 1$ and BPSK constellation. All the links are assumed to be perturbed by different random carrier offsets uniformly distributed over $[-\pi, \pi)$. It is seen from Figure 4 that the proposed DDAAF scheme outperforms the direct DD transmission at all SNRs. It can be seen from Figure 4 that the proposed scheme has higher diversity as compared to the direct transmission scheme and a performance gain of more than 5 dB is observed at SER = 10^{-2}. It can also be observed that there is a collapse in the performance of the conventional differential scheme [4] because of the random carrier offsets.

6.2. Comparison of analytical and experimental performances

Figure 5 shows the analytical and experimental performances of the proposed DDDAF-based cooperative scheme with random carrier offsets. We have plotted the approximate analytical BER (26) for BPSK constellation, $P_1/P = P_2/P = 0.5$, $\sigma^2_{s,d} = \sigma^2_{s,r} = \sigma^2_{r,d} = 1$, and $m_{s,d} = m_{s,r} = m_{r,d} \in \{1, 1.5, 2, 3\}$. The simulation results of the proposed DDAAF scheme are shown under the same conditions. From Figure 5, it is seen that the experimental data closely follows the analytical results from moderate to high SNR values.Hence, this justifies the assumption taken in (13) and (18).

6.3. Power allocation for DDAAF system

It can be seen from (27) that the BER of the DDAAF system depends nonlinearly upon P_1 and P_2. Therefore, using the power constraint $P_1 + P_2 = P$, we can obtain the values

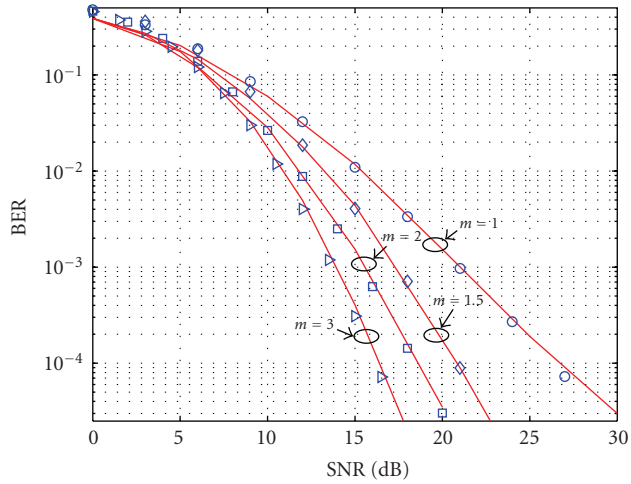


FIGURE 5: Analytical and experimental BERs versus SNR performance of DDAAF cooperative system over Nakagami-m channels with ∘ $m = 1$, ◇ $m = 1.5$, □ $m = 2$, and Δ $m = 3$.

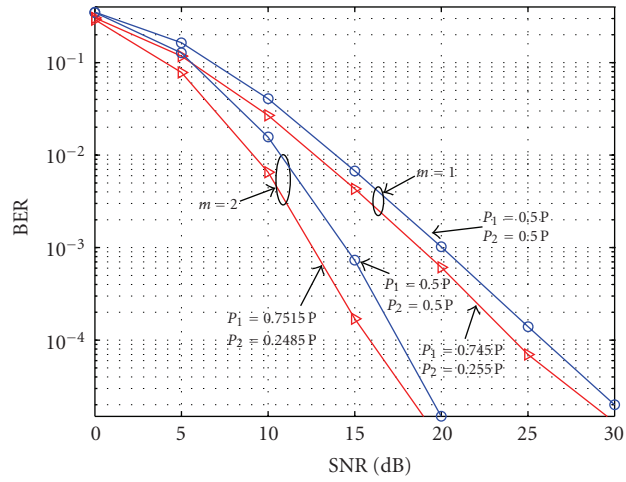


FIGURE 6: BER versus SNR performance of DDAAF cooperative system over Nakagami-m channels with uniform ∘ and optimized Δ power distribution to minimize the BER, $\sigma^2_{s,d} = \sigma^2_{s,r} = 1$, and $\sigma^2_{r,d} = 10$.

of P_1 and P_2 which minimize the BER. We have calculated the power distribution for SNR = 20 dB by numerically minimizing (27) subject to the power constraint $P = P_1 + P_2 = 2$, $\sigma^2_{s,d} = \sigma^2_{s,r} = 1$, and $\sigma^2_{r,d} = 10$. Figure 6 shows the performance of the proposed DDAAF scheme using uniform and numerically calculated power allocation over Nakagami-m channels with $m \in \{1, 2\}$. It can be seen from Figure 6 that the DDAAF scheme with optimized power distribution outperforms the DDAAF scheme with uniform power distribution $P_1 = P_2 = 0.5P$.

6.4. *Comparison of DDAAF with trained cooperative system*

Figure 7 shows the comparison of the proposed DDAAF system with the trained cooperative system from Section 5. The simulations are performed using the QPSK constellation, $P_1 = P_2 = 1$, and Rayleigh fading channels with $\sigma^2_{s,d} = \sigma^2_{s,r} = \sigma^2_{r,d} = 1$. It can be seen that the proposed DDAAF system outperforms the trained cooperative system of Section 5 for all SNR values. The upper bound of the BER of the proposed DDAAF system is calculated from (43) and is also plotted in Figure 7. We have also numerically calculated the power distribution $P_1 = 0.69P, P_2 = 0.31P$, which minimizes the upper bound of (43) at SNR = 20 dB. It can be seen that the DDAAF system with optimized power distribution performs better than the one with uniform power distribution. The performance of an AAF system with perfect channel state information (CSI) and perfect carrier offset knowledge (COK) at the relay and the destination is also shown in Figure 7. It can be seen from Figure 7 that the proposed DDAAF system performs approximately 7.5 dB poorer than the *ideal* AAF system at BER of 10^{-2}. However, with optimized power allocation, 1 dB improvement can be obtained at BER of 10^{-2}, as shown in Figure 7.

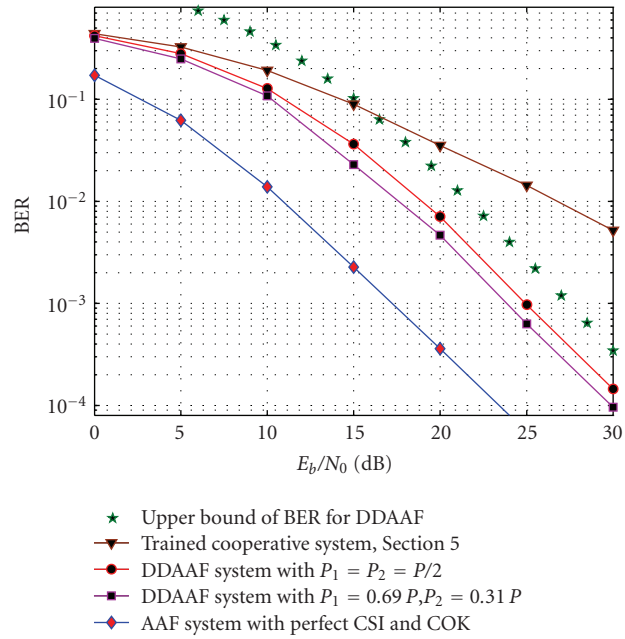


FIGURE 7: Comparison of DDAAF cooperative system with trained cooperative system.

7. CONCLUSIONS

We have implemented double-differential modulation in cooperative communication system with amplify-and-forward protocol. The proposed double-differentially modulated cooperative system can overcome the problem of carrier offsets in Nakagami-m fading channels. Our scheme performs well in the practical scenario, where the conventional differential modulation schemes fail. With our scheme, the users are still able to decode their data without knowing the channel gains or carrier offsets. We have also performed

the BER analysis to predict the behavior of the cooperative system. In addition, we have done a numerical power allocation based on this analysis to further improve the performance of the system. The proposed double-differential system also outperforms the similar rate trained cooperative system.

APPENDICES

A. PROOF OF (42)

From (1), (24), and using the analogy between double- and single-differential systems in Section 4.1, the MGF of $\gamma_{s,d}$ can be written as

$$M_{\gamma_{s,d}}(\nu(\psi)) = \int_0^\infty \exp\left(-\left(\frac{\gamma}{2} - \frac{1}{4}\right)\nu(\psi)\right) f_{\gamma'_{s,d}}(\gamma)d\gamma. \tag{A.1}$$

The integral of (A.1) can be solved with the help of [16, Equations (3.351.3)].

B. PROOF OF COROLLARY 2

In order to obtain an upper bound of BER, we need to maximize $u(\psi)$ and minimize $\nu(\psi)$ with respect to ψ. It can be observed from (37) that $\nu(\psi)$ has its minimum value at $\psi = -\pi/2$. To maximize $u(\psi)$, we need to find the first-order derivative of $u(\psi)$ with respect to ψ and equate it to zero. The first-order derivative of $u(\psi)$ can be written as

$$\begin{aligned}\frac{\partial u(\psi)}{\partial \psi} = \frac{(1-\delta^2)}{a}\Bigg[& -2(1+\delta^2)\sin(2\psi) \\ & -4\delta\sin\psi\sin(2\psi) - \left(\frac{\delta+1}{\delta}\right)\cos\psi \\ & +2\delta\left(\frac{\delta+1}{\delta-1}\right)\sin\psi\cos\psi \\ & -\delta(\delta^2+7)\cos\quad\psi - 2\delta\cos\psi\cos(2\psi)\Bigg],\end{aligned} \tag{B.2}$$

where $a = (1+2\delta\sin\psi+\delta^2)^2$. It can be seen from (B.2) that $u(\psi)$ is maximized at $\psi = -\pi/2$. The maximization can be verified by plotting $u(\psi)$ versus ψ graph. Therefore, the upper bound over BER can be obtained at $\psi = -\pi/2$. By substituting $\psi = -\pi/2$ in (36) and (37), we obtain the values of $\widetilde{u}$ and $\widetilde{\nu}$, respectively. Then, from (38), (39), and (42), we can obtain (43).

ACKNOWLEDGMENT

This work was supported by the Research Council of Norway, Project no. 176773/S10 called OptiMO, which belongs to the VERDIKT program.

REFERENCES

[1] A. Sendonaris, E. Erkip, and B. Aazhang, "User cooperation diversity—part I: system description," *IEEE Transactions on Communications*, vol. 51, no. 11, pp. 1927–1938, 2003.

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

[3] J. N. Laneman, D. N. C. Tse, and G. W. Wornell, "Cooperative diversity in wireless networks: efficient protocols and outage behaviour," *IEEE Transactions on Information Theory*, vol. 50, no. 12, pp. 3062–3080, 2004.

[4] T. Himsoon, W. Su, and K. J. R. Liu, "Differential transmission for amplify-and-forward cooperative communications," *IEEE Signal Processing Letters*, vol. 12, no. 9, pp. 597–600, 2005.

[5] H. Li and Q. Zhao, "Distributed modulation for cooperative wireless communications," *IEEE Signal Processing Magazine*, vol. 23, no. 5, pp. 30–36, 2006.

[6] Q. Zhao and H. Li, "Performance of differential modulation with wireless relays in Rayleigh fading channels," *IEEE Communications Letters*, vol. 9, no. 4, pp. 343–345, 2005.

[7] Q. Zhao and H. Li, "Performance analysis of an amplify-based differential modulation for wireless relay networks under Nakagami-*m* fading channels," in *Proceedings of the 6th IEEE Workshop on Signal Processing Advances in Wireless Communications (SPAWC '05)*, pp. 211–215, New York, NY, USA, June 2005.

[8] M. K. Simon and D. Divsalar, "On the implementation and performance of single and double differential detection schemes," *IEEE Transactions on Communications*, vol. 40, no. 2, pp. 278–291, 1992.

[9] P. Stoica, J. Liu, and J. Li, "Maximum-likelihood double differential detection clarified," *IEEE Transactions on Information Theory*, vol. 50, no. 3, pp. 572–576, 2004.

[10] M. K. Simon, S. M. Hinedi, and W. C. Lindsey, *Digital Communication Techniques: Signal Design and Detection*, Prentice-Hall, Upper Saddle River, NJ, USA, 1994.

[11] M. R. Bhatnagar and A. Hjørungnes, "SER expressions for double differential modulation," in *Proceedings of the IEEE Information Theory Workshop on Information Theory for Wireless Networks*, pp. 203–207, Bergen, Norway, July 2007.

[12] Z. Liu, G. B. Giannakis, and B. L. Hughes, "Double differential space-time block coding for time-selective fading channels," *IEEE Transactions on Communications*, vol. 49, no. 9, pp. 1529–1539, 2001.

[13] A. Cano, E. Morgado, A. Caamano, and J. Ramos, "Distributed double-differential modulation for cooperative communications under CFO," in *Proceedings of IEEE Global Communications Conference (GLOBECOM '07)*, pp. 3447–3441, Washington, DC, USA, November 2007.

[14] M. Nakagami, "The *m*-distribution: a general formula of intensity distribution of rapid fading," in *Statistical Methods in Radio Wave Propagation*, pp. 3–36, Pergamon Press, Oxford, UK, 1960.

[15] M. K. Simon and M.-S. Alouini, *Digital Communication over Fading Channels*, John Wiley & Sons, Hoboken, NJ, USA, 2005.

[16] I. S. Grandshteyn and I. M. Ryzhik, *Table of Integrals, Series, and Products*, Academic Press, San Diego, Calif, USA, 6th edition, 2000.

[17] E. G. Larsson and P. Stoica, *Space-Time Block Coding for Wireless Communications*, Cambridge University Press, Cambridge, UK, 2003.

[18] D. G. Brennan, "Linear diversity combining techniques," *Proceedings of the IEEE*, vol. 91, no. 2, pp. 331–356, 2003.

[19] M. O. Hasna and M.-S. Alouini, "A performance study of dual-hop transmissions with fixed gain relays," *IEEE Transactions on Wireless Communications*, vol. 3, no. 6, pp. 1963–1968, 2004.

[20] A. Papoulis, *Probability, Random Variables, and Stochastic Processes*, McGraw-Hill, Singapore, 3rd edition, 1991.

[21] M. Abramowitz and I. A. Stegun, *Handbook of Mathematical Functions*, Dover, New York, NY, USA, 1972.

[22] T. A. Tsiftsis, G. K. Karagiannidis, P. T. Mathiopoulos, and S. A. Kotsopoulos, "Nonregenerative dual-hop cooperative links with selection diversity," *EURASIP Journal on Wireless Communications and Networking*, vol. 2006, Article ID 17862, 8 pages, 2006.

[23] J. G. Proakis, *Digital Communications*, McGraw-Hill, Singapore, 4th edition, 2001.

[24] C. S. Patel and G. L. Stüber, "Channel estimation for amplify and forward relay based cooperation diversity systems," *IEEE Transactions on Wireless Communications*, vol. 6, no. 6, pp. 2348–2356, 2007.

Hindawi Publishing Corporation
EURASIP Journal on Advances in Signal Processing
Volume 2008, Article ID 736818, 9 pages
doi:10.1155/2008/736818

Research Article

Delay Optimization in Cooperative Relaying with Cyclic Delay Diversity

Slimane Ben Slimane, Bo Zhou, and Xuesong Li

Radio Communication Systems, Department of Communication Systems, The Royal Institute of Technology (KTH), Electrum 418, 164 40 KISTA, Sweden

Correspondence should be addressed to Slimane Ben Slimane, slimane@radio.kth.se

Received 1 November 2007; Accepted 5 March 2008

Recommended by J. Wang

Cooperative relaying has recently been recognized as an alternative to MIMO in a typical multicellular environment. Inserting random delays at the nonregenerative fixed relays further improve the system performance. However, random delays result in limited performance gain from multipath diversity. In this paper, two promising delay optimization schemes are introduced for a multicellular OFDM system with cooperative relaying with stationary multiple users and fixed relays. Both of the schemes basically aim to take the most advantages of the potential frequency selectivity by inserting predetermined delays at the relays, in order to further improve the system performance (coverage and throughput). Evaluation results for different multipath fading environments show that the system performance with delay optimization increases tremendously compared with the case of random delay.

1. INTRODUCTION

A practical method called cooperative communications has been proposed recently in order to approach the theoretical limits of MIMO technology [1]. Mobile units or relays cooperate by sharing their antennas, so as to create a *virtual MIMO* system [2], thus enabling to exploit diversity and reducing end-to-end path loss [3].

To achieve greater coverage and capacity, relaying has been proved to be a valuable alternative [4–7] for future generations of wireless networks. There are fundamentally two kinds of relays depending upon whether the received signal is only amplified and forwarded or is processed before forwarding, the former is called a nonregenerative relay (amplify-and-forward relay) and the later is called a regenerative relay (decode-and-forward relay). A relay can also be mobile or stationary.

Inserting delays at the relays can make the channel more frequency selective and enhances system performance [8]. These delays can be totally random or can be predetermined. In order to take advantage of the obtained frequency selectivity of the channel, we can either use coded OFDM signalling or single-carrier system with frequency domain equalization [3]. However, the equivalent relay channel will still experience fading dips that may not be resolved by channel coding or equalization. With channel feedback from the mobile unit to the relays, optimal coherent combining of the relayed signals can be obtained and considerable performance improvement can be achieved [6]. However, such an improvement is obtained at the expense of huge feedback information as full channel state information is needed at the different relays.

The aim of this paper is to optimize the cyclic delays in a cooperative OFDM relaying scheme with cyclic delay diversity. Our objective is to improve the coverage and throughput of the system while minimizing the feedback information from the mobile unit to the relay stations. For this purpose, two algorithms are proposed and studied, one is based on the strongest path and the other is based on linear approximation of the channel phase. The obtained results show that both algorithms provide very good performance which make them very promising for future wireless communications.

The paper is organized as follows: Section 2 presents the cellular/relay system model. Section 3 introduces the

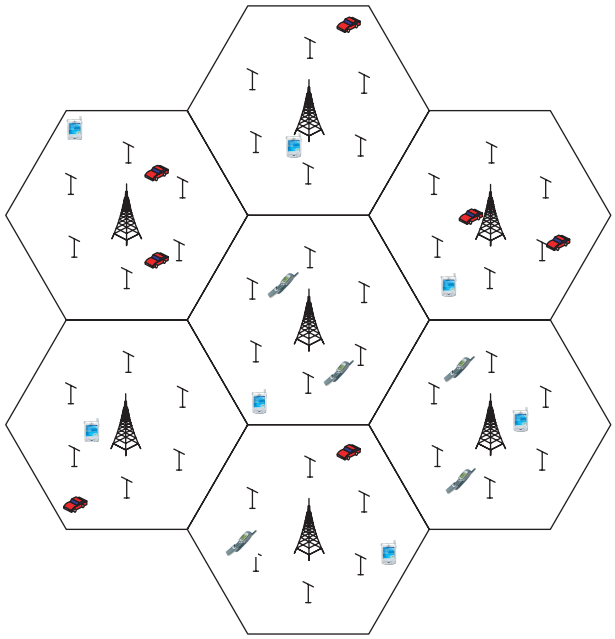


Figure 1: System layout of cooperative relaying communication.

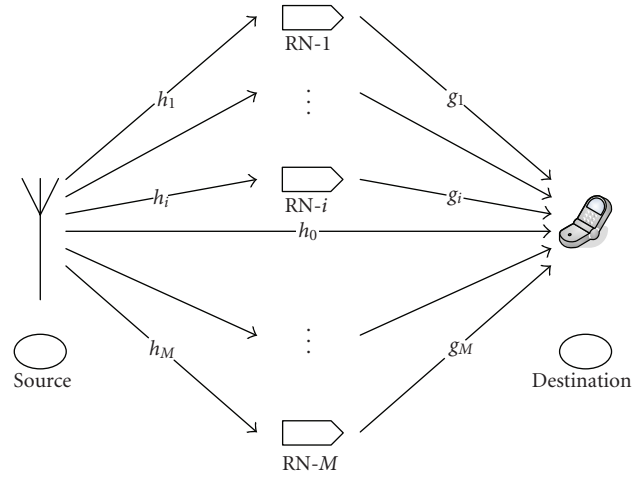


Figure 2: Cooperative relaying communication in a single cell.

delay optimization procedure used in this paper where two different algorithms are given. Section 4 gives a mathematical model of the received signal-to-interference+noise-ratio (SINR) as a function of the number of relays and the different radio channels. Section 5 gives some numerical results to illustrate the behaviour of the algorithms and their performance. Section 6 summarizes the work and provide some suggestions for further studies.

2. SYSTEM MODEL

Figure 1 shows the cellular/relay system where each cell consists of a base-station at the center of the cell with omnidirectional antenna and M relays (placed at half the distance from the boundary of the cell). Mobile users are uniformly distributed over the service area. We limit our study to the downlink and assume an OFDM access scheme where the same frequency is used in all the cells (reuse 1).

To better illustrate the system, the communication link within one cell is shown in Figure 2. We assume that the relays operate in a duplex mode where the first time slot is used to receive the OFDM signal from the base station and the second time slot is used to forward a cyclic delayed version (blockwise) of the signal to the mobile unit while the base station is silent. We assume nonregenerative relays where each relay introduces a predetermined cyclic delay, amplifies the signal, and then forwards it. Hence the mobile unit receives two versions of the useful signal that can be combined using maximum ratio combining (MRC) before decoding.

The cyclic delay is usually assumed predetermined or totally arbitrary [2, 3]. In this paper, we try to identify the delays that can be used at the different relays such that the system throughput is improved.

3. DELAY OPTIMIZATION ALGORITHM

Random cyclic delays at the relays do not make the full use of the multipath channel's feature since it increases the frequency selectivity of the radio channel, but does not remove the fading dips. To optimize the delays at a given relay, some information about the channel state between the relay and the mobile unit is needed at the relay. Perfect knowledge of the channel state will provide the best performance, but at the expense of a huge overhead where the channel transfer function at each OFDM subcarrier needs to be sent to the relay [6]. In this paper, we try to reduce this overhead by considering the dominant part of the channel only.

3.1. Delay optimization based on the strongest path

Inserting random delay does not make the full use of the multipath channel's feature. An optimal delay allocation approach using coherent combining in large-scale cooperative relaying networks was introduced in [6], but it is well suited for unlimited feedback communications with perfect knowledge of the channel, which is hard to achieve in the practical case. Besides, what we gain from the delay optimization will be lost on the feedback concerning the spectrum efficiency. Although we obtain an optimal delay through this scheme, it comes at expense of the feedback information required.

In this paper, we only take the best segment of the signal from each relay into account and thus only a fractional feedback is required. The benefit of this is with low complexity of the system and high spectrum efficiency; a significant performance gain can be obtained by making the most of the frequency and delay diversity. The idea is locate the strongest path from each relay, cophase it at the relay, and adjust the cyclic delay such that they are in phase and aligned at the mobile unit. This procedure will increase the power of one path of the equivalent relay channel and average the powers of the other paths making the equivalent relay

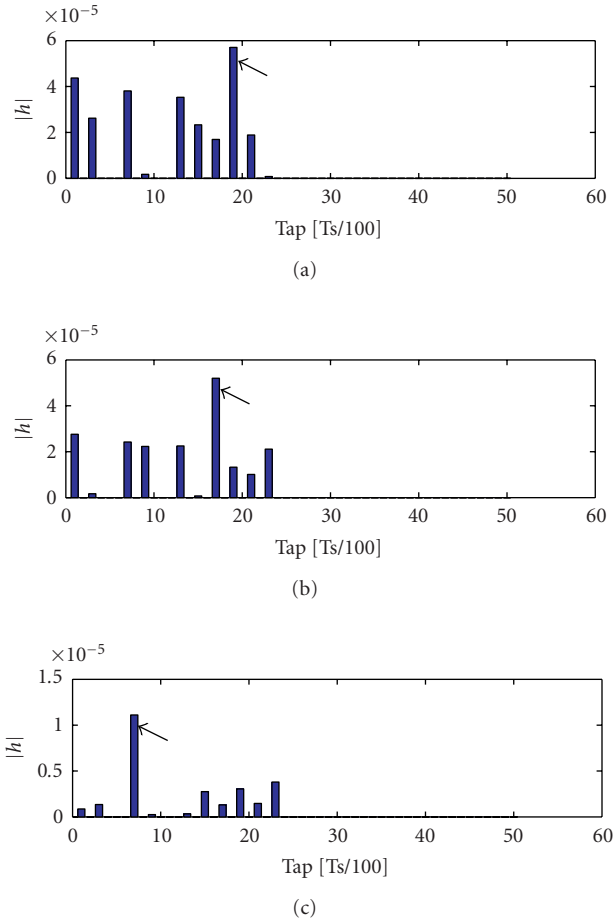


Figure 3: Impulse responses of the initial relay channel i, $i = 1, 2, 3$ in a typical urban environment.

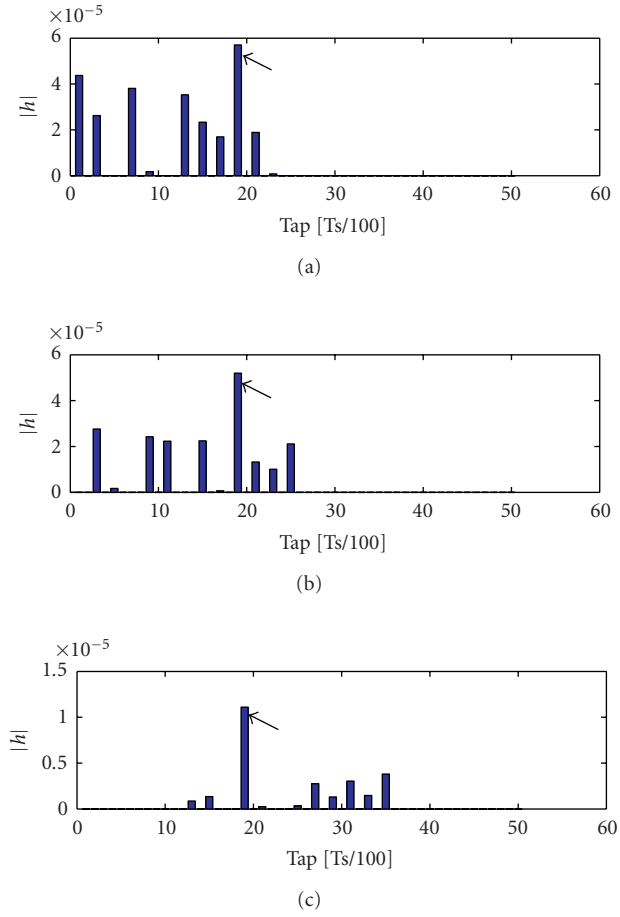


Figure 4: Impulse responses of the three relay channels after delay optimization in a typical urban environment.

channel appears as Rician fading. Hence each relay requires information about the time delay of the strongest path and its phase only. The benefit of this is a good diversity gain with very limited feedback information.

In order to give a basic introduction to this scheme, let us consider the case of three relay stations with the channel impulse responses shown in Figure 3. The strongest path of each channel is indicated with an arrow.

With the received signals from the relays, the algorithm operates in the following way.

(i) The receiver (MS) locates the strongest path from the three different relays, generates an index of their locations $\{l_1, l_2, l_3\}$ and phases as $\{\theta_1, \theta_2, \theta_3\}$, and feed them back to the relays;

(ii) Based on the feedback information about the position of the strongest path and its phase, each relay introduces the proper cyclic shift (optimized delay) to the signal so as the peaks of all signals are aligned at the receiver.

Figure 4 shows how the different channels appear at the receiver side after delay optimization where it is observed that the strongest paths of the different channels are now aligned in time.

(iii) Each relay compensates for the phase of the strongest path such that when the different signals are multiplexed in the air, they will add coherently at the receiver. Hence the total received power of the useful signal will be enhanced. Figure 5 shows the resulting equivalent low pass of the fading multipath relay channel where it is observed that the strongest paths have been added coherently while the secondary paths have been averaged out and kept low values.

3.2. *Linear approximation of the channel phase*

The method discussed in Section 3.1 requires channel state information in the time domain which requires an extra IFFT operation at the mobil unit since channel estimation for OFDM is usually done in the frequency domain. One way to avoid this is by investigating and approximating the channel transfer function phase directly.

Based on the multipath fading channel model, the frequency selective channel can be written as

$$h(t) = \sum_{k=0}^{K-1} h_k \delta(t - \tau_k). \tag{1}$$

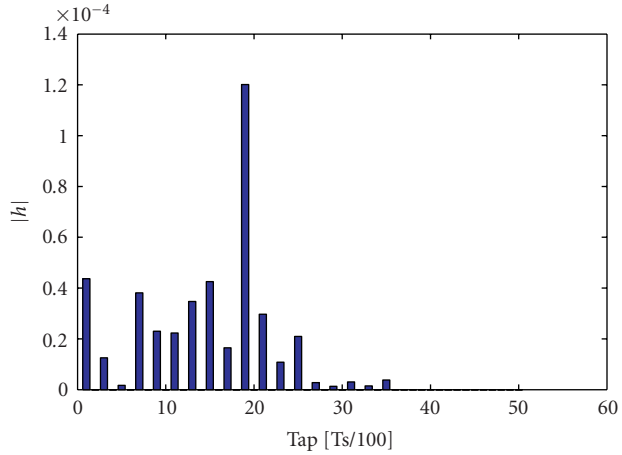


FIGURE 5: Ultimate channel impulse response of the equivalent relay channel.

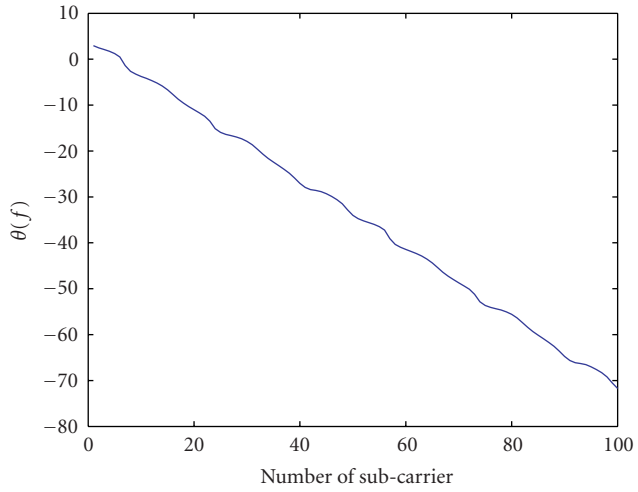


FIGURE 6: Phase variation of multipath fading channel in a typical urban environment with respect to OFDM spectrum.

Assuming h_k is slowly varying, the channel transfer function between relay i and the mobile unit can be approximated as

$$H_i(f) = \sum_{k=0}^{K-1} h_k e^{-j2\pi f \tau_k} = |H_i(f)|\, e^{-j\theta_i(f)}. \tag{2}$$

Figure 6 shows how the $\theta_i(f)$ varies with respect to the frequency f. From Figure 6, we notice that the phase $\theta_i(f)$ can be approximated as a linear function:

$$\theta_i(f) = \theta_i - 2\pi\tau_i f, \tag{3}$$

where $-2\pi\tau_i$ is the slope of the phase and θ_i is a constant phase. Thus the corresponding channel in frequency domain is given by

$$\begin{aligned} H_i(f) &= |H_i(f)|\, e^{j\theta_i(f)} \\ &\approx |H_i(f)|\, e^{j(-2\pi f\tau_i + \theta_i)}. \end{aligned} \tag{4}$$

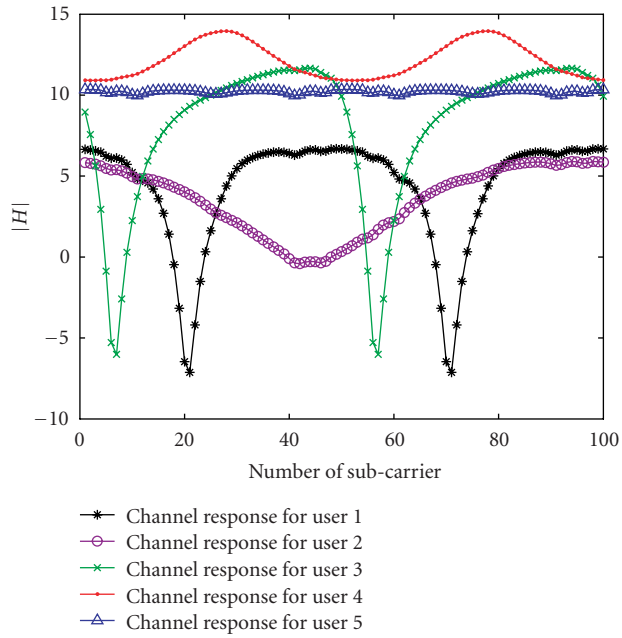


FIGURE 7: Channel responses in a typical urban environment experienced by 5 different users.

Based on the above formulas, the optimized delay τ_i for relay i can be approximated as

$$\tau_i(f) = -\frac{1}{2\pi}\frac{d\theta_i(f)}{df} \approx \tau_i. \tag{5}$$

Having the estimated delay and the initial phase, each relay channel will make the necessary cophasing and cyclic shifting before signal forwarding. The cyclic delayed signals from the different relays multiplex in the air providing an overall received signal with higher signal amplitude as compared to the case of no delay optimization.

3.3. Scheduling

Multiaccess scheme is required to arrange the multiple users sharing the limited resource. In the interest of maximizing the spectrum efficiency thus to limit the cost of the system, which is the main issue from the operators' standpoint [9], an OFDMA scheme with frequency scheduling is considered here.

It should be noted that this scheduling scheme is implemented with priority: one has to give the first priority to the user suffering the most frequency selective channel (with the highest standard deviation) and give second priority to the user having the second highest standard deviation, and so on. This is not the optimal channel allocation algorithm with respect to system throughput, but a fair system from the user's point of view and at the same time the spectrum efficiency remains at a high level.

As illustration of the scheduling scheme, we consider five users per cell. Figure 7 shows the channel frequency response with respect to different users.

By means of the scheduling scheme presented above, we give higher priority to those users who are not more sensitive to the channel, so as to allocate the subcarriers in a more efficient way.

Applying the scheduling algorithm, we notice from Figure 8 that the users are related well with each other on the spectrum with the help of scheduling.

4. MATHEMATICAL MODEL

To model the system, we consider one communication link between the base station and the mobile unit within a given cell. As indicated earlier, the communication is done in two steps: in the first step (first-time slot), the base station transmits information to both mobile unit and the relays and in the second step (second-time slot), the relays forward the information to the mobile unit while the base station is silent.

Considering cell 0, the received signal at the mobile unit directly from base stations (BS) during the first time slot can be written as

$$r_0(t) = \sum_{i=0}^{N_c-1} \sum_{p=0}^{P-1} h_{0,p}^{(i)} s_i\left(t - v_{0,p}^{(i)}\right) + z_0(t), \tag{6}$$

where $s_i(t)$ is the signal coming from base station i, $h_{0,p}^{(i)}$ and $v_{0,p}^{(i)}$ are the channel attenuation and time delay of path p between base station i and the mobile unit, respectively, N_c is the total number of base stations, and $z_0(t)$ represents thermal noise.

Assuming that the base stations are synchronized, the demodulated output sample at subcarrier n can be written as

$$R_{0,n} = H_0^{(0)}(n) s_{0,n} + \sum_{i=1}^{N_c-1} H_0^{(i)}(n) s_{i,n} + Z_{0,n}, \tag{7}$$

where

$$H_0^{(i)}(n) = \sum_{p=0}^{P-1} h_{0,p}^{(i)} e^{-j2\pi v_{0,p}^{(i)} n/T} \tag{8}$$

is the channel transfer function at subcarrier n, $s_{i,n}$ is the received symbol from base station i at subcarrier n, and $Z_{0,n}$ is zero-mean complex Gaussian random variable with variance N_0.

The received signals at the different relays from the base station within cell 0 are given by

$$y_{0,m}(t) = \sum_{i=0}^{N_c-1} \sum_{p=0}^{P-1} c_{m,p}^{(i)} s_i\left(t - \nu_{m,p}^{(i)}\right) + z_m(t), \quad m = 0, 1, \ldots, M-1. \tag{9}$$

Each relay amplifies and retransmits its received signal with the appropriate cyclic delay while the base stations are silent. Hence the received signal at the mobile unit from the different relays during the second time slot can be written as

$$r_1(t) = \sum_{i=0}^{N_c-1} \sum_{m=0}^{M-1} \beta_{i,m} \sum_{p=0}^{P-1} g_{m,p}^{(i)} y'_{i,m}\left(t - \tau_{m,p}^{(i)}\right) + z_1(t), \tag{10}$$

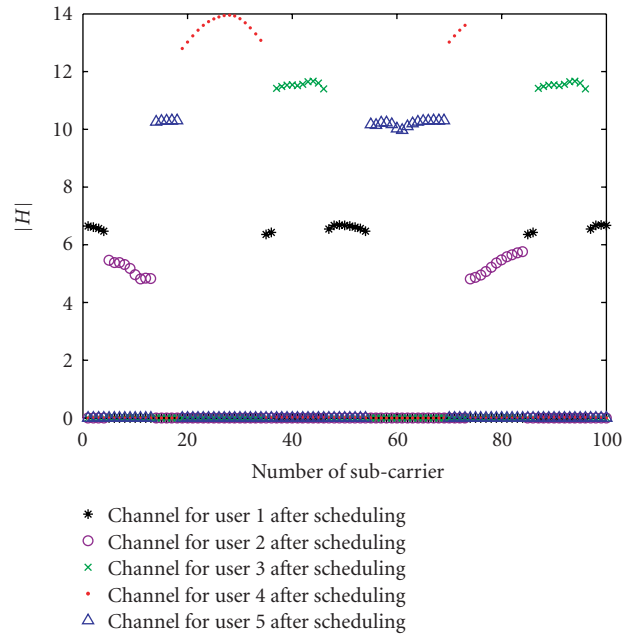


FIGURE 8: Channel allocation to the 5 users after scheduling in a typical urban environment.

where $y'(t)$ is the cyclic delay version (blockwise) of $y(t)$ and $\beta_{i,m}$ is the amplification factor used at relay node m within cell i with

$$\beta_{i,m} = \frac{1}{\sqrt{\sum_{p=0}^{P-1} \left| c_{m,p}^{(i)} \right|^2 + N_0/E_i}}, \tag{11}$$

and $E_i = p_i T$ is the average energy per transmitted symbol of cell i.

Assuming that the relays are synchronized, the demodulated signal sample at subcarrier n can be written as

$$R_{1,n} = H_{1,e}(n) s_{0,n} + \sum_{i=1}^{N_c-1} H_{i,e}(n) s_{i,n} + \sum_{i=1}^{N_c-1} G_{i,e}(n) s_{i,n} + Z_{1,n}, \tag{12}$$

where

$$\begin{aligned} H_{1,e}(n) &= \sum_{m=0}^{M-1} \beta_{0,m} G_m^{(0)}(n) C_m^{(0)}(n) e^{-j(\theta_{0,m} + 2\pi n l_{0,m}/N)}, \\ H_{i,e}(n) &= \sum_{m=0}^{M-1} \beta_{0,m} G_m^{(0)}(n) C_m^{(i)}(n) e^{-j(\theta_{0,m} + 2\pi n l_{0,m}/N)}, \\ G_{i,e}(n) &= \sum_{m=0}^{M-1} \beta_{i,m} G_m^{(i)}(n) \sum_{k=0}^{N_c-1} C_m^{(k)}(n) e^{-j(\theta_{i,m} + 2\pi n l_{i,m}/N)}. \end{aligned} \tag{13}$$

$G_m^{(i)}(n)$ is the channel transfer function between relay m of base station i and the mobile unit at subcarrier n, $C_m^{(i)}(n)$ is the channel transfer function between base station i and its relay m at subcarrier n, $(l_{i,m}, \theta_{i,m})$ are the optimized cyclic

Table 1: Simulation parameters.

Parameter	Value
Number of subcarriers	128
Noise floor	−105 dBm
Fast multipath fading	Urban/suburban
Path loss exponent (α)	3.5
Shadow fading standard deviation	6 dB
Transmit power (base station and relay)	33 dBm
Users per cell	5
Cell radius	500 m/1 km
Channels per cell	20
Number of relays per cell	6

shift and the phase employed at relay m within cell i, and $Z_{1,n}$ is zero-mean complex Gaussian with variance N_0.

Combining the direct received signal in (7) and that from the relays in (12), the signal-to-interference+noise-ratio (SINR) can be written as

$$\Gamma = \frac{|H_0^{(0)}|^2 p_0}{\sum_{i=1}^{N_c-1} |H_0^{(i)}|^2 p_i + N_0 W} + \frac{|H_{1,e}|^2 p_0}{\sum_{i=1}^{N_c-1} (|H_{i,e}|^2 + |G_{i,e}|^2) p_i + N_0 W}, \tag{14}$$

where without loss of generality, we have dropped the subcarrier index n, p_i is the average transmitted power of signal $s_i(t)$, and W is the signal bandwidth.

The throughput is derived from the received SINR using the following expression:

$$R = C_B W \log_2\left(1 + \frac{\Gamma}{2}\right), \tag{15}$$

where $C_B = 1/2$ to account for the half duplex operation of the relay node and the factor 2 is to account for practical implementation of channel coding and modulation.

5. NUMERICAL RESULTS

Numerical evaluation is performed by system simulation of a two-tier (19 cells) hexagonal cellular system with omnidirectional antenna and 6 relay nodes per cell as illustrated in Figure 1. The proposed algorithms are evaluated by snapshot simulation for the OFDMA system. We assume that users are uniformly distributed over the whole cells. The number of active relays for each user is set to $M = 3$. Mobile units are uniformly distributed within the area. The multipath fading channel is modelled as a tapped delay line and based on the models proposed in [10]. A more detailed list of the simulation parameters is given in Table 1.

With fractional feedback, delay optimization based on strongest path further enhances the channel response compared to inserting random delays at relays. Two different types of channel are considered here: (1) flat fading channel and (2) frequency selective fading channel. By adding predetermined delays and retransmitting the signal with proper amplification at relay nodes, this delay diversity scheme leads a substantial improvement to the system performance.

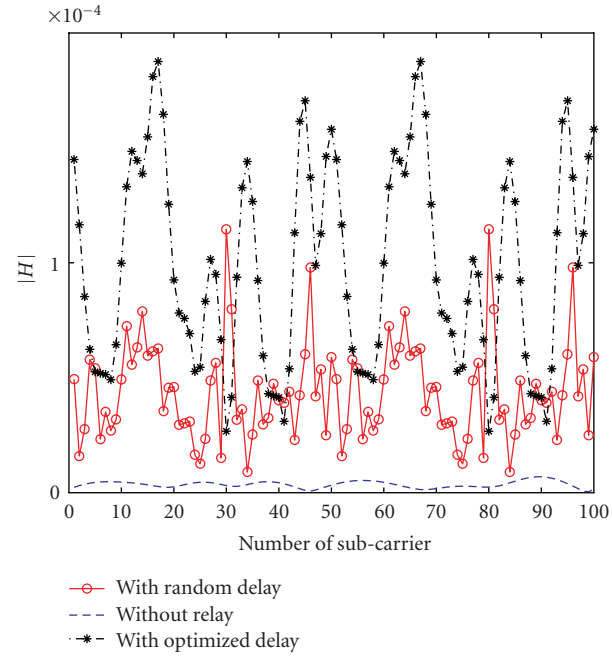


Figure 9: A snapshot of a frequency selective channel before and after adding cyclic delays for the case of three active relays.

By properly selecting the cyclic delay for each relay node, we expect to get a good relay channel that can improve the communication link of the mobile unit. Figure 9 illustrates the channel transfer function of the relay channel with and without cyclic delay diversity for a typical urban environment. It is observed that the initial channel has been improved and the optimized delays have improved the channel gains of the different OFDM subcarriers which make the channel more robust as compared to the case with random delays.

A performance improvement of the OFDMA scheme with frequency scheduling is then expected. As our objective is to assess the performance of the optimized delay scheme, we limit our study to the case of having the same statistical channels between the source and relays, as well as between the relays and the mobile unit. We have investigated the performance of our system in a typical urban and rural area environments [10].

The number of active relays within the cell will affect the received SINR experienced by the user. Figure 10 shows the received SINR at 5 percentile for a given user and with different number of active relays. We notice that having 3 active relays is a good compromise between increased received power and experienced interference.

As we can see from Figure 11, the performance can be improved by increasing the total number of relays per cell, but it can be noted that with more than six relays the system performance has not been improved much. Due to

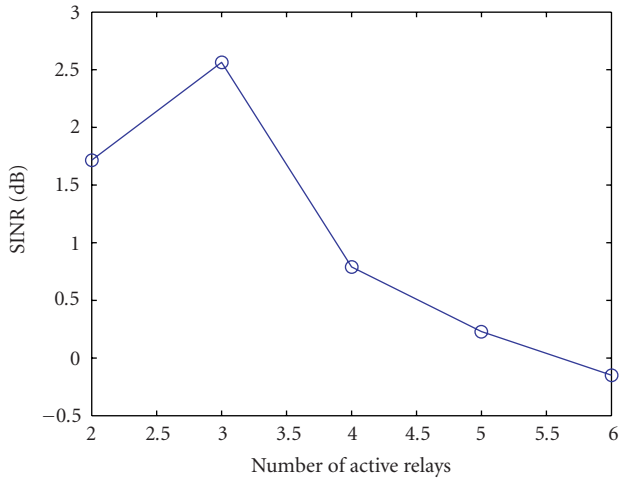


Figure 10: The received SINR at 5 percentile with the different number of active relays on an urban environment with a cell radius of 500 m.

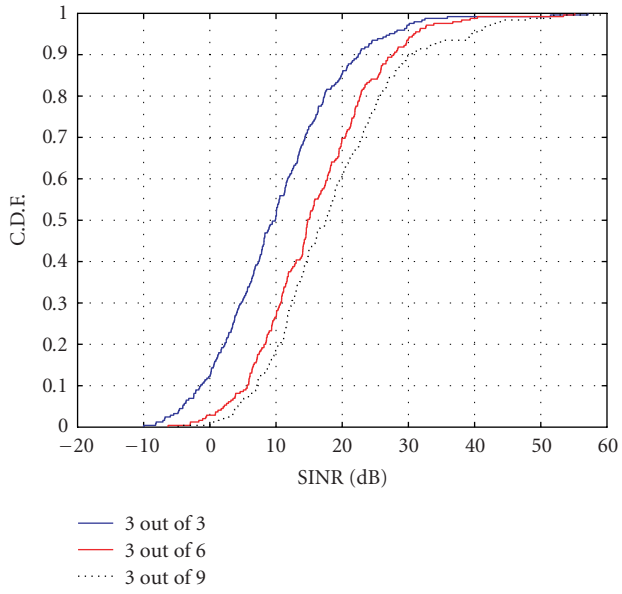


Figure 11: Cumulative distribution function of the received SINR with the different number of total relays on an urban environment with a cell radius of 500 m.

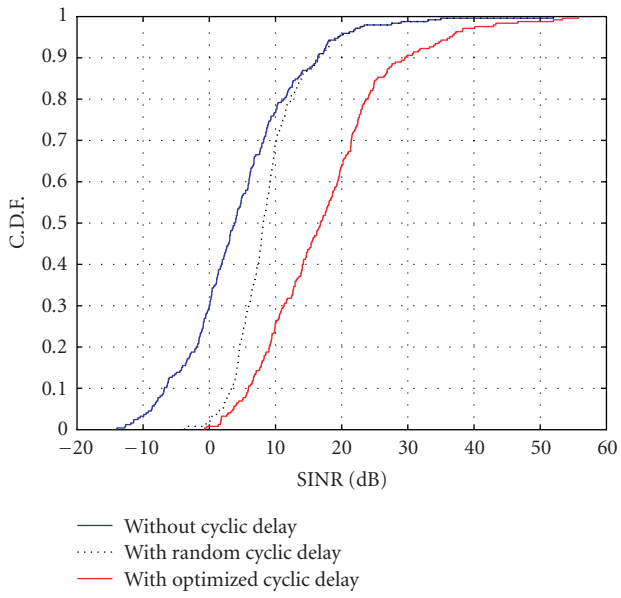


Figure 12: Cumulative distribution function of the received SINR on an urban environment with a cell radius of 500 m.

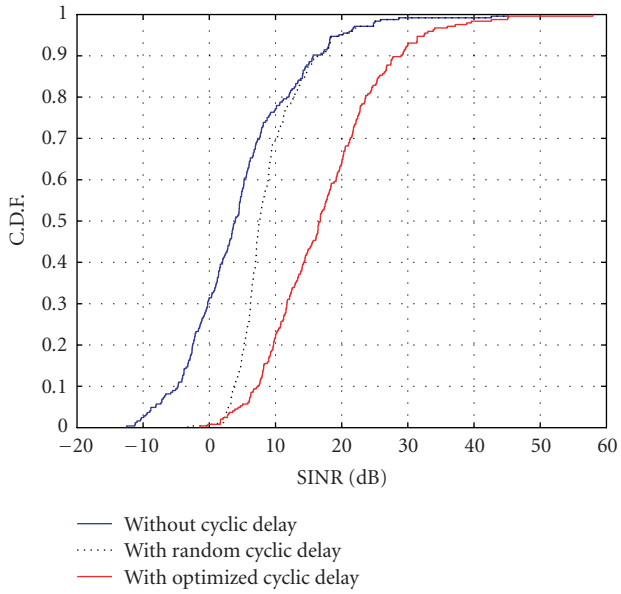


Figure 13: Cumulative distribution function of the received SINR on a rural environment with a cell radius of 1000 m.

the infrastructure cost issue, we considered six relays per cell and we assumed that only three are active at a time. The following simulations are based on this relay selectivity scheme.

Figure 12 shows the cumulative distribution function (CDF) of the combined received SINR with and without delay diversity over an urban environment when the optimized delay is based on the strongest path and with three active relays out of 6 relays. Clearly, the optimized delay algorithm improves the system performance considerably. An improvement of about 3 dB at 5 percentile of the CDF compared to random delay is obtained.

In a rural environment, we can see that (Figure 13) the system has also been greatly improved by about 3 dB at 5 percentile of the CDF when introducing optimized delay as compared to the random delay scheme.

Comparing the results of Figures 12 and 13, we notice that when the cell radius increases, the performance of system with optimized delay still remains at a high level. This feature offers us a good solution to guarantee the

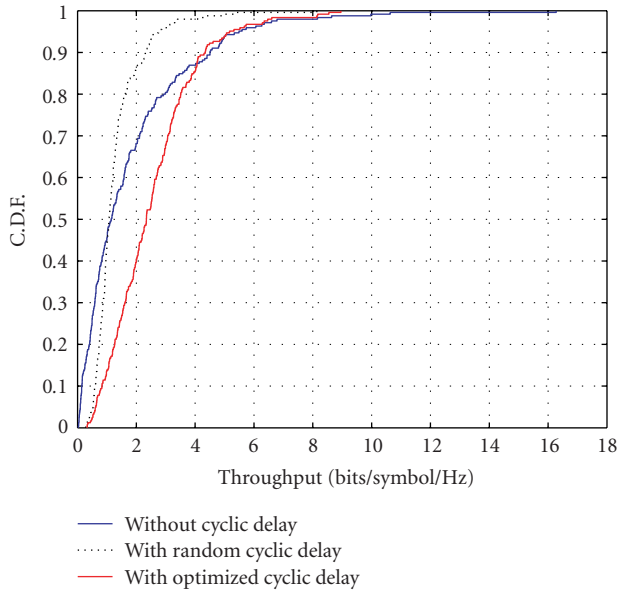


Figure 14: Normalized system throughput on an urban environment with a cell radius of 500 m.

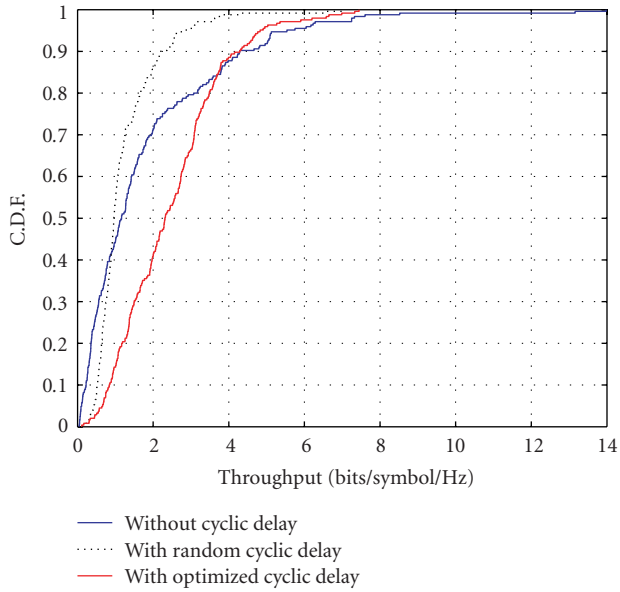


Figure 15: Normalized system throughput on a rural environment with a cell radius of 1000 m.

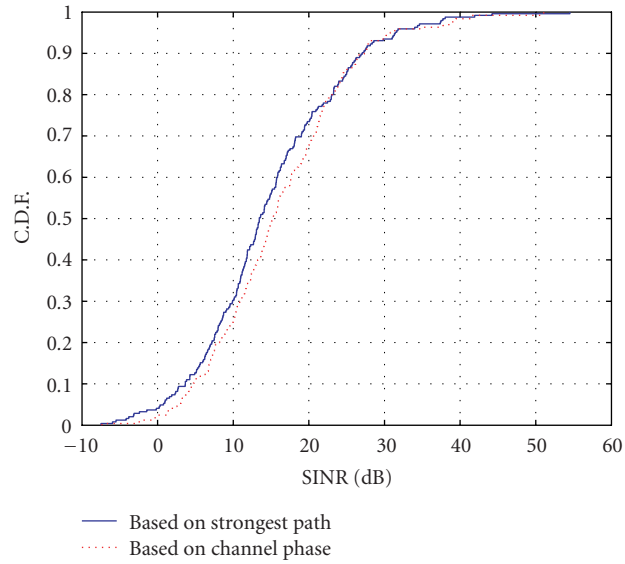


Figure 16: Cumulative distribution function of the received SINR for the two optimized algorithms on an urban environment with a cell radius of 500 m.

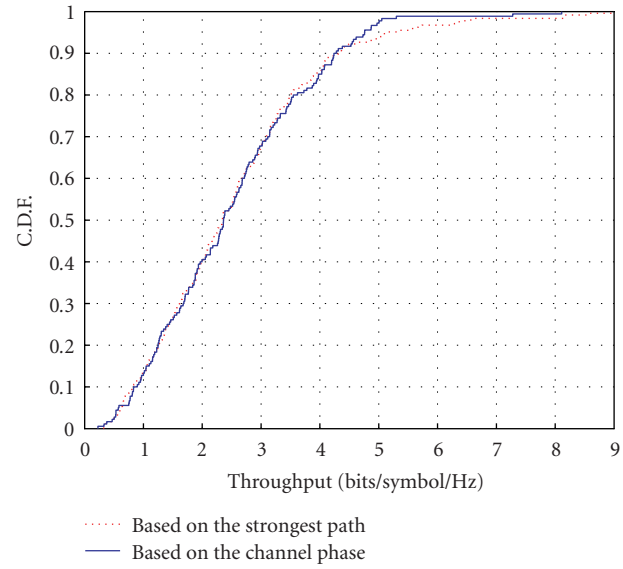


Figure 17: Normalized system throughput for the two optimized algorithms on an urban environment with a cell radius of 500 m.

service quality in large coverage case and can reduce the infrastructure cost and at the same time improve the system performance. The system performance is further evaluated in terms of system throughput to support our theoretical derivation. The corresponding normalized throughput for an urban environment has been evaluated and is illustrated in Figure 14 and that on a rural environment is shown in Figure 15. From these simulation results, it is clear that the system throughput increases when using the optimized cyclic delay algorithm on different environments. It is interesting to note that for both environments, relays with random delay do not improve the system throughput in comparison to the case without relay. The crossover in the two curves for random delay and no relay situation occurs due to the interference behaviour. At high coverage, the SINR decreases for random delay compared to the case without relay.

By implementing the two delay optimization schemes proposed in this paper, the corresponding results both in SINR and throughput are shown below. Figure 16 shows the cumulative distribution function of the received SINR for the two optimized algorithms on an urban environment, while Figure 17 shows the normalized throughput. It is observed

that the two algorithms perform almost in the same way with respect to received SINR as well as throughput. The algorithm based on strongest path performs a little better than the second algorithm. For the linear approximation algorithm, we take an approximate of the slope phase curve (which contains variations) which does not give that accurate optimum delay but gives us a rough idea of how to estimate it. On the other hand, the method based on strongest path tends to give a better approximation of the delay as it adds the paths coherently.

6. CONCLUSIONS

In this paper, two promising delay optimization schemes have been proposed based on linear approximation of the channel phase and the strongest path, for a multicellular OFDM system with cooperative relays, in order to take the most advantages of the multipath fading channel by means of exploiting the potential frequency selectivity. The obtained results show that the system performance with delay optimization increases tremendously compared with random delay diversity. Evaluations in different environments further shows that the delay of these optimization schemes is well suited for diverse environments and supports a large coverage. It should be noted that the relays work in a distributed manner and no coordination is needed; besides, both of the delay optimization schemes only require a fractional feedback to substantially improve the system performance. It is quite attractive to the operators who hope to improve the service as well as reduce the system complexity and cost.

We focused in this paper on the delay optimization with limited feedback only relying on the strongest path. One of the interesting points is to investigate how the feedback affects the system performance and what is the optimum degree of feedback with respect to the performance/cost ratio. Implementation of sector antennas will also affect the results by reducing the interference. In addition, the introduction of different scheduling algorithms, for example, always assigning the channel to the user holding the best SINR, could improve the system performance as well. These are some points that can be further explored and studied in the future.

REFERENCES

[1] F. H. P. Fitzek and M. D. Katz, *Cooperation in Wireless Networks: Principles and Applications*, Springer, New York, NY, USA, 2006.

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

[3] S. B. Slimane and A. Osseiran, "Relay communication with delay diversity for future communication systems," in *Proceedings of the 64th IEEE Vehicular Technology Conference (VTC '06)*, pp. 321–325, Montreal, Canada, September 2006.

[4] N. Ahmed, M. A. Khojastepour, and B. Aazhang, "Outage minimization and optimal power control for the fading relay channel," in *Proceedings of the IEEE Information Theory Workshop (ITW '04)*, pp. 458–462, San Antonio, Tex, USA, October 2004.

[5] J. N. Laneman, D. N. C. Tse, and G. W. Wornell, "Cooperative diversity in wireless networks: efficient protocols and outage behaviour," *IEEE Transactions on Information Theory*, vol. 50, no. 12, pp. 2062–3080, 2004.

[6] P. Larsson, "Large-scale cooperative relay network with optimal coherent combining under aggregate relay constraints," in *Proceedings of the Future Telecommunications Conference*, pp. 166–170, Beijing, China, December 2003.

[7] A. Sendonaris, E. Erkip, and B. Aazhang, "User cooperation diversity—part I: system description," *IEEE Transactions on Communications*, vol. 51, no. 11, pp. 1927–1938, 2003.

[8] A. Yaver, A. Anto, M. U. Khattak, and P. Nagarajan, "Cooperative relaying with cyclic delay diversity," Wireless Networks Course Project, 2006.

[9] J. Zander and S.-L. Kim, *Radio Resource Management for Wireless Networks*, Artech House, Norwood, Mass, USA, 2001.

[10] GPP TR 25.943 V5.1.0 Technical Report—Release 5, 3rd Generation Partnership Project, June 2002.

Hindawi Publishing Corporation
EURASIP Journal on Advances in Signal Processing
Volume 2008, Article ID 872151, 11 pages
doi:10.1155/2008/872151

Research Article

Pragmatic Space-Time Codes for Cooperative Relaying in Block Fading Channels

Andrea Conti,[1,2] Velio Tralli,[1,2] and Marco Chiani[3]

[1] *Engineering Department in Ferrara (ENDIF), University of Ferrara, Via Saragat 1, 44100 Ferrara, Italy*
[2] *Wirless Communication Laboratory (WiLAB), University of Bologna, Viale Risorgimento 2, 40136 Bologna, Italy*
[3] *Marco Chiani WiLAB/DEIS, University of Bologna, via Venezia 52, 47023 Cesena, Italy*

Correspondence should be addressed to Andrea Conti, a.conti@ieee.org

Received 13 December 2007; Accepted 23 April 2008

Recommended by Moe Win

We address the problem of construction of space-time codes for cooperative communications in block fading channels. More precisely, we consider a pragmatic approach based on the concatenation of convolutional codes and BPSK/QPSK modulation to obtain cooperative codes for relay networks, for which we derive the pairwise error probability, an asymptotic bound for frame error probability, and a design criterion to optimize both diversity and coding gain. Based on this framework, we set up a code search procedure to obtain a set of good pragmatic space-time codes (P-STCs) with overlay construction, suitable for cooperative communication with a variable number of relays in quasistatic channel, which outperform in terms of coding gain other space-time codes (STCs) proposed in the literature. We also find that, despite the fact that the implementation of pragmatic space-time codes requires standard convolutional encoders and Viterbi decoders with suitable generators and branch metric, thus having low complexity, they perform quite well in block fading channels, including quasistatic channel, even with a low number of states and relays.

1. INTRODUCTION

In wireless communications, signal fading arising from multipath is one of the main impairments and it can be mitigated through the use of diversity. A classical diversity solution is given by the adoption of multiple receiving antennas, spaced sufficiently apart from each other to obtain independent copies of the transmitted signal (see, e.g., [1–3]). In addition, also the use of multiple transmitting antennas can give similar improvements [4–6].

Cooperative communications are gaining increasing interest as a new communication paradigm involving both transmission and distributed processing which promises significant increase of capacity and diversity gain in wireless networks, by counteracting fading channels with cooperative diversity.

Several issues arise with cooperative diversity schemes such as, among others, channel modeling and implementation aspects [7, 8], protocols and resource management [9], the choice of proper relays [10], power allocation among cooperating nodes [11], and cooperative/distributed STCs [12, 13]. This work is devoted to this latter aspect.

In addition to physical antenna arrays, the relay channel model [14] enables the exploitation of distributed antennas belonging to multiple relaying terminals. This form of space diversity is referred to as cooperative diversity because terminals share antennas and other resources to create a virtual array through distributed transmission and signal processing [15, 16].

With the introduction of STCs, it has been shown how, with the use of proper trellis codes, multiple transmitting antennas can be exploited to improve system performance obtaining both diversity and coding gain, without sacrificing spectral efficiency [6, 17–21].

In particular, the design of STCs over quasistatic flat fading (i.e., fading level constant over a frame and independent frame by frame) have been addressed in [18], where some handcrafted trellis codes for two transmitting antennas have been proposed. A number of extensions of this work have eventually appeared in the literature to design good codes for

different scenarios, and STCs with improved coding gain has been presented in [22–24]. In [25–27], a pragmatic approach to STC, called P-STC, has been proposed; it simplifies the encoder and decoder structures and also allows a feasible method to search for good codes in block fading channel (BFC). P-STCs consist in the use of standard convolutional encoders and Viterbi decoders over multiple transmitting and receiving antennas, achieving maximum diversity and excellent performance, with no need of specific encoder or decoder; the Viterbi decoder requires only a simple modification in the metrics computation.

The parallel between spatial diversity and cooperative diversity encouraged researchers to investigate design criteria for STCs in relay networks, in most cases by considering only one relaying node, a quasistatic channel and limitations on the number of antennas per node. In this paper, a methodology to design P-STCs for relay networks is provided, resulting in increased flexibility with respect to the above issues. We model the channel between transmitting and receiving as BFC [28–30] that represents a simple and powerful model to include a variety of fading rates, from fast fading (i.e., ideal symbol interleaving) to quasistatic. Moreover, after the proposal of the P-STC structure for cooperative communication with various numbers of relays and transmitting antennas, we will derive the pairwise error probability, asymptotic error probability bounds, and design criteria to optimize diversity and coding gain. Finally, we will perform an efficient search for P-STCs with overlay construction over BFC to provide good (with respect to our performance bound) convolutional generators for various constraint lengths and number of relays.

The paper is organized as follows. In Section 2, we describe the system model and assumptions for the cooperative scheme. In Section 3, we describe the P-STCs approach for relay networks. Then in Section 5, we address design and search procedures for cooperative codes. The performance of P-STCs for relay networks is then given in Section 7, and our conclusions are in Section 8.

2. SYSTEM MODEL

The cooperative scheme is depicted in Figure 1 and follows time-division channel allocations with orthogonal cooperative diversity transmission [31]. Each user (i.e., the source) divides its own time-slot into two equal segments, the first from time t_1 to $t_1 + \Delta$ and the second from $t_2 = t_1 + \Delta$ to $t_2 + \Delta$, where Δ is the segment duration. In the first segment, the source broadcasts its coded symbols; in the second all the active relays which are able to decode the message forward the information through proper encoding trying to take advantage of the overall available diversity. Thus, the design of proper STCs for the two phases is crucial to maximize both achievable diversity and coding gain.

We assume n transmitting antennas at each terminal and m receiving antennas at the destination. Hence, $n_1 = n$ antennas will be used in the first phase and a total of $n_2 = Rn$ antennas will be used in the second phase, where R is the number of relays able to decode and forward the source message.

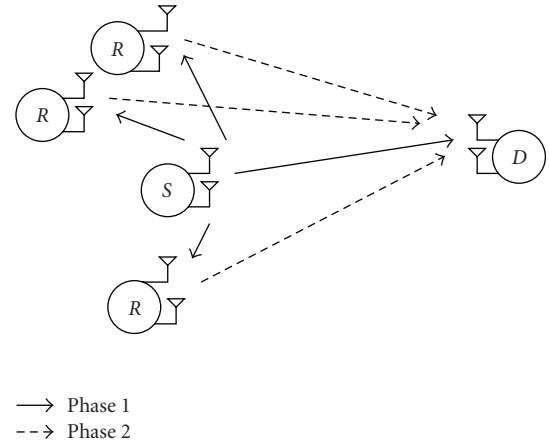


Figure 1: Two-phase relaying scheme: phase 1 (continuous line), phase 2 (dashed line). Source, relays, and destination nodes are denoted with S, R, D, respectively.

We indicate with $c_{r,i}^{(t)}$ the modulation symbol transmitted by relay r ($0 \le r \le R$, and $r = 0$ is the source) on the antenna i at discrete time t, that is, at the tth instant of the encoder clock. With superscripts H, T, and $*$ we denote conjugation and transposition, transposition only, and conjugation only, respectively. Each symbol is assumed to have unit norm and to be generated according to the modulation format by suitable mapping. Note that symbol $c_{0,i}^{(t)}$ is transmitted at time $t_1 + t$, while symbols $c_{r,i}^{(t)}$ for $r > 0$ are transmitted at time $t_2 + t$. The received signals corresponding to all symbols $c_{r,i}^{(t)}$ are jointly processed by the decoder at the reference time t. We also denote with $\mathbf{C}^{(t)} = [c_{0,1}^{(t)}, c_{0,2}^{(t)}, \ldots, c_{R,n}^{(t)}]^T$ a supersymbol, which is the vector of the $(R+1)n$ outputs of the overall "virtual encoder" constituted by the source encoder and the relays' encoders.

A codeword is a sequence $\underline{c} = (\mathbf{C}^{(1)}, \ldots, \mathbf{C}^{(N)})$ of N supersymbols generated by the source and relays' encoders. This codeword $\underline{c}$ is interleaved before transmission to obtain the sequence $\underline{c}_I = I(\underline{c}) = (\mathbf{C}^{(\sigma_1)}, \ldots, \mathbf{C}^{(\sigma_N)})$, where $\sigma_1, \ldots, \sigma_N$ is a permutation of the integers $1, \ldots, N$, and $I(\cdot)$ is the interleaving function. Note that with this notation the permutation is the same for all the transmitting terminals in the two phases.

The channel model includes additive white Gaussian noise (AWGN) and multiplicative flat fading, with Rayleigh distributed amplitudes assumed constant over blocks of B consecutive transmitted space-time symbols and independent from block to block [28–30]. Perfect channel state information is assumed at the decoder for each node.

The transmitted supersymbol at time σ_t goes through a compound channel described by the $((n_1 + n_2) \times m)$ channel matrix $\mathbf{H}^{(\sigma_t)} = [H_0^{(\sigma_t)}, \ldots, H_R^{(\sigma_t)}]^T$, where $H_r^{(\sigma_t)} = \{h_{r,i,s}^{(\sigma_t)}\}$, and $h_{r,i,s}^{(\sigma_t)}$ is the channel gain between transmitting antenna i, with $i = 1, \ldots, n$, of the terminal r and receiving antenna s, with $s = 1, \ldots, m$, at time σ_t.

In the BFC model, these channel matrices do not change for B consecutive transmissions, thus we actually have only

$L = N/B$ possible distinct channel matrix instances per codeword (for the sake of simplicity, we assume that N and B are such that L is an integer). Denoting by $\mathcal{Z} = \{\mathbf{Z}_1, \ldots, \mathbf{Z}_L\}$ the set of L channel instances, we have

$$\mathbf{H}^{(\sigma_t)} = \mathbf{Z}_l \quad \text{for } \sigma_t = (l-1)B + 1, \ldots, lB,\ l = 1, \ldots, L. \tag{1}$$

When the fading block length, B, is equal to one, we have the ideally interleaved fading channel (i.e., independent fading levels from symbol to symbol), while for $L = 1$, we have the quasistatic fading channel (fading level constant over a codeword); by varying L, we can describe channels with different correlation degrees [28–30].

At the receiving side, the sequence of received signal vectors is $\underline{r}_I = (\mathbf{R}^{(\sigma_1)}, \ldots, \mathbf{R}^{(\sigma_N)})$, and after deinterleaving we have $\underline{r} = I^{-1}(\underline{r}_I) = (\mathbf{R}^{(1)}, \ldots, \mathbf{R}^{(N)})$, where the received vector at time t is $\mathbf{R}^{(t)} = [r_1^{(t,1)}, r_1^{(t,2)}, \ldots, r_m^{(t,2)}]^T$ with components

$$r_s^{(t,1)} = \sqrt{E_s}\sum_{i=1}^{n} h_{0,i,s}^{(\sigma_t)} c_{0,i}^{(t)} + \eta_s^{(t,1)}, \qquad s = 1, \ldots, m, \tag{2}$$

in the first phase and

$$r_s^{(t,2)} = \sqrt{E_s}\sum_{r=1}^{R}\sum_{i=1}^{n} h_{r,i,s}^{(\sigma_t)} c_{r,i}^{(t)} + \eta_s^{(t,2)}, \qquad s = 1, \ldots, m, \tag{3}$$

in the second phase. In this equation, $r_s^{(t,l)}$ is the signal-space representation of the signal received by antenna s at time t in phase l, the noise terms $\eta_s^{(t,l)}$ are independent, identically distributed (i.i.d.) complex Gaussian random variables (r.v.s), with zero mean and variance $N_0/2$ per dimension, and the r.v.s $h_{r,i,s}^{(\sigma_t)}$ represent the deinterleaved complex Gaussian fading coefficients. Since we assume spatially uncorrelated channels, these are i.i.d. with zero mean and variance 1/2 per dimension, and consequently $|h_{r,i,s}^{(\sigma_t)}|$ are Rayleigh distributed r.v.s with unit power. The constellations are multiplied by a factor $\sqrt{E_s}$ in order to have a transmitted energy per symbol equal to E_s, which is also the average received symbol energy (per transmitting antenna) due to the normalization adopted on the fading gains. This is motivated by the use of a power control technique which keeps constant the average received symbol energy.

The total energy transmitted per supersymbol is $E_{s_T} = (n_1 + n_2)E_s$ and the energy transmitted per information bit is $E_b = E_s/(hR_c)$, where h is the number of bits per modulation symbol and R_c is the overall code rate of the cooperative space-time code. Thus with ideal pulse shaping, the spectral efficiency is $nhR_c/2$ [bps/Hz].

For following discussions sections, it is worthwhile to recall that, over a Rayleigh fading channel, the system achieves a diversity $\mathcal{D}$ if the asymptotic error probability is $P_e \approx K(E_s/N_0)^{-\mathcal{D}}$, where K is a constant depending on the asymptotic coding gain [1, 32]. In other words, a system with diversity $\mathcal{D}$ is described by a curve of error probability with a slope approaching $10/\mathcal{D}$ [dB/decade] for large signal-to-noise ratio (SNR).

3. PRAGMATIC SPACE-TIME CODES FOR COOPERATIVE RELAYING

In the case of the two-phase relaying scheme shown in Figure 1, the probability of transmission failure over the two phases depends on the number of relays available for cooperation and on the link qualities on source-destination, source-relays, and relays-destination. We envisage two main applications.

(a) *With static set of relays*. The set of relays is initialized at the beginning of a data communication session and is kept unchanged over a long period of many slots. The set of relays is chosen by looking at active terminals able to guarantee a good average link quality (depending on terminal position and slow fading) with the source terminal. During this period, a cooperative coding scheme is used by the source and the set of relays to protect the transmission of data frames between source and destination. Sometimes, due to fast fading fluctuations, it may happen that one or more relays are not able to decode the source codewords in phase 1. In the simplified case of equal quality on all source-relay links, denoting by $P_e^{(S-D)}$ the error probability for the link from source to destination, $P_e^{(S-R)}$ the error probability for the source-relay link (i.e., $P_e^{(S-R_1)} = \cdots = P_e^{(S-R_r)} = P_e^{(S-R)}$), and with $P_e^{(SR_1\cdots R_k-D)}$ the error probability for the link from the source plus k relays to destination, the overall error probability P_e is given by

$$\begin{aligned} P_e &= P_e^{(S-R)}P_e^{(S-D)} + \left(1 - P_e^{(S-R)}\right)P_e^{(SR_1-D)}, \\ P_e &= \left(P_e^{(S-R)}\right)^2 P_e^{(S-D)} + 2P_e^{(S-R)}\left(1 - P_e^{(S-R)}\right)P_e^{(SR_1-D)} \\ &\quad + \left(1 - P_e^{(S-R)}\right)^2 P_e^{(SR_1R_2-D)}, \end{aligned} \tag{4}$$

for the cases of 1 and 2 potential relays, respectively. These can be generalized to the case of R potential relays as

$$\begin{aligned} P_e &= \left(P_e^{(S-R)}\right)^R P_e^{(S-D)} \\ &\quad + \sum_{k=1}^{R}\binom{R}{k}\left(P_e^{(S-R)}\right)^{R-k}\left(1 - P_e^{(SR-D)}\right)^k P_e^{(SR_1\cdots R_k-D)}. \end{aligned} \tag{5}$$

(b) *With dynamic set of relays*. The set of cooperating relays is determined frame by frame as those relays which are active as well as able to hear and decode the source in phase 1. By means of a suitable signaling, they agree on the cooperative coding scheme and complete transmission in phase 2. In this case, the error probability P_e is given by

$$P_e = \sum_{k=0}^{R}\mathbb{P}\{k \text{ cooperating relays}\}P_e^{(SR_1\cdots R_k-D)}, \tag{6}$$

where $\mathbb{P}\{\cdot\}$ indicates probability, R is the total number of relays, and $\mathbb{P}\{k \text{ cooperating relays}\}$ depends on the spatial distribution of the nodes in the network as well as fast and slow fading statistics.

For the design of the coding scheme with cooperative relays, it is generally recognized that the code components

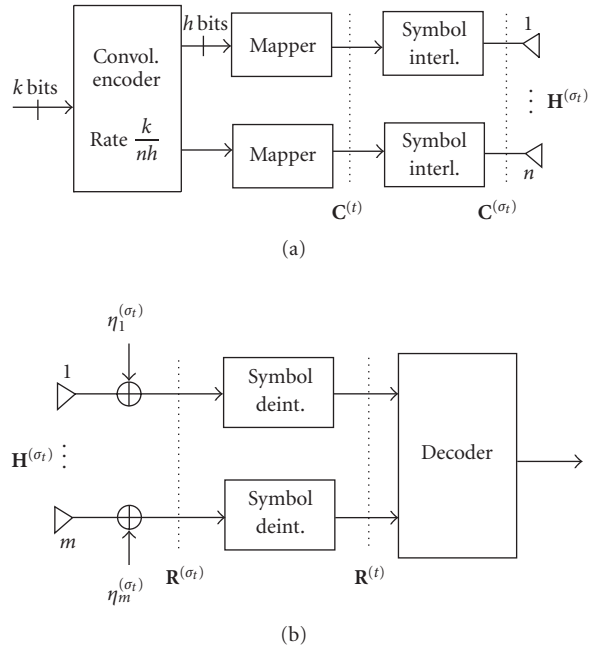


Figure 2: Architecture of pragmatic space-time codes. For cooperative P-STCs, the "distributed" convolutional encoder is the ensamble of $R+1$ single encoders, one for each transmitter; hence, instead of n, we must consider the overall number of antennas $(R+1)n$.

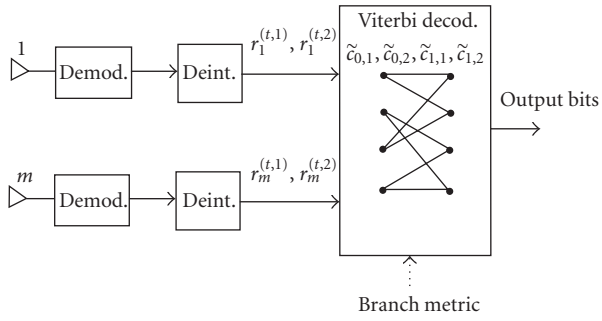


Figure 3: Receiver structure for the proposed P-STCs in cooperative communications. The Viterbi decoder is the same as for the single convolutional code adopted in transmission, but the metric on the generic branch is, for example, for $n=2$ and $R=1$, $\sum_{s=1}^{m}|r_s^{(t,1)} - \sqrt{E_s}(h_{0,1,s}^{(t)}\widetilde{c}_{0,1} + h_{0,2,s}^{(t)}\widetilde{c}_{0,2})|^2 + |r_s^{(t,2)} - \sqrt{E_s}(h_{1,1,s}^{(t)}\widetilde{c}_{1,1} + h_{1,2,s}^{(t)}\widetilde{c}_{1,2})|^2$, where $\widetilde{c}_{0,1}, \widetilde{c}_{0,2}, \widetilde{c}_{1,1}, \widetilde{c}_{1,2}$, are the four symbols associated to the branch. Note that $r_s^{(t,1)}$ is received at time $t_1 + t$ whereas $r_s^{(t,2)}$ is received at time $t_2 + t$.

used by the source in phase 1 should maximize diversity and coding gain for each link connecting the source to relays and destination. The other code components should be designed to maximize diversity and coding gain of the entire cooperative code, that is, the code including all the code components transmitted during phases 1 and 2, for any possible number of cooperative relays [12].

In this paper, we are considering the design of space-time trellis codes for relaying networks by using the pragmatic approach of [25, 27]. Our proposed "pragmatic" approach uses a low-complexity architecture for STCs where the code components are built by the concatenation of a binary convolutional encoder and binary phase shift keying (BPSK) or quaternary phase shift keying (QPSK) modulator. This code architecture was also referred to as algebraic STCs in [33]. Our "pragmatic" approach thus consists in using common convolutional codes as space-time codes, with the architecture presented in Figure 2. Here, k information bits are encoded by a convolutional encoder with rate $k/(nh)$. The nh output bits are divided into n streams, one for each transmitting antenna, of BPSK ($h=1$) or QPSK ($h=2$) symbols that are obtained from a natural (Gray) mapping of h bits. By natural mapping; we mean that for BPSK an information bit $b \in \{0,1\}$ is mapped into the antipodal symbol $c = 2b-1$, giving $c \in \{-1,+1\}$; for QPSK a pair of information bits, a, b, is mapped into a complex symbol $c = (2a-1)/\sqrt{2} + j(2b-1)/\sqrt{2}$, giving $c \in \{\pm 1/\sqrt{2} \pm j/\sqrt{2}\}$, with $j = \sqrt{-1}$. Then, each stream of symbols is eventually interleaved (we focus our attention on symbol interleaving: bit interleaving is addressed in [34]). If μ is the encoder constraint length, then the associated trellis has $N_s = 2^{k(\mu-1)}$ states.

We can describe P-STCs for cooperative communication, obtained by joining the $R+1$ code components used by the cooperating transmitters, by using the trellis of each encoder (the same as for the convolutional codes (CC)), labelling the generic branch from state S_i to state S_j with the supersymbol $\widetilde{\mathbf{C}}_{S_i \to S_j} = [\widetilde{c}_{0,1}, \ldots, \widetilde{c}_{R,n}]^T$, where for BPSK, the symbol $\widetilde{c}_{r,i}$ is the output (in antipodal form) of the ith generator of the rth transmitter.

One of the advantages of the pragmatic architecture is that the maximum likelihood (ML) decoder is the same Viterbi decoder of the convolutional encoder adopted in transmission (same trellis), with a simple modification of the branch metrics. Being $\{\widetilde{c}_{r,i}\}$ the set of output symbols labelling the branch, the branch metric for the Viterbi decoder is thus given by

$$\sum_{s=1}^{m}\left|r_s^{(t,1)} - \sqrt{E_s}\sum_{i=1}^{n}h_{0,i,s}^{(t)}\widetilde{c}_{0,i}\right|^2 + \left|r_s^{(t,2)} - \sqrt{E_s}\sum_{r=1}^{R}\sum_{i=1}^{n}h_{r,i,s}^{(t)}\widetilde{c}_{r,i}\right|^2. \tag{7}$$

For example, in Figure 3 we show the receiver architecture for the cooperative P-STCs, that simply consists in the usual Viterbi decoder for the convolutional code adopted in transmission, with the only change of the metric on a generic trellis branch, as illustrated in the caption. Thus, the advantages of P-STCs with respect to STCs are as follows:

(i) the encoder is a common convolutional encoder;

(ii) the (Viterbi) decoder is the same as for a convolutional code, except for a change in the metric evaluation;

(iii) P-STCs are easy to study and optimize, even over BFC.

These advantages apply also when P-STCs are used for cooperative communications, as it will be further investigated in the next sections.

4. PERFORMANCE ANALYSIS FOR COOPERATIVE SPACE-TIME CODES OVER BFC

We first consider the derivation of the pairwise error probability (PEP). Given the transmitted codeword $\underline{c}$ and another codeword $\underline{g} \neq \underline{c}$, the PEP, that is, the probability that the ML decoder favors the codeword $\underline{g}$ over $\underline{c}$, conditional to the set of fading levels $\mathcal{Z}$, can be written as

$$\mathbb{P}\{\underline{c} \longrightarrow \underline{g} \mid \mathcal{Z}\} = \frac{1}{2}\operatorname{erfc}\sqrt{\frac{E_s}{4N_0}d^2(\underline{c},\underline{g} \mid \mathcal{Z})}, \tag{8}$$

where $\operatorname{erfc}(x) \triangleq (2/\sqrt{\pi})\int_x^\infty e^{-t^2}dt$ is the complementary error function, and the conditional Euclidean squared distance at the channel output, $d^2(\underline{c},\underline{g} \mid \mathcal{Z})$, is given by [18]

$$d^2(\underline{c},\underline{g} \mid \mathcal{Z}) = \sum_{t=1}^{N}\sum_{s=1}^{m}\left[\left|\sum_{i=1}^{n} h_{0,i,s}^{(\sigma_t)}\cdot\left(c_{0,i}^{(t)} - g_{0,i}^{(t)}\right)\right|^2 + \left|\sum_{r=1}^{R}\sum_{i=1}^{n} h_{r,i,s}^{(\sigma_t)}\cdot\left(c_{r,i}^{(t)} - g_{r,i}^{(t)}\right)\right|^2\right]. \tag{9}$$

To specialize this expression for BFC, we first rewrite the squared distance as follows:

$$d^2(\underline{c},\underline{g} \mid \mathcal{Z}) = \sum_{t=1}^{N}\sum_{s=1}^{m}\mathbf{h}_s^{(\sigma_t)}\mathbf{A}^{(t)}(\underline{c},\underline{g})\mathbf{h}_s^{(\sigma_t)H}, \tag{10}$$

where $\mathbf{h}_s^{(\sigma_t)} = [h_{0,1,s}^{(\sigma_t)}, h_{0,2,s}^{(\sigma_t)}, \ldots, h_{R,n,s}^{(t\sigma_t)}]$ is the $(1 \times (R+1)n)$ vector of fading coefficients related to the receiving antenna s. In (10), the $((n_1+n_2)\times(n_1+n_2))$ matrix $\mathbf{A}^{(t)}(\underline{c},\underline{g})$ is Hermitian nonnegative definite [27] and has the following block structure:

$$\mathbf{A}^{(t)}(\underline{c},\underline{g}) = \begin{bmatrix}\mathbf{a}^{(t,1)} & \mathbf{0}\\ \mathbf{0} & \mathbf{a}^{(t,2)}\end{bmatrix}, \tag{11}$$

where

$$\begin{aligned}\mathbf{a}^{(t,1)} &= \left(\mathbf{c}^{(t,1)} - \mathbf{g}^{(t,1)}\right)\left(\mathbf{c}^{(t,1)} - \mathbf{g}^{(t,1)}\right)^H,\\ \mathbf{a}^{(t,2)} &= \left(\mathbf{c}^{(t,2)} - \mathbf{g}^{(t,2)}\right)\left(\mathbf{c}^{(t,2)} - \mathbf{g}^{(t,2)}\right)^H,\end{aligned} \tag{12}$$

after having split the generic supersymbol $\mathbf{C}^{(t)}$ in the two parts transmitted during phase 1 and phase 2, respectively, that is, $\mathbf{C}^{(t)} = [\mathbf{c}^{(t,1)}\ \mathbf{c}^{(t,2)}]^T$, where $\mathbf{c}^{(t,1)} = [c_{0,1}, \ldots, c_{0,n}]$ and $\mathbf{c}^{(t,2)} = [c_{1,1}, \ldots, c_{R,n}]$.

Due to the BFC assumption, for each frame and each receiving antenna, the fading channel is described by only L different vectors $\mathbf{h}_s^{(t)} \in \{\mathbf{z}_s^{(1)}, \mathbf{z}_s^{(2)}, \ldots, \mathbf{z}_s^{(L)}\}$, $s = 1, \ldots, m$, where $\mathbf{z}_s^{(l)}$ is the sth row of $\mathbf{Z}_l$. By grouping these vectors, we can rewrite (10) as

$$d^2(\underline{c},\underline{g} \mid \mathcal{Z}) = \sum_{l=1}^{L}\sum_{s=1}^{m}\mathbf{z}_s^{(l)}\mathbf{F}^{(l)}(\underline{c},\underline{g})\mathbf{z}_s^{(l)H}, \tag{13}$$

where

$$\mathbf{F}^{(l)}(\underline{c},\underline{g}) \triangleq \sum_{t\in T(l)}\mathbf{A}^{(t)}(\underline{c},\underline{g}), \quad l = 1, \ldots, L, \tag{14}$$

and $T(l) \triangleq \{t : \mathbf{H}^{(\sigma_t)} = \mathbf{Z}_l\}$ is the set of indexes t where the channel fading gain matrix is equal to $\mathbf{Z}_l$. This set depends on the interleaving strategy adopted. Note that in our scheme (Figure 2), the interleaving is done "horizontally" for each transmitting antenna and in the same way for each transmitter, and that the set $T(l)$ is independent of s, in other words, that the interleaving rule is the same for all antennas.

The matrix $\mathbf{F}^{(l)}(\underline{c},\underline{g})$ is also Hermitian nonnegative definite, being the sum of Hermitian nonnegative definite matrices. It has, therefore, real nonnegative eigenvalues. Moreover, it can be written as $\mathbf{F}^{(l)}(\underline{c},\underline{g}) = \mathbf{U}^{(l)}\mathbf{\Lambda}^{(l)}\mathbf{U}^{(l)H}$, where $\mathbf{U}^{(l)}$ is a unitary matrix and $\mathbf{\Lambda}^{(l)}$ is a real diagonal matrix, whose diagonal elements $\lambda_i^{(l)}$ with $i = 1, \ldots, \tilde{n} = n_1 + n_2$ are the eigenvalues of $\mathbf{F}^{(l)}(\underline{c},\underline{g})$ counting multiplicity. Note that $\mathbf{F}^{(l)}$ and its eigenvalues $\lambda_i^{(l)}$ are functions of $\underline{c} - \underline{g}$. As a result, we can express the squared distance $d^2(\underline{c},\underline{g} \mid \mathcal{Z})$ in terms of the eigenvalues of $\mathbf{F}^{(l)}(\underline{c},\underline{g})$ as follows:

$$\begin{aligned}d^2(\underline{c},\underline{g} \mid \mathcal{Z}) &= \sum_{l=1}^{L}\sum_{s=1}^{m}\mathbf{z}_s^{(l)}\mathbf{U}^{(l)}\mathbf{\Lambda}^{(l)}\mathbf{U}^{(l)H}\mathbf{z}_s^{(l)H}\\ &= \sum_{l=1}^{L}\sum_{s=1}^{m}\mathbf{B}_s^{(l)}\mathbf{\Lambda}^{(l)}\mathbf{B}_s^{(l)H}\\ &= \sum_{l=1}^{L}\sum_{s=1}^{m}\sum_{i=1}^{\tilde{n}}\lambda_i^{(l)}\left|\beta_{i,s}^{(l)}\right|^2,\end{aligned} \tag{15}$$

where $\mathbf{B}_s^{(l)} = [\beta_{1,s}^{(l)}, \beta_{2,s}^{(l)}, \ldots, \beta_{\tilde{n},s}^{(l)}] = \mathbf{z}_s^{(l)}\mathbf{U}^{(l)}$.

It should be observed that the form of matrix $\mathbf{F}^{(l)}(\underline{c},\underline{g})$ is different from the matrix of the same space-time code working on a system with $n_1 + n_2$ transmit antennas defined in [27] due to the use of two distinct transmission phases in the cooperative system. Therefore, the same code used in the cooperative system may achieve different diversity and coding gains. It should also be noted that this matrix is diagonal (hence full-rank) only when $n = 1$ and $R = 1$. When transmitters have more than one antenna or more than one relay cooperate to transmission, only a suitable choice of the code may lead to a full-rank matrix, as shown later.

Vector $\mathbf{z}_s^{(l)}$ has independent, complex Gaussian elements, with zero mean and variance 1/2 per dimension. Since $\mathbf{U}^{(l)}$ represents a unitary transformation, $\mathbf{B}_s^{(l)}$ has the same statistical description of $\mathbf{z}_s^{(l)}$. Moreover, for BFC, vectors $\mathbf{B}_s^{(l)}$ and $\mathbf{B}_s^{(j)}$ are independent for all $l \neq j$. Hence, the unconditional pairwise error probability (PEP) becomes

$$\mathbb{P}\{\underline{c} \longrightarrow \underline{g}\} = \mathbb{E}\left\{\frac{1}{2}\operatorname{erfc}\sqrt{\frac{E_s}{4N_0}\sum_{s=1}^{m}\sum_{l=1}^{L}\sum_{i=1}^{\tilde{n}}\lambda_i^{(l)}\left|\beta_{i,s}^{(l)}\right|^2}\right\}, \tag{16}$$

where $\mathbb{E}\{\cdot\}$ indicates expectation with respect to fading (i.e., over the distribution of the $\beta_{i,s}^{(l)}$). By evaluating the asymptotic behavior for large SNR, we obtain (see [35])

$$P(\underline{c} \longrightarrow \underline{g}) \leq K(m\eta)\left[\prod_{l=1}^{L}\prod_{i=1}^{\eta_l}\lambda_i^{(l)}\left(\frac{E_s}{4N_0}\right)^{\eta}\right]^{-m}, \tag{17}$$

where

$$K(d) = \frac{1}{2^{2d}}\binom{2d-1}{d}. \tag{18}$$

A looser bound can be obtained by observing that $K(d) \leq 1/4$. The integer $\eta_l = \eta_l(\underline{c},\underline{g})$ is the number of nonzero eigenvalues of $\mathbf{F}^{(l)}(\underline{c},\underline{g})$, and η (which we call the *pairwise transmit diversity*) is the sum of the ranks of $\mathbf{F}^{(l)}(\underline{c},\underline{g})$, that is,

$$\eta = \eta(\underline{c},\underline{g}) = \sum_{l=1}^{L}\operatorname{rank}[\mathbf{F}^{(l)}(\underline{c},\underline{g})] = \sum_{l=1}^{L}\eta_l. \tag{19}$$

The PEP between $\underline{c}$ and $\underline{g}$ shows a diversity $m\eta$, that is, the product of transmit and receive diversity.

Therefore, it is clear that the performance analysis of a cooperative space-time codes with fixed number of cooperating relays is similar to the analysis of common space-time codes as in [18, 27]. The only difference lies in the structure of the matrix $\mathbf{F}^{(l)}(\underline{c},\underline{g})$ which has some zero off-diagonal elements and may therefore have different rank and eigenvalues. Bearing in mind this fact, it is easy to derive an error probability bound as

$$\widetilde{P}_w = \sum_{\underline{c}}\mathbb{P}\{\underline{c}\}P_w(\underline{c}) \leq \sum_{\underline{c}}\sum_{\underline{g}\neq\underline{c}}\mathbb{P}\{\underline{c}\}\mathbb{P}\{\underline{c}\longrightarrow\underline{g}\}, \tag{20}$$

where $\mathbb{P}\{\underline{c}\}$ is the probability of transmitting the codeword $\underline{c}$ (i.e., for P-STCs, equal to 2^{-kN} for equiprobable input bit sequence and $2^{-k(N-\mu+1)}$ for a zero-tailed code) and $\mathrm{P}_w(\underline{c})$ is the codeword error probability for the transmitted codeword $\underline{c}$. By using the asymptotic approximation (17), and by observing that the retained dominant terms are those with transmit diversity $\widetilde{\eta}_{\min} = \min_{\underline{c}}\eta_{\min}(\underline{c})$, where $\eta_{\min}(\underline{c}) = \min_{\underline{g}}\eta(\underline{c},\underline{g})$ and $\mathcal{E}(\underline{c},x) = \{\underline{g}\neq\underline{c} : \eta(\underline{c},\underline{g}) = x\}$ is the set of codeword sequences at minimum diversity, the asymptotic error probability bound can be written as

$$\widetilde{P}_{w_\infty} \approx K(\widetilde{\eta}_{\min}m)\left(\frac{E_s}{4N_0}\right)^{-\widetilde{\eta}_{\min}\cdot m}\sum_{\underline{c}}\mathbb{P}\{\underline{c}\}\sum_{\underline{g}\in\mathcal{E}(\underline{c},\widetilde{\eta}_{\min})}\left[\prod_{l=1}^{L}\prod_{i=1}^{\eta_l}\lambda_i^{(l)}\right]^{-m}. \tag{21}$$

From (21), we observe that the asymptotic performance of STCs in BFC depends on both the achievable diversity, $\widetilde{\eta}_{\min}\cdot m$, and the performance factor

$$\begin{aligned}\widetilde{F}_{\min}(m) &= \sum_{\underline{c}}\mathbb{P}\{\underline{c}\}F_{\min}(\underline{c},m)\\ &\triangleq \sum_{\underline{c}}\mathbb{P}\{\underline{c}\}\sum_{\underline{g}\in\mathcal{E}(\underline{c},\widetilde{\eta}_{\min})}\left[\prod_{l=1}^{L}\prod_{i=1}^{\eta_l}\lambda_i^{(l)}\right]^{-m},\end{aligned} \tag{22}$$

which is related to the coding gain in (21).

Note also that $\widetilde{\eta}_{\min}$ and the weights $\prod_{l=1}^{L}\prod_{i=1}^{\eta_l}\lambda_i^{(l)}$ for each $\underline{c}$ and $\underline{g}$ do not depend on the number of receiving antennas. Therefore, when a code is found to reach the maximum diversity $\widetilde{\eta}_{\min}$ in a system with one receiving antenna, the same code reaches the maximum diversity $\widetilde{\eta}_{\min}\cdot m$ when used with multiple receiving antennas. However, due to the presence of the exponent m in each term of the sum in (22), the best code (i.e., the code having the smallest performance factor) for a given number of antennas is not necessarily the best for a different number of receiving antennas. Thus, a search for optimum codes in terms of both diversity and performance factor must in principle be pursued for each m.

To summarize, the derivation of the asymptotic behavior of a given STC with a given length requires the computation of the matrices $\mathbf{F}^{(l)}(\underline{c},\underline{g})$ in (14) with their rank and product of nonzero eigenvalues. Moreover, according to [36], by restricting in the bound the set of sequences $\underline{g}$ to those corresponding to paths in the trellis diagram of the code diverging only once from the path of codeword $\underline{c}$, the union bound becomes tighter and can be evaluated in an effective way, by using the methodology illustrated in [27] through the concept of space-time generalized transfer function.

5. PRAGMATIC SPACE-TIME CODE DESIGN FOR RELAYING

In this section, we address the issues of how to set up design criteria for good cooperative STCs and how to perform an efficient search for the optimum (in a sense defined later) generators for the code components of cooperative STCs in BFC.

In general the design of good cooperative STCs may be based on one of the following approaches.

(a) By assuming that the cooperative code is working with a predefined number of cooperative relays R, it may be designed as P-STC with k binary inputs and $n(R+1)$ output symbols which maximize diversity gain and coding gain. A pragmatic suboptimal solution to this problem may be to build the code using the rate $k/(nh(R+1))$ maximum free distance convolutional code, optimum for the AWGN channel. This design method does not guarantee that the first rate k/nh component code used in phase 1 is the best performing code. It also does not guarantee good performance when some code components are not used by relays unable in some frames to decode the source message. Moreover, the pragmatic solution may be not optimal even in terms of diversity gain and therefore should be checked by means of simulations. However, we observed that in many cases this solution leads to quite good results.

(b) By assuming that the cooperative code is obtained by joining code components in phase 2 from every relay able to decode the source message, the code may be designed as STC with overlay construction [37]. With this method, a good code for R relays is designed starting from a good code for $R-1$ relays and by adding the best code component that maximizes diversity and coding gain of the final code. In this way, the first code component used by the source in phase 1 is always a good code. In the case of a fixed set of cooperating relays, the sequence of additional code components can be

assigned to the relays ranked in order of average link quality in such a way that the second code component is assigned to the relay with the best link quality and so on, thus they are used with high probability in the same combinations for which they have been designed. Moreover, it is easier to design the additional code components than the entire cooperative code.

The design of STCs with overlay construction was addressed in [37], but not in the special case of cooperative codes. Algebraic design criteria were derived for maximizing diversity gain, without addressing coding gain issues. The work in [12] proposed to use this STC with overlay construction as a cooperative STC but without specializing the design for the cooperative scenario. In this section, we propose to set up an STC code search that aims at seeking cooperative codes covering both the outlined design approaches, that is, the design of an entire rate $k/(nh(R+1))$ P-STCs, and the design of rate k/nh code components in an overlay structure.

The search criterion proposed here is based on the asymptotic error probability in (21), so that the optimum code with fixed parameters (n, k, h, μ), among the set of non-catastrophic codes, is the code that

(i) maximizes the achieved diversity, $\widetilde{\eta}_{\min}$;

(ii) minimizes the performance factor $\widetilde{F}_{\min}(m)$;

where the values of $\widetilde{\eta}_{\min}$ and $\widetilde{F}_{\min}(m)$ can be extracted from the space-time generalized transfer function (ST-GTF) of the code [27]. Therefore, an exhaustive search algorithm should evaluate the ST-GTF for each code of the set.

Another search criterion for STC has been addressed in [22, 24] where a method based on the evaluation of the worst PEP was proposed. Although the worst PEP carries information about the achievable diversity, $\widetilde{\eta}_{\min}$, it is incomplete with respect to coding gain, thus producing a lower bound for the error probability. Even though our method based on the union bound is still approximate with respect to coding gain (giving an upper bound), it includes more information than the other method, leading often to the choice of codes with better performance.

When applying our search criterion, we must consider that, as shown in [38], the union bound for the average error probability is loose and in some cases (long codes and small diversity) is very far from the actual value. This problem can be partially overcome by truncating the sum to the most significant terms, but this technique leads to an approximation. However, this approach gives good results in reproducing the correct performance ranking of the codes among those achieving the same diversity $\widetilde{\eta}_{\min}$, as will be checked in the numerical results section.

Of course, the achievable diversity is the most important design parameter. Since $\widetilde{\eta}_{\min}$ cannot be larger than $\eta(\underline{c}, \underline{g}) \leq (R+1)nL$ and the free distance d_f of the convolutional code used to build the P-STCs, it appears that to capture the maximum diversity per receiving antenna offered by the channel, $(R+1)nL$, the free distance of a good code for a given BFC should be at least $(R+1)nL$ or larger. On the other hand, there is a fundamental limit on the achievable diversity related to the Singleton bound for BFC [30].

Table 1: Optimum overlays for rate $1/(2R)$ COP-STC with BPSK, $n = 2$, $m = 1$, $R = 1$, in quasistatic channel. The basic code for a single transmitter is STC as in [27].

μ	Generators $r = 0$	Generators $r = 1$	$\widetilde{F}_{\min}(1)/N$
2	$(1,2)_8$	$(1,2)_8$	0.0044
3	$(3,4)_8$	$(5,7)_8$	0.0015
4	$(13,15)_8$	$(11,17)_8$	0.0008
5	$(23,31)_8$	$(27,35)_8$	0.0006

Let us define the reference block fading channel (RBFC) for the system as the ideal equivalent BFC with $(R+1)nL$ fading blocks that would describe the space-time fading channel if the $(R+1)n$ transmitters determine $(R+1)n$ independent channels. The achievable diversity, which cannot be larger than the diversity achievable on the reference BFC, is bounded by

$$\widetilde{\eta}_{\min} \leq 1 + \left\lfloor L(R+1)n\left(1 - \frac{k}{(R+1)nh}\right)\right\rfloor. \tag{23}$$

As an example, to achieve full diversity $(R+1)n$ with P-STCs in a quasistatic channel ($L = 1$), the value k/h cannot be larger than 1, thus the code rate of each convolutional code component cannot be larger than $1/n$, or the value of h cannot be smaller than k (see also [18]).

6. CODE SEARCH RESULTS

In this section, we report the results obtained in our search for good cooperative STCs with overlay construction for different system configurations with $R = 1, 2, 3$ relays, $n = 1, 2$ transmitting antennas, $m = 1$ receiving antennas. Note that according to the analytical framework in Section 4 the diversity gain of the proposed codes increases as $\mathcal{D}m$ for $m > 1$. All the codes proposed are full-diversity codes. Two approaches are considered for overlay construction: the first considers the use of the maximum free distance (optimum for AWGN channel) rate k/hn code as the first code component; the second as the first code component considers the best rate k/hn P-STCs reported in [27]. When possible, these codes are compared with the cooperative STCs proposed in [12].

Suboptimal cooperative STCs, working with a predefined number of cooperative relays R and constructed by pragmatically choosing the best (maximum distance) convolutional codes for AWGN, can be easily obtained by using the convolutional code generators reported in [2, Section 8.2]. It has been checked that this easy approach leads in most cases to acceptable results. However, it sometimes leads to cooperative STCs not achieving full diversity. As an example, this is the case of the rate 1/4, 4-state code with generators $(5,7,7,7)_8$ for systems with BPSK, $n = 2$ and $R = 1$ cooperative relays, which achieves a maximum diversity of 3. Note that, according to the Singleton bound, full-diversity rate $k/(h(R+1)n)$ codes can be constructed if $k \leq h$.

The results are collected in Tables 1–5. It is interesting to note that the proposed codes are able to capture the maximum available diversity with only 2–4 states in the

Table 2: Optimum overlays for rate $1/(2R)$ COP-STC with BPSK, $n = 2$, $m = 1$, $R = 1$, in quasistatic channel. The basic code for a single transmitter is STC with the best convolutional code for the AWGN channel. Superscript (1) refers to C-STC as in [12].

μ	Generators $r = 0$	Generators $r = 1$	$\widetilde{F}_{\min}(1)/N$	Generators$^{(1)}$ $r = 1$	$\widetilde{F}_{\min}(1)^{(1)}/N$
2	$(1,3)_8$	$(1,2)_8$	0.0044	—	—
3	$(5,7)_8$	$(5,6)_8$	0.00135	$(5,7)_8$	0.00155
4	$(15,17)_8$	$(13,16)_8$	0.00095	$(13,15)_8$	0.00121
5	$(23,35)_8$	$(27,31)_8$	0.00061	$(25,37)_8$	0.00072

Table 3: Optimum overlays for rate $1/(2R)$ COP-STC with QPSK, $n = 2$, $m = 1$, $R = 1$ in quasistatic channel. The basic code for a single transmitter is STC as in [27].

μ	Generators $r = 0$	Generators $r = 1$	$\widetilde{F}_{\min}(1)/N$
2	$(1,2,3,1)_8$	$(1,2,3,1)_8$	0.00305
3	$(2,5,7,6)_8$	$(2,7,5,3)_8$	0.00109
4	$(11,15,17,13)_8$	$(06,15,13,12)_8$	0.00053

Table 4: Optimum overlays for rate $1/(2R)$ with QPSK, $n = 2$, $m = 1$, $R = 1$, in quasistatic channel. The basic code for single transmitter is STC with best convolutional code for the AWGN channel.

μ	Generators $r = 0$	Generators $r = 1$	$\widetilde{F}_{\min}(1)/N$
2	$(1,3,3,3)_8$	$(1,2,3,1)_8$	0.0075
3	$(5,7,7,7)_8$	$(2,6,5,3)_8$	0.00239
4	$(13,15,15,17)_8$	$(11,17,16,12)_8$	0.00069

trellis. By increasing the number of trellis states, only a small coding gain improvement is obtained. It is also worth noting that cooperative codes obtained by using the best P-STCs as a first code component usually perform better than the others, including the best available one from the literature [12]. It is also found in Table 5 that the 4-state code in [12] for $R = 2$ does not achieve full diversity.

7. NUMERICAL RESULTS

In this section, we report the performance results, in terms of frame error rate (FER) as a function of SNR, for cooperative pragmatic space-time codes (CP-STCs) and cooperative overlay pragmatic space-time codes (COP-STCs) in different conditions. The SNR is defined as E_b/N_0 per receiving-antenna element where, for a fair comparison among situations with different number of relays, E_b is the total energy per information bit over all transmitting nodes and averaged with respect to fading. We refer here to applications with a static set of relays in the simplified case of equal quality on all source-relay links. The probability that a relay cooperates with the source, P_{coop}, is given by

$$P_{\text{coop}} = 1 - P_e^{(S-R)}. \tag{24}$$

We first investigate in Figure 4 the effect of the number of states (ranging from 4 to 64) on the achievable diversity, in case of two relays cooperating with the source (i.e., $R = 2$ and

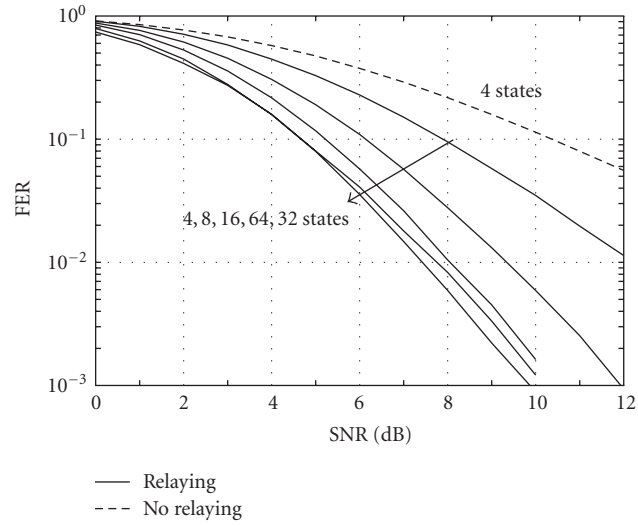


Figure 4: BPSK, optimal generators for AWGN, 4 to 64 states, 2 relays, 2 transmitting antennas per node, 1 receiving antenna, quasistatic fading channel $L = 1$.

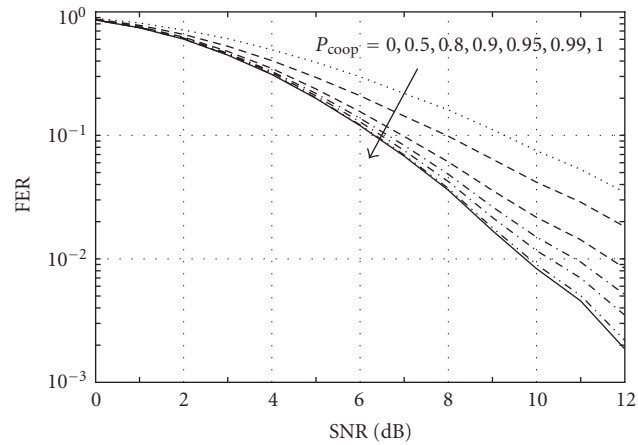


Figure 5: FER versus SNR for BPSK modulation, optimal generators for AWGN, 8 states, 1 relay, 2 transmitting antennas per node, 1 receiving antenna, respectively, in quasistatic fading channel ($L = 1$).

$P_{\text{coop}} = 1$). We assume CP-STCs with the rate 1/6 generators optimal for the AWGN channel, two transmitting antennas per node, one receiving antenna, a quasistatic fading channel (i.e., $L = 1$), and BPSK modulation. It is noticeable that only a portion of the available space-time diversity can be

Table 5: Optimum overlays for rate $1/(2R)$ COP-STC with QPSK, $n = 1$, $m = 1$, $R \in \{1, 2, 3\}$, in quasistatic channel. The basic code for a single transmitter is the best convolutional code for the AWGN channel. The bottom part of the table refers to C-STC as in [12].

μ	Generators $r = 0$	Generators $r = 1$	Generators $r = 2$	Generators $r = 3$	$\widetilde{F}_{\min}(1)/N$ $R = 1$	$\widetilde{F}_{\min}(1)/N$ $R = 2$	$R = 3$	FER ($R = 1$ @12 dB)	FER ($R = 2$ @12 dB)
1	$(1,3)_8$	$(1,3)_8$	$(2,1)_8$	$(2,2)_8$	0.060	0.012	div < 4	0.041	0.040
2	$(5,7)_8$	$(1,3)_8$	$(6,4)_8$	$(2,1)_8$	0.101	0.0113	0.00237	0.026	0.012
3	$(15,17)_8$	$(11,13)_8$	$(05,16)_8$	$(16,13)_8$	0.189	0.0136	0.00148	0.015	0.0051
5	$(23,35)_8$	$(27,31)_8$	$(21,37)_8$		0.315	0.0157		0.013	0.0038
3	$(5,7)_8$	$(5,7)_8$	$(5,7)_8$	—	0.125	div < 3	—	0.020	0.022
4	$(15,17)_8$	$(13,15)_8$	$(17,13)_8$	—	0.323	0.0191	—	0.015	0.0069
5	$(23,35)_8$	$(25,37)_8$	$(27,33)_8$	—	0.401	0.0169	—	0.013	0.038

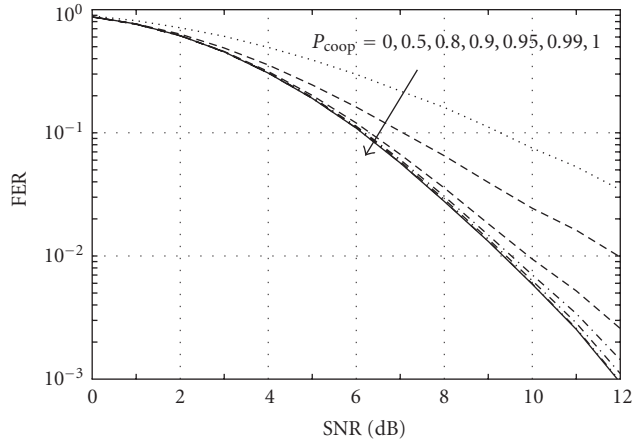


Figure 6: FER versus SNR for BPSK modulation, optimal generators for AWGN, 8 states, two relay, 2 transmitting antennas per node, one receiving antenna, respectively, in quasistatic fading channel ($L = 1$).

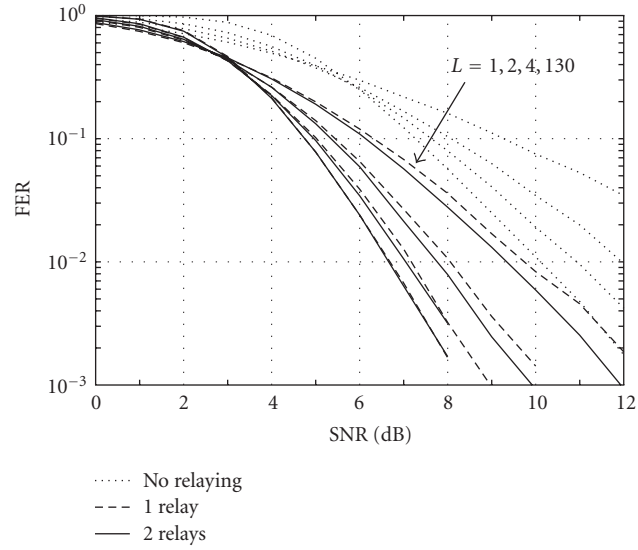


Figure 7: BPSK, optimal generators for AWGN, 16 states, 1 or 2 relays with $P_{\text{coop}} = 1, 2$ transmitting antennas per node, 1 receiving antenna, in BFC with various L.

achieved depending on the number of states, that is, $\widetilde{\eta}_{\min} = 3$ for 4-state codes, $\widetilde{\eta}_{\min} = 4$ for 8 and 64 states, $\widetilde{\eta}_{\min} = 5$ for 16 and 32 states, respectively. Note also that the generators for 64 state codes achieve a diversity smaller than that for 32 state codes. This is due to the the fact that pragmatic suboptimal construction does not always lead to the best possible generators for the direct and the relay phases. For understanding the impact of space-time diversity given by the cooperation phase, we also plot the case of rate 1/6 code with 4 states in the absence of relaying, showing that this code is not able to capture the same diversity degree as with relaying (achieving diversity 2). This is due to the fact that, in this case, the cooperative diversity of relaying is not available as can be seen from the branch metric evaluation.

In Figures 5 and 6, we show the FER versus SNR for CP-STCs with BPSK modulation, optimal generators for AWGN, 8 states, one relay, 2 transmitting antennas per node, and one receiving antenna in the quasistatic fading channel ($L = 1$). Here, the probability of cooperation P_{coop} takes values 0% (no cooperation), 50%, 80%, 90%, 95%, 99%, and 100% (i.e., certainty of cooperation). We investigate both the situations where up to 1 relay is available and optimal rate 1/4 generators for AWGN are used, and up to 2 relaying nodes are available and optimal rate 1/6 generators for AWGN are used. We can note from the figures that to approach the best performance a probability of cooperation larger than 0.95 and (0.9) are needed for 1 and 2 relays, respectively. On the other hand, the code used for 2 relays only achieves diversity 4.

We now study the impact of fading velocity, related to L, in the case of CP-STCs with BPSK modulation, optimal 8 state code generators in AWGN, two transmitting antennas per node, one receiving antenna at the destination. We consider the extreme cases of absence of cooperation (i.e., $P_{\text{coop}} = 0$), as well as of perfect cooperation (i.e., $P_{\text{coop}} = 1$) when varying L. The performance is reported in Figure 7. It is possible to note that, for the given number of states, as the available temporal diversity L increases, the performance with one relay approaches that with two relays, while a significant performance improvement is obtained with respect to the situation of absence of relaying. It is easy to see that the best results are obtained with the second

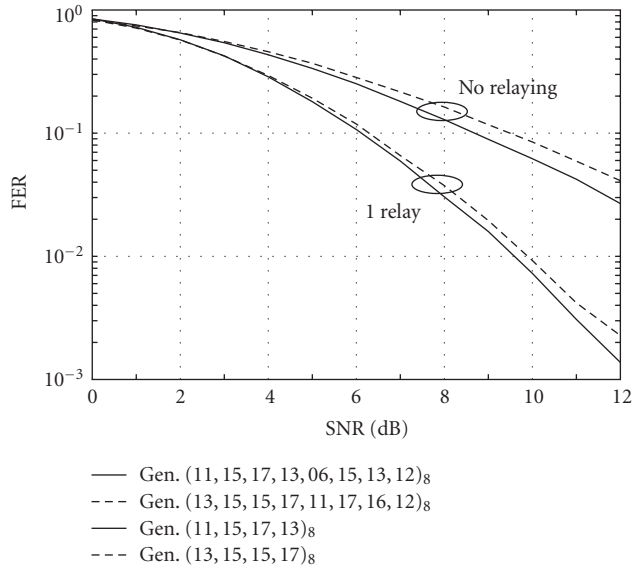


Figure 8: QPSK, 8 states ($\mu = 4$), generators as in Tables 3 and 4 with and without one relay, 2 transmitting antennas per node, 1 receiving antenna, in quasistatic fading channel.

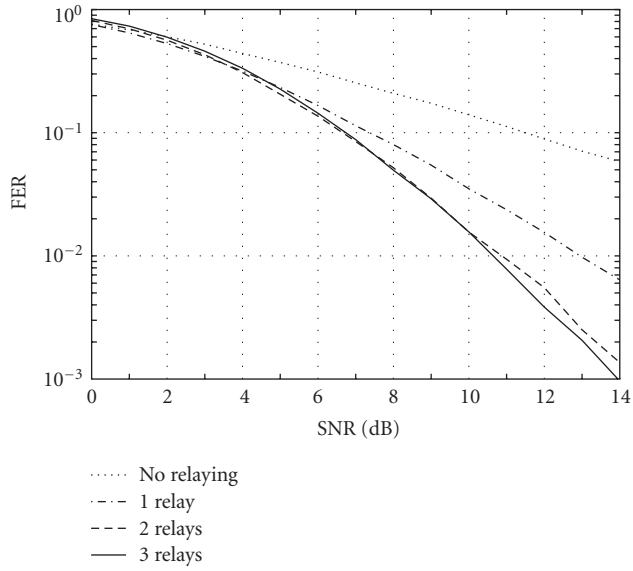


Figure 9: QPSK, 8 states ($\mu = 4$), generators as in Table 5, various number of relaying nodes, 1 transmitting antenna per node, 1 receiving antenna, in quasistatic fading channel.

approach, with a performance gain in agreement with the values of the performance factor ($\widetilde{F}_{\min}(1)/N$ is 0.00069 and 0.00053, resp.).

With the next figures we verify the performance of COP-STCs codes for QPSK modulation obtained through the design and search criterion explained in Section 5. In Figure 8, we show the performance without relaying and with one relay in the quasistatic fading channel (i.e., $L = 1$). Generators for rate $1/(4R)$ and 8 state codes are obtained through the two different approaches for overlay construction explained in Section 5, that is, designing the overall code starting from first code component taken as the the best rate 1/4 code for AWGN or as the best P-STCs reported in [27]. It is easy to see that the best results are obtained with the second approach.

Finally, we investigate in Figure 9 the impact of the number of relaying nodes, ranging from 0 to 3, for the 8 state COP-STCs with QPSK with one transmitting antenna per node and one receiving antenna in the quasistatic fading channel. The generators are those in Table 5 third line, and $P_{\text{coop}} = 1$ is assumed to fully exploit cooperation.

8. CONCLUSIONS

In this paper, we investigated the feasibility of a pragmatic approach to space-time coding for wireless cooperative relay networks, where standard convolutional encoders and decoders are used with suitably defined branch metrics.

We also proposed a design criterion to rank different codes with an efficient algorithm, based on the asymptotic error probability union bound. A search methodology to obtain optimum generators for different fading rates has then been given.

It has been shown that P-STCs applied to cooperative communication systems achieve comparable or improved performance when compared to previously known STCs and that they are suitable for systems with different spectral efficiencies, number of antennas and fading rates, making them a suitable choice in terms of both implementation complexity and performance.

ACKNOWLEDGMENTS

The authors would like to thank Dr. Mark Flanagan for the careful reading of the manuscript. This research was supported by the FP7 European project OPTIMIX. Part of this work was presented at the IEEE ICC 2008 Workshop on Cooperative Communications and Networking: Theory, Practice and Applications, Beijing, China.

REFERENCES

[1] M. Schwarz, W. R. Bennett, and S. Stein, *Communications Systems and Techniques*, IEEE Press, Piscataway, NJ, USA, 1996.

[2] J. G. Proakis, *Digital Communications*, McGraw-Hill, New York, NY, USA, 4th edition, 2001.

[3] J. H. Winters, "Smart antennas for wireless systems," *IEEE Personal Communications*, vol. 5, no. 1, pp. 23–27, 1998.

[4] S. M. Alamouti, "A simple transmit diversity technique for wireless communications," *IEEE Journal on Selected Areas in Communications*, vol. 16, no. 8, pp. 1451–1458, 1998.

[5] A. Narula, M. D. Trott, and G. W. Wornell, "Performance limits of coded diversity methods for transmitter antenna arrays," *IEEE Transactions on Information Theory*, vol. 45, no. 7, pp. 2418–2433, 1999.

[6] A. Wittneben, "New bandwidth efficient transmit antenna modulation diversity scheme for linear digital modulation," in *Proceedings of the IEEE International Conference on Communications (ICC '93)*, pp. 1630–1634, Geneva, Switzerland, May 1993.

[7] A. Sendonaris, E. Erkip, and B. Aazhang, "User cooperation diversity—part I: system description," *IEEE Transactions on Communications*, vol. 51, no. 11, pp. 1927–1938, 2003.

[8] A. Sendonaris, E. Erkip, and B. Aazhang, "User cooperation diversity—part II: implementation aspects and performance analysis," *IEEE Transactions on Communications*, vol. 51, no. 11, pp. 1939–1948, 2003.

[9] J. N. Laneman and G. W. Wornell, "Distributed space-time-coded protocols for exploiting cooperative diversity in wireless networks," *IEEE Transactions on Information Theory*, vol. 49, no. 10, pp. 2415–2425, 2003.

[10] Z. Lin, E. Erkip, and A. Stefanov, "Cooperative regions and partner choice in coded cooperative systems," *IEEE Transactions on Communications*, vol. 54, no. 7, pp. 1323–1334, 2006.

[11] J. Luo, R. S. Blum, L. J. Cimini, L. J. Greenstein, and A. M. Haimovich, "Decode-and-forward cooperative diversity with power allocation in wireless networks," *IEEE Transactions on Wireless Communications*, vol. 6, no. 3, pp. 793–799, 2007.

[12] A. Stefanov and E. Erkip, "Cooperative space-time coding for wireless networks," *IEEE Transactions on Communications*, vol. 53, no. 11, pp. 1804–1809, 2005.

[13] M. Dohler, Y. Li, B. Vucetic, A. H. Aghvami, M. Arndt, and D. Barthel, "Performance analysis of distributed space-time block-encoded sensor networks," *IEEE Transactions on Vehicular Technology*, vol. 55, no. 6, pp. 1776–1789, 2006.

[14] T. A. Cover and J. A. Thomas, *Elements of Information Theory*, John Wiley & Sons, New York, NY, USA, 1st edition, 1991.

[15] J. N. Laneman, D. N. C. Tse, and G. W. Wornell, "Cooperative diversity in wireless networks: efficient protocols and outage behavior," *IEEE Transactions on Information Theory*, vol. 50, no. 12, pp. 3062–3080, 2004.

[16] A. Bletsas, H. Shin, and M. Z. Win, "Cooperative communications with outage-optimal opportunistic relaying," *IEEE Transactions on Wireless Communications*, vol. 6, no. 9, pp. 3450–3460, 2007.

[17] J.-C. Guey, M. P. Fitz, M. R. Bell, and W.-Y. Kuo, "Signal design for transmitter diversity wireless communication systems over Rayleigh fading channels," in *Proceedings of the 46th IEEE Vehicular Technology Conference (VTC '96)*, vol. 1, pp. 136–140, Atlanta, Ga, USA, April-May 1996.

[18] V. Tarokh, N. Seshadri, and A. R. Calderbank, "Space-time codes for high data rate wireless communication: performance criterion and code construction," *IEEE Transactions on Information Theory*, vol. 44, no. 2, pp. 744–765, 1998.

[19] V. Tarokh, A. Naguib, N. Seshadri, and A. R. Calderbank, "Space-time codes for high data rate wireless communication: performance criteria in the presence of channel estimation errors, mobility, and multiple paths," *IEEE Transactions on Communications*, vol. 47, no. 2, pp. 199–207, 1999.

[20] B. Vucetic and J. Yuan, *Space-Time Coding*, John Wiley & Sons, Chichester, UK, 2003.

[21] H. Jafarkhani, *Space-Time Coding Theory and Practice*, Cambridge University Press, Cambridge, UK, 2005.

[22] S. Baro, G. Bauch, and A. Hansmann, "Improved codes for space-time trellis-coded modulation," *IEEE Communications Letters*, vol. 4, no. 1, pp. 20–22, 2000.

[23] Z. Chen, J. Yuan, and B. Vucetic, "Improved space-time time trellis coded modulation scheme on slow Rayleigh fading channels," *Electronics Letters*, vol. 37, no. 7, pp. 440–441, 2001.

[24] Q. Yan and R. S. Blum, "Improved space-time convolutional codes for quasi-static slow fading channels," *IEEE Transactions on Wireless Communications*, vol. 1, no. 4, pp. 563–571, 2002.

[25] M. Chiani, A. Conti, and V. Tralli, "A pragmatic approach to space-time coding," in *Proceedings of the IEEE International Conference on Communications (ICC '01)*, vol. 9, pp. 2794–2799, Helsinki, Finland, June 2001.

[26] M. Chiani, A. Conti, and V. Tralli, "Design and performance of bit-interleaved pragmatic space-time codes in block fading channels," in *Proceedings of the 13th IEEE International Symposium on Personal, Indoor and Mobile Radio Communications (PIMRC '02)*, vol. 2, pp. 683–687, Lisbon, Portugal, September 2002.

[27] M. Chiani, A. Conti, and V. Tralli, "Pragmatic space-time trellis codes for block fading channels," submitted to *IEEE Transactions on Information Theory*, available on-line at arxiv.org.

[28] R. J. McEliece and W. E. Stark, "Channels with block interference," *IEEE Transactions on Information Theory*, vol. 30, no. 1, pp. 44–53, 1984.

[29] M. Chiani, "Error probability for block codes over channels with block interference," *IEEE Transactions on Information Theory*, vol. 44, no. 7, pp. 2998–3008, 1998.

[30] E. Malkamäki and H. Leib, "Coded diversity on block-fading channels," *IEEE Transactions on Information Theory*, vol. 45, no. 2, pp. 771–781, 1999.

[31] A. Stefanov and E. Erkip, "Cooperative coding for wireless networks," *IEEE Transactions on Communications*, vol. 52, no. 9, pp. 1470–1476, 2004.

[32] J. H. Winters, J. Salz, and R. D. Gitlin, "Impact of antenna diversity on the capacity of wireless communication systems," *IEEE Transactions on Communications*, vol. 42, no. 234, part 3, pp. 1740–1751, 1994.

[33] H. El Gamal and A. R. Hammons Jr., "On the design of algebraic space-time codes for MIMO block-fading channels," *IEEE Transactions on Information Theory*, vol. 49, no. 1, pp. 151–163, 2003.

[34] M. Chiani, A. Conti, and V. Tralli, "Bit-interleaved pragmatic space-time codes: design and code construction," in *Proceedings of the IEEE Global Telecommunications Conference (GLOBECOM '02)*, vol. 2, pp. 1940–1944, Taipei, Taiwan, November 2002.

[35] M. Chiani, A. Conti, and V. Tralli, "Further results on convolutional code search for block-fading channels," *IEEE Transactions on Information Theory*, vol. 50, no. 6, pp. 1312–1318, 2004.

[36] G. Caire and E. Viterbo, "Upper bound on the frame error probability of terminated trellis codes," *IEEE Communications Letters*, vol. 2, no. 1, pp. 2–4, 1998.

[37] H. El Gamal, A. R. Hammons Jr., and A. Stefanov, "Space-time overlays for convolutionally coded systems," *IEEE Transactions on Communications*, vol. 51, no. 9, pp. 1603–1612, 2003.

[38] E. Malkamäki and H. Leib, "Evaluating the performance of convolutional codes over block fading channels," *IEEE Transactions on Information Theory*, vol. 45, no. 5, pp. 1643–1646, 1999.

Hindawi Publishing Corporation
EURASIP Journal on Advances in Signal Processing
Volume 2008, Article ID 125735, 11 pages
doi:10.1155/2008/125735

Research Article

Interference Mitigation in Cooperative SFBC-OFDM

D. Sreedhar and A. Chockalingam

Department of Electrical Communication Engineering, Indian Institute of Science, Bangalore 560012, India

Correspondence should be addressed to A. Chockalingam, achockal@ece.iisc.ernet.in

Received 15 November 2007; Accepted 28 March 2008

Recommended by Andrea Conti

We consider cooperative space-frequency block-coded OFDM (SFBC-OFDM) networks with amplify-and-forward (AF) and decode-and-forward (DF) protocols at the relays. In cooperative SFBC-OFDM networks that employ DF protocol, (i), intersymbol interference (ISI) occurs at the destination due to violation of the "quasistatic" assumption because of the frequency selectivity of the relay-to-destination channels, and (ii) intercarrier interference (ICI) occurs due to imperfect carrier synchronization between the relay nodes and the destination, both of which result in error-floors in the bit-error performance at the destination. We propose an interference cancellation algorithm for this system at the destination node, and show that the proposed algorithm effectively mitigates the ISI and ICI effects. In the case of AF protocol in cooperative networks (without SFBC-OFDM), in an earlier work, we have shown that full diversity can be achieved at the destination if phase compensation is carried out at the relays. In cooperative networks using SFBC-OFDM, however, this full-diversity attribute of the phase-compensated AF protocol is lost due to frequency selectivity and imperfect carrier synchronization on the relay-to-destination channels. We propose an interference cancellation algorithm at the destination which alleviates this loss in performance.

1. INTRODUCTION

Cooperative communications have become popular in recent research, owing to the potential for several benefits when communicating nodes in wireless networks are allowed to cooperate [1]. A classical benefit that arises from cooperation among nodes is the possibility of achieving spatial diversity, even when the nodes have only one antenna. That is, cooperation allows single-antenna nodes in a multiuser environment to share their antennas with other nodes in a distributed manner so that a given node can realize a virtual multiantenna transmitter that provides transmit diversity benefits. Such techniques, termed as "cooperative diversity" techniques, have widely been researched [2, 3]. Achieving cooperative diversity benefits based on a relay node merely repeating the information sent by a source node comes at the price of loss of throughput because the relay-to-destination transmission requires a separate time slot [3]. This loss in throughput due to repetition-based cooperation can be alleviated by integrating channel coding with cooperation [4]. Also, cooperation methods using distributed space-time coding are widely being researched [5, 6].

Recent investigations on cooperative communications focus on space-time cooperative systems based on OFDM [7–11]. Since space-time codes were developed originally for frequency-flat channels, an effective way to use them on frequency selective channels is to use them along with OFDM. A major advantage of space-time OFDM (ST-OFDM) is that a frequency selective channel is converted into multiple frequency flat channels [12], and with a proper outer code applied along with ST-OFDM code as an inner code, the full diversity of a frequency selective channel (i.e., multipath diversity) can be exploited as well. In addition to multipath diversity, user-cooperation diversity can be achieved in cooperative ST-OFDM (CO-ST-OFDM) systems, where space-time block codes (STBC) can be used in the relaying phase of cooperation [7, 8]. Accurate time and frequency synchronization, however, are crucial in achieving the promised potential of CO-ST-OFDM [8–11]. For example, in the context of cooperative OFDM, the relays-to-destination transmissions during the relaying phase of the protocol resemble transmissions from multiple noncooperating users in an upink OFDMA system [13, 14]. Hence nonzero carrier frequency offsets (CFOs) arising due

to imperfect carrier synchronization between the relays and the destination results in multiuser interference (multiple relays viewed as virtual multiple users) at the destination. A similar effect will occur if the timing synchronization is imperfect, that is, with nonzero timing offset. Without any effort to handle this interference, the performance of cooperative OFDM may end up being worse than that of OFDM without cooperation, particularly when the synchronization errors (in terms of CFOs and timing offsets) are large, and hence interference cancellation (IC) techniques employed at the destination will be of interest. Equalization techniques to alleviate the effect of carrier frequency offsets in distributed STBC-OFDM have been reported in the literature [10]. Practical timing and frequency synchronization algorithms and channel estimation for CO-ST-OFDM using Alamouti code [15] have been investigated in [8].

An alternate way to employ space-time codes in MIMO OFDM is to perform coding across space and frequency (instead of coding across space and time), which is often referred to as space-frequency coding (SFC) [16–19]. One way to do space-frequency coding is to take space-time codes and apply them in frequency dimension instead of time dimension [16]. The advantages of using space-frequency codes along with OFDM are low delays and robustness to time-selectivity of the channel [19]. Our focus, accordingly, in this paper is on cooperative OFDM systems when space-frequency block codes (SFBC) are employed; we refer to these systems as cooperative SFBC-OFDM (CO-SFBC-OFDM) systems.

Our new contribution in this paper can be highlighted as follows. In CO-SFBC-OFDM networks that employ decode-and-forward (DF) protocol, (i) intersymbol interference (ISI) occurs at the destination due to violation of the "quasistatic" assumption because of the frequency selectivity of the relay-to-destination channels, and (ii) intercarrier interference (ICI) occurs due to imperfect carrier synchronization between the relay nodes and the destination, both of which result in errorfloors in the bit error performance at the destination. We propose an interference cancellation algorithm for this system at the destination node, and show that the proposed algorithm effectively mitigates the ISI and ICI effects. In the case of amplify-and-forward (AF) protocol in cooperative networks (without SFBC-OFDM), in our earlier work in [20], we have shown that full diversity can be achieved at the destination if phase compensation is carried out at the relays. In cooperative networks using SFBC-OFDM, however, this full-diversity attribute of the phase-compensated AF protocol is lost due to frequency selectivity and imperfect carrier synchronization on the relay-to-destination channels. To address this problem, we propose an interference cancellation algorithm at the destination which alleviates this loss in performance.

The rest of this paper is organized as follows. In Section 2, we present the CO-SFBC-OFDM system model with AF protocol and phase compensation at the relays, and illustrate the ISI and ICI effects. The proposed IC algorithm for this system is presented in Section 2.2. Section 3 presents the system model for CO-SFBC-OFDM system with DF protocol at the relays, and illustrates the associated ISI and ICI effects. The proposed IC algorithm for this DF protocol system is presented in Section 3.2. Results and discussions for both AF and DF protocols are presented in Section 4. Conclusions are given in Section 5.

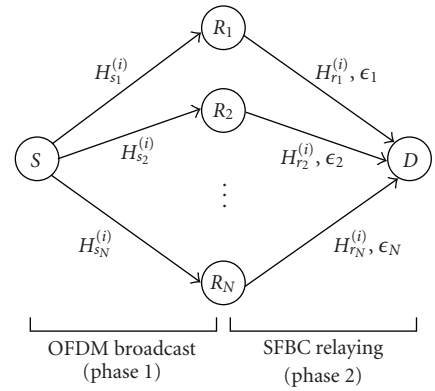


Figure 1: A cooperative SFBC-OFDM network consisting of one source, one destination, and N relays.

2. COOPERATIVE SFBC-OFDM WITH AF PROTOCOL

Consider a wireless network as depicted in Figure 1 with $N+2$ nodes consisting of a source, a destination and N relays. All nodes are half duplex nodes, that is, a node can either transmit or receive at a time. OFDM is used for transmission on the source-to-relays and relays-to-destination links. The destination is assumed to know (i) source-to-relays channel state information (CSI) and (ii) relays-to-destination CSI. Each relay is assumed to know the phase information of the channel from the source to itself. We employ amplification and channel phase compensation on the received signals at the relays. The transmission protocol is as follows (see Figures 1 and 2):

(i) In the first time slot (i.e., phase 1), the source transmits information symbols $X^{(k)}, 1 \le i \le M$ using an M subcarrier OFDM symbol. All the N relays receive this OFDM symbol. This phase is called the *OFDM broadcast phase*.

(ii) In the second time slot (i.e., phase 2), N relays forward the received information. (We assume that all the relays participate in the cooperative transmission. It is also possible that some relays do not participate in the transmission based on whether the channel state is in outage or not. We do not consider such a partial participation scenario here.) For the AF protocol, the relays perform channel phase compensation and amplification on the received signal, followed by space-frequency block coding. This phase is called *AF-SFBC relay phase*. The destination receives these transmissions, performs ICI/ISI cancellation and SFBC decoding.

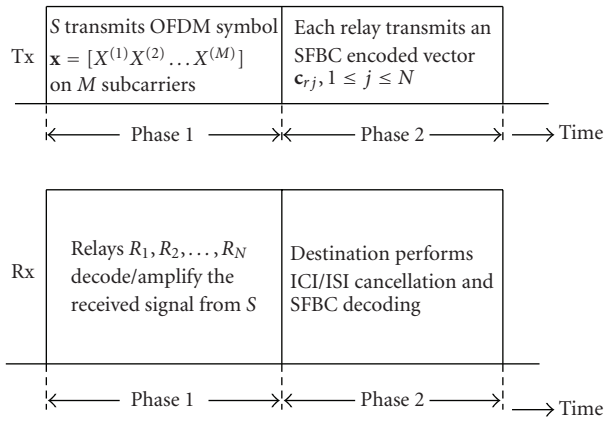


Figure 2: AF/DF transmission protocol in a cooperative SFBC-OFDM network.

Broadcast reception at the relays

Let $\mathbf{x} = [X^{(1)}, X^{(2)}, \ldots, X^{(M)}]$ denote the information symbol vector transmitted by the source on M subcarriers. (We use the following notation in this paper: Bold letter uppercase is used to represent matrices and bold letter lower case is used to represent vectors. $\Re(\cdot)$ denotes real value of a complex argument and $\Im(\cdot)$ denotes imaginary value. $x^{(I)}$ and $x^{(Q)}$ denote the real and imaginary parts of the complex number x. $(\cdot)^{\mathcal{H}}$ and $(\cdot)^T$ denote matrix conjugate transposition and matrix transposition, respectively. $(\cdot)^*$ denotes matrix conjugation. $\text{diag}\{a_1, a_2, \ldots, a_N\}$ is a diagonal matrix having diagonal entries $a_1, a_2, \ldots, a_N$. $\mathbf{j}$ denotes $\sqrt{-1}$. $E\{\cdot\}$ denotes expectation operation.) The received signal, $v_{rj}^{(k)}$, on the kth subcarrier at the jth relay during the *OFDM broadcast phase* can be written as

$$v_{rj}^{(k)} = \sqrt{E_1} H_{sj}^{(k)} X^{(k)} + Z_{rj}^{(k)}, \quad 1 \le i \le M,\ 1 \le j \le N, \quad (1)$$

where $H_{sj}^{(k)}$ is the frequency response on the kth subcarrier of the channel from source to jth relay, given by $H_{sj}^{(k)} = \text{DFT}_M(h_{sj}^{(n)})$, where $h_{sj}^{(n)}$ is the time-domain impulse response of the channel from source to jth relay. (In all the source-to-relay and relay-to-destination links, we assume frequency-selective block fading channel model [21, 22]. The maximum delay spread of the channel is assumed to be less than the added guard interval. The channel is assumed to be static for one OFDM symbol duration.) $Z_{rj}^{(k)}$ is additive white Gaussian noise with zero mean and variance σ^2, and $E\{|X^{(k)}|^2\} = 1$. E_1 is the energy per symbol spent in the broadcast phase. On the source-to-relay links, all the relays listen to the source and each relay can compensate for its CFO individually. Hence there is no ISI/ICI on the source-to-relay links.

Space-frequency block coding at the relay in AF protocol

At the relay j, first, phase compensation followed by an amplification of the received signal is done. Let $H_{sj}^{(k)} = |H_{sj}^{(k)}| e^{\mathbf{j}\theta_{sj}^{(k)}}$. The operation at the relay can then be described as (i) phase compensation (i.e, multiplication by $e^{-\mathbf{j}c\theta_{sj}^{(k)}}$), and (ii) amplification on $v_{rj}^{(k)}$ such that energy per transmission is E_2, that is,

$$\hat{v}_{rj}^{(k)} = \sqrt{\frac{E_2}{E_1 + \sigma^2}} e^{-\mathbf{j}\theta_{sj}^{(k)}} v_{rj}^{(k)}, \quad (2)$$

$$= \sqrt{\frac{E_1 E_2}{E_1 + \sigma^2}} \left| H_{sj}^{(k)} \right| X^{(k)} + \hat{Z}_{rj}^{(k)}, \quad (3)$$

where

$$\hat{Z}_{rj}^{(k)} = \sqrt{\frac{E_2}{E_1 + \sigma^2}} e^{-\mathbf{j}\theta_{sj}^{(k)}} Z_{rj}^{(k)}. \quad (4)$$

The space-frequency block encoding at the relays is illustrated in Figure 3. An $N \times K$ space-time block code (STBC) matrix with P information symbols is used across subcarriers in N-relays. For the *AF-SFBC relay phase* transmission, we divide the M subcarriers into M_g groups such that $M = M_g K + \kappa$. If M is not a multiple of K then, there will not be any transmission on κ subcarriers, and accordingly the source will transmit only $M_g P$ information symbols and there will be no transmission on $M - M_g P$ subcarriers from the source. Note that $M_g P \le M$ since $P/K \le 1$ for the STBC codes considered. Now, for each relay j, we form M_g groups out of the $M_g P$ values in $\hat{v}_{rj}^{(k)}$, and, for each group q, we form the $2P \times 1$ vector $\hat{\mathbf{v}}_{rj}^{(q)}$, given by

$$\hat{\mathbf{v}}_{rj}^{(q)} = \left[\hat{v}_{rj}^{((q-1)P+1)(I)}, \hat{v}_{rj}^{((q-1)P+1)(Q)}, \hat{v}_{rj}^{((q-1)P+2)(I)}, \hat{v}_{rj}^{((q-1)P+2)(Q)}, \right.$$
$$\left. \ldots \hat{v}_{rj}^{(qP)(I)}, \hat{v}_{rj}^{(qP)(Q)} \right]^T. \quad (5)$$

The space-frequency coded symbols for the qth group of the jth relay can be obtained as

$$\mathbf{c}_{rj}^{(q)} = \mathbf{A}_j \hat{\mathbf{v}}_{rj}^{(q)} = \sqrt{\frac{E_1 E_2}{E_1 + \sigma^2}} \mathbf{A}_j \mathbf{H}_{sj}^{(q)} \mathbf{x}^{(q)} + \mathbf{A}_j \hat{\mathbf{z}}_{rj}^{(q)}, \quad 1 \le q \le M_g, \quad (6)$$

where the $2P \times 2P$ matrix $\mathbf{H}_{sj}^{(q)} = \text{diag}[|H_{sj}^{((q-1)P+1)}|, |H_{sj}^{((q-1)P+1)}|, \ldots, |H_{sj}^{(qP)}|, |H_{sj}^{(qP)}|]$, the $2P \times 1$ vector $\hat{\mathbf{z}}_{rj}^{(q)} = [\hat{Z}_{rj}^{((q-1)P+1),(I)}, \hat{Z}_{rj}^{((q-1)P+1),(Q)}, \ldots, \hat{Z}_{rj}^{(qP),(I)}, \hat{Z}_{rj}^{(qP),(Q)}]^T$, and the $2P \times 1$ vector $\mathbf{x}^{(q)} = [X^{((q-1)P+1),(I)}, X^{((q-1)P+1),(Q)}, \ldots, X^{(qP),(I)}, X^{(qP),(Q)}]^T$. The $\mathbf{A}_j$ matrices perform the space-frequency encoding. For example, for the 2-relay case (i.e., $N = 2$) using Alamouti code:

$$\mathbf{A}_1 = \begin{bmatrix} 1 & 0 & \mathbf{j} & 0 \\ 0 & -1 & 0 & \mathbf{j} \end{bmatrix}, \qquad \mathbf{A}_2 = \begin{bmatrix} 0 & 1 & 0 & \mathbf{j} \\ 1 & 0 & \mathbf{j} & 0 \end{bmatrix}. \quad (7)$$

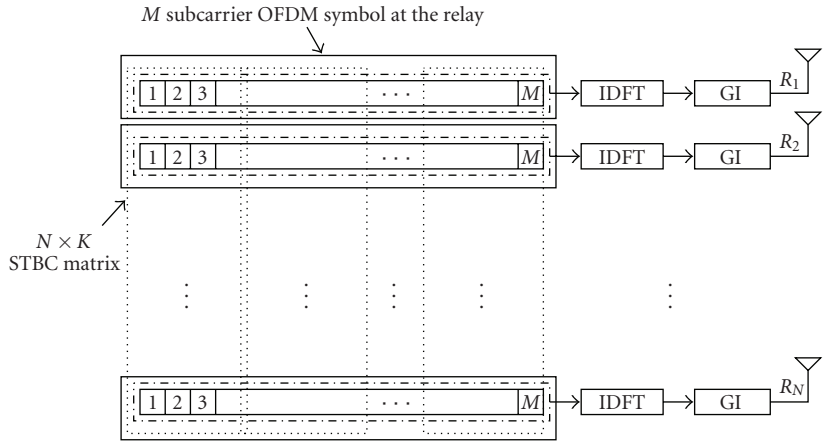


Figure 3: Space-frequency block coding at the relays.

The overall space-frequency coded symbol vector from the jth relay can be written as

$$\mathbf{c}_{rj} = \begin{bmatrix} \mathbf{c}_{rj}^{(1)} \\ \vdots \\ \mathbf{c}_{rj}^{(M_g)} \\ \mathbf{0}^{\kappa\times 1} \end{bmatrix}. \tag{8}$$

Finally, the inverse Fourier transform of $\mathbf{c}_{rj}$, that is, $\mathbf{t}_{rj} = \text{IDFT}(\mathbf{c}_{rj})$ is transmitted by the jth relay.

Received signal at the destination

The received time-domain baseband signal at the destination, after coarse carrier frequency synchronization and guard time removal, is given by

$$y^{(n)} = \sum_{j=1}^{N} \left(t_{rj}^{(n)} \star h_{jd}^{(n)}\right) e^{\mathrm{j}2\pi\epsilon_j n/N} + z_d^{(n)}, \quad 0 \le n \le M-1, \tag{9}$$

where $\star$ denotes linear convolution, h_{jd}^{n} is the channel impulse response from the jth relay to the destination. It is assumed that h_{jd}^{n} is nonzero only for $n = 0,\ldots,L-1$, where L is the maximum channel delay spread. It is also assumed that the added guard interval is greater than L. ϵ_j, $j = 1,\ldots N$, $0 \le |\epsilon_j| \le 0.5$, denotes residual carrier frequency offset (CFO) from the jth relay normalized by the subcarrier spacing, and $z_d^{(n)}$ is the AWGN with zero mean and variance σ_d^2. We assume that all the nodes are time synchronized and that ϵ_j, $j = 1,\ldots,N$ are known at the destination. At the destination, $y^{(n)}$ is first fed to the DFT block. The $M\times 1$ DFT output vector, $\mathbf{y}$, can be written in the form

$$\mathbf{y} = \sum_{j=1}^{N} \mathbf{\Psi}_j \mathbf{H}_{jd} \mathbf{c}_{rj} + \mathbf{z}_d, \tag{10}$$

where $\mathbf{\Psi}_j$ is a $M \times M$ circulant matrix given by

$$\mathbf{\Psi}_j = \begin{bmatrix} \psi_j^{(0)} & \psi_j^{(1)} & \cdots & \psi_j^{(M-1)} \\ \psi_j^{(M-1)} & \psi_j^{(0)} & \cdots & \psi_j^{(M-2)} \\ \vdots & \vdots & \ddots & \vdots \\ \psi_j^{(1)} & \psi_j^{(2)} & \cdots & \psi_j^{(0)} \end{bmatrix}, \tag{11}$$

where

$$\psi_j^{(k)} = \text{DFT}_M\left(e^{\mathrm{j}2\pi n\epsilon_j/M}\right). \tag{12}$$

$\mathbf{H}_{jd}$ is the $M \times M$ diagonal channel matrix given by $\mathbf{H}_{jd} = \text{diag}[H_{jd}^{(1)}, H_{jd}^{(2)},\ldots, H_{jd}^{(M)}]$, and the channel coefficient in frequency domain $H_{jd}^{(k)}$ is given by $H_{jd}^{(k)} = \text{DFT}_M(h_{jd}^{(n)})$. Similarly, $\mathbf{z}_d = [Z_d^{(1)}, Z_d^{(2)},\ldots, Z_d^{(M)}]$, where $Z_d^{(k)} = \text{DFT}_M(z_d^{(n)})$. Equation (10) can be rewritten as

$$\mathbf{y} = \sum_{j=1}^{N} \psi_j^{(0)} \mathbf{H}_{jd}\mathbf{c}_{rj} + \underbrace{\sum\nolimits_{j=1}^{N} (\mathbf{\Psi}_j - \psi_j^{(0)}\mathbf{I})\mathbf{H}_{jd}\mathbf{c}_{rj}}_{\text{ICI}} + \mathbf{z}_d. \tag{13}$$

If we collect the K entries of $\mathbf{y}$ corresponding to the qth SFBC block and form a $K \times 1$ vector $\mathbf{y}^{(q)}$, then we can write

$$\mathbf{y}^{(q)} = \sum_{j=1}^{N} \psi_j^{(0)} \mathbf{H}_{jd}^{(q)} \mathbf{c}_{rj}^{(q)} + \sum_{j=1}^{N} \left(\mathbf{\Psi}_j - \psi_j^{(0)}\mathbf{I}\right)^{[q]} \mathbf{H}_{jd}\mathbf{c}_{rj} + \mathbf{z}_d^{(q)}, \tag{14}$$

where $\mathbf{H}_{jd}^{(q)} = \text{diag}[H_{jd}^{((q-1)K+1)},\ldots, H_{jd}^{(qK)}]$, $\mathbf{z}_d^{(q)} = [Z_d^{((q-1)K+1)}, \ldots, Z_d^{(qK)}]^T$ and $(\cdot)^{[q]}$ denotes picking the K rows of a matrix starting from $(q-1)K+1$.

Optimal ML detector and zero-forcing detector

Using (6), the $\mathbf{c}_{rj}$ vector in (8) can be written as

$$\mathbf{c}_{rj} = \underbrace{\sqrt{\frac{E_1E_2}{E_1+\sigma^2}}\begin{bmatrix} \mathbf{A}_j\mathbf{H}_{sj}^{(1)} & \mathbf{0} & \cdots & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{A}_j\mathbf{H}_{sj}^{(2)} & \cdots & \mathbf{0} & \mathbf{0} \\ \vdots & \mathbf{0} & \ddots & \vdots & \vdots \\ \mathbf{0} & \mathbf{0} & \cdots & \mathbf{A}_j\mathbf{H}_{sj}^{(M_g)} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \cdots & \mathbf{0} & \mathbf{0} \end{bmatrix}}_{\Omega_j} \underbrace{\begin{bmatrix} \mathbf{x}^{(1)} \\ \mathbf{x}^{(2)} \\ \vdots \\ \mathbf{x}^{(M_g)} \\ \mathbf{0} \end{bmatrix}}_{\mathbf{x}} + \underbrace{\begin{bmatrix} \mathbf{A}_j\hat{\mathbf{z}}_{rj}^{(1)} \\ \mathbf{A}_j\hat{\mathbf{z}}_{rj}^{(2)} \\ \vdots \\ \mathbf{A}_j\hat{\mathbf{z}}_{rj}^{(M_g)} \\ \mathbf{0} \end{bmatrix}}_{\eta_j}. \tag{15}$$

Substituting this in (10), we get

$$\mathbf{y} = \underbrace{\left(\sum_{j=1}^{N}\mathbf{\Psi}_j\mathbf{H}_{jd}\mathbf{\Omega}_j\right)}_{\mathbf{\Phi}}\mathbf{x} + \sum_{j=1}^{N}\mathbf{\Psi}_j\mathbf{H}_{jd}\boldsymbol{\eta}_j + \mathbf{z}_d. \tag{16}$$

The optimal ML detection of $\mathbf{x}$ is given by

$$\tilde{\mathbf{x}} = \arg\min_{\mathbf{x}}(\mathbf{y} - \mathbf{\Phi}\mathbf{x})^{\mathcal{H}}\mathbf{\Sigma}^{-1}(\mathbf{y} - \mathbf{\Phi}\mathbf{x}), \tag{17}$$

where $\mathbf{\Sigma}$ is the covariance matrix of $\sum_{j=1}^{N}\mathbf{\Psi}_j\mathbf{H}_{jd}\eta_j + \mathbf{z}_d$. This has complexity of the order $O(\mathcal{M}^{\lfloor M/K\rfloor P})$, where $\mathcal{M}$ is the cardinality of the signal set used. A suboptimal zero-forcing detection can be carried out using

$$\tilde{\mathbf{y}} = (\mathbf{\Phi}^{\mathcal{H}}\mathbf{\Phi})^{-1}\mathbf{\Phi}^{\mathcal{H}}\mathbf{y}. \tag{18}$$

Since $\mathbf{\Phi}$ is of size $M \times M$, the inversion operation is of complexity $O(M^4)$. Interference cancellers at much lesser complexity can be adopted for the detection. In the following, we formulate the proposed ISI-ICI cancellation approach.

Detection in frequency-flat channel in the absence of CFO

For a frequency-flat channel, all the diagonal entries of $\mathbf{H}_{sj}^{(q)}$ and $\mathbf{H}_{jd}^{(q)}$ become equal. Hence in frequency-flat channel with no CFO, (14) reduces to

$$\mathbf{y}^{(q)} = \sum_{j=1}^{N}\left|H_{sj}^{((q-1)2P+1)}\right|\left|H_{jd}^{((q-1)K+1)}\right|\mathbf{A}_j\mathbf{x}^{(q)} + \sum_{j=1}^{N}\mathbf{A}_j\hat{\mathbf{z}}_{rj}^{(q)} + \mathbf{z}_d^{(q)}. \tag{19}$$

Define $\mathbf{H}_{\text{eq}}^{(q)} = \sum_{j=1}^{N}|H_{sj}^{((q-1)2P+1)}||H_{jd}^{((q-1)K+1)}|\mathbf{A}_j$. It can then be verified from the results in [20] that $\Re(\mathbf{H}_{\text{eq}}^{(q)^{\mathcal{H}}}\mathbf{H}_{\text{eq}}^{(q)})$ is a block diagonal matrix, and hence with the operation $\Re(\mathbf{H}_{\text{eq}}^{(q)^{\mathcal{H}}}\mathbf{y}^{(q)})$ it is possible to do full-diversity symbol-by-symbol detection of $\mathbf{y}^{(q)}$. But when the channel is frequency-selective and CFOs are nonzero, this detection gives rise to ISI and ICI, which we will analyze in the following Section 2.1.

2.1. ICI and ISI in AF protocol

Now we analyze the ICI and ISI at the output of the detection scheme described in Section 2, when the relays-to-destination channels as well as the source-to-relays channels are frequency-selective and when CFOs are not equal to zero. Define

$$\mathbf{H}_{\text{eq-af}}^{(q)} = \sum_{j=1}^{N}\sqrt{\frac{E_1E_2}{E_1+\sigma^2}}\psi_j^{(0)}\left|H_{sj}^{((q-1)2P+1)}\right|\left|H_{jd}^{((q-1)K+1)}\right|\mathbf{A}_j. \tag{20}$$

Since $\sqrt{E_1E_2/(E_1+\sigma^2)}\psi_j^{(0)}$ is a scalar, it is easily verified from the results in [20] that $\Re(\mathbf{H}_{\text{eq-af}}^{(q)^{\mathcal{H}}}\mathbf{H}_{\text{eq-af}}^{(q)})$ is a block diagonal matrix. Next, we split the channel matrices $\mathbf{H}_{sj}^{(q)}$ and $\mathbf{H}_{jd}^{(q)}$ into a quasistatic part and a nonquasistatic part, as

$$\begin{aligned} \mathbf{H}_{sj}^{(q)} &= \underbrace{\left|H_{sj}^{((q-1)2P+1)}\right|\mathbf{I}}_{\mathbf{H}_{sj,\text{qs}}^{(q)}} + \underbrace{\begin{bmatrix} 0 & 0 & \cdots & 0 \\ 0 & V & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \left|H_{sj}^{(q2P)}\right| - \left|H_{sj}^{((q-1)2P+1)}\right| \end{bmatrix}}_{\mathbf{H}_{sj,\text{nqs}}^{(q)}}, \\ \mathbf{H}_{jd}^{(q)} &= \underbrace{H_{jd}^{((q-1)K+1)}\mathbf{I}}_{\mathbf{H}_{jd,\text{qs}}^{(q)}} + \underbrace{\begin{bmatrix} 0 & 0 & \cdots & 0 \\ 0 & S & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & H_{jd}^{(qK)} - H_{jd}^{((q-1)K+1)} \end{bmatrix}}_{\mathbf{H}_{jd,\text{nqs}}^{(q)}}, \end{aligned} \tag{21}$$

where V denotes $|H_{sj}^{((q-1)2P+2)}| - |H_{sj}^{((q-1)2P+1)}|$, and S denotes $H_{jd}^{((q-1)K+2)} - H_{jd}^{((q-1)K+1)}$.

Using this, the output of the operation $\Re\left(\mathbf{H}_{\text{eq-af}}^{(q)\,\mathcal{H}}\mathbf{y}^{(q)}\right)$ on (14) can be written as

$$\begin{aligned}\hat{\mathbf{y}}^{(q)} = &\underbrace{\Re(\mathbf{H}_{\text{eq-af}}^{(q)\,\mathcal{H}}\mathbf{H}_{\text{eq-af}}^{(q)})\mathbf{x}^{(q)}}_{\text{Signal part}}\\ &+\underbrace{\Re(\mathbf{H}_{\text{eq-af}}^{(q)\,\mathcal{H}}\textstyle\sum_{j=1}^{N}\psi_j^{(0)}W)}_{\substack{\text{ISI due to frequency-selectivity}\\ \text{of broadcast and relay channels}}}\mathbf{x}^{(q)}\\ &+\underbrace{\Re\left(\mathbf{H}_{\text{eq-af}}^{(q)\,\mathcal{H}}\textstyle\sum_{j=1}^{N}(\boldsymbol{\Psi}_j-\psi_j^{(0)}\,\mathbf{I}\,)^{[q]}\mathbf{H}_{jd}\mathbf{c}_{rj}\right)}_{\text{ICI due to CFOs}}\\ &+\underbrace{\Re\left(\mathbf{H}_{\text{eq-af}}^{(q)\,\mathcal{H}}\left(\textstyle\sum_{j=1}^{N}\mathbf{A}_j\hat{\mathbf{z}}_{rj}^{(q)}+\mathbf{z}_d^{(q)}\right)\right)}_{\text{Total noise}},\end{aligned}\tag{22}$$

where W denotes that $(\mathbf{H}_{jd,\text{nqs}}^{(q)}\mathbf{A}_j\mathbf{H}_{sj,\text{qs}}^{(q)}+\mathbf{H}_{jd,\text{qs}}^{(q)}\mathbf{A}_j\mathbf{H}_{sj,\text{nqs}}^{(q)}+\mathbf{H}_{jd,\text{nqs}}^{(q)}\mathbf{A}_j\mathbf{H}_{sj,\text{nqs}}^{(q)})$.

As pointed out earlier, the optimum detector in this case would be a joint maximum-likelihood detector in PM_g variables, which has a prohibitive exponential receiver complexity.

2.2. Proposed ISI-ICI cancelling detector for AF protocol

In this section, we propose a two-step parallel interference canceling (PIC) receiver that cancels the frequency-selectivity-induced ISI, and the CFO-induced ICI. The proposed detector estimates and cancels the ISI (caused due to the violation of the quasistatic assumption) in the first step, and then estimates and cancels the ICI (caused due to loss of subcarrier orthogonality because of CFO) in the second step. This two-step procedure is then carried out in multiple stages. The proposed detector is presented in the following.

As can be seen, (22) identifies the desired signal, ISI, ICI, and noise components present in the output $\hat{\mathbf{y}}^{(q)}$. Based on this received signal model and the knowledge of the matrices $\mathbf{H}_{jd,\text{nqs}}^{(q)}$, $\mathbf{H}_{jd,\text{qs}}^{(q)}$, $\mathbf{H}_{sj,\text{nqs}}^{(q)}$, $\mathbf{H}_{sj,\text{qs}}^{(q)}$, and $\mathbf{H}_{\text{eq-af}}^{(q)}$, for all q, j we formulate the proposed interference estimation and cancellation procedure as follows.

(1) For each space-frequency code block q, estimate the information symbols $\hat{\mathbf{x}}^{(q)}$ from (22), ignoring ISI and ICI.

(2) For each space-frequency code block q, obtain an estimate of the ISI (i.e., an estimate of the ISI term in (22)) from the estimated symbols $\hat{\mathbf{x}}^{(q)}$ in the previous step.

(3) Cancel the estimated ISI from $\hat{\mathbf{y}}^{(q)}$.

(4) Using $\hat{\mathbf{x}}^{(q)}$ from step 1, regenerate $\hat{\mathbf{c}}^{(q)}$ using (6). Then, using $\hat{\mathbf{c}}^{(q)}$, obtain an estimate of the ICI (i.e., an estimate of the ICI term in (22)).

(5) Cancel the estimated ICI from the ISI-canceled output in step 3.

(6) Take the ISI- and ICI-canceled output from step 5 as the input back to step 1 (for the next stage of cancellation).

Based on the above, and $\boldsymbol{\Lambda}_{\text{af}}^{(q)} = \Re(\mathbf{H}_{\text{eq-af}}^{(q)\,\mathcal{H}}\mathbf{H}_{\text{eq-af}}^{(q)})$, the cancellation algorithm for the mth stage can be summarized as in Algorithm 1.

It is noted that Algorithm 1 has polynomial complexity. Also, $\boldsymbol{\Lambda}_{\text{af}}^{(q)}$ is a full-rank block diagonal matrix, and its inversion in the second equation in Algorithm 1 is simple. Assuming that the multiplication of the matrices $\mathbf{A}_j$ with $\mathbf{H}_{sj}, \mathbf{H}_{jd}$ could be precomputed, the total number of complex multiplications required for m stages of the proposed iterative interference cancellation is $2P\lfloor M/K\rfloor(K+2P+(m-1)(4P+2K+NK))$, which is much less complex than the zero-forcing detector complexity of $O(M^4)$.

3. COOPERATIVE SFBC-OFDM WITH DF PROTOCOL

The broadcast phase of the transmission protocol is the same for both AF protocol as well as DF protocol. In the relay phase of the DF protocol, however, the relays decode the information (instead of merely amplifying it) sent by the source, and transmits a space-frequency encoded version of this decoded information. This phase is called *DF-SFBC relay phase*. The destination receives this transmission, does ISI and ICI cancellation, followed by SFBC decoding.

Space-frequency block coding at the relay in DF protocol

We employ the same space-frequency encoding strategy as in AF protocol, except that instead of an amplification operation in (2) at the relay j, a decoding of the information symbols is done, that is, the decoded symbol on the kth subcarrier at the jth relay, denoted by $\widetilde{X}_j^{(k)}$, is obtained as

$$\widetilde{X}_j^{(k)} = \sqrt{E_2}\left(\arg\min_{X^{(k)}}\left|\left|v_{rj}^{(k)}-\sqrt{E_1}H_{sj}^{(k)}X^{(k)}\right|\right|^2\right),\qquad 1\le i\le M_gP,\ 1\le j\le N,\tag{23}$$

where E_2 is the energy per transmission in the relay phase. The corresponding space-frequency coded symbols for the qth group of subcarriers of the jth relay is obtained as

$$\mathbf{c}_{rj}^{(q)} = \mathbf{A}_j\widetilde{\mathbf{x}}_j^{(q)},\tag{24}$$

where $\widetilde{\mathbf{x}}_j^{(q)} = [\widetilde{X}_j^{((q-1)P+1),(I)}, \widetilde{X}_j^{((q-1)P+1),(Q)}, \ldots, \widetilde{X}_j^{(qP),(I)}, \widetilde{X}_j^{(qP),(Q)}]^T$. The received signal model at the destination in the DF protocol is the same as in (14), with $\mathbf{c}_{rj}^{(q)}$ generated as in (24). It is possible that the symbol vector $\mathbf{x}$ is detected differently at each relay. For the purpose of developing the IC algorithm, however, and henceforth in this paper, we assume that $\widetilde{\mathbf{x}}_j^{(q)} = \widetilde{\mathbf{x}}_k^{(q)}\ \forall\, j,k$ and drop the j index from $\widetilde{\mathbf{x}}_j^{(q)}$. In all our simulations, however, we will use the actual $\widetilde{\mathbf{x}}_j^{(q)}$'s at the relays.

Initialization: Set $m = 1$.
Evaluate

$$\hat{\mathbf{y}}^{(q,m)} = \Re\left(\mathbf{H}_{\text{eq-af}}^{(q)\,\mathcal{H}}\mathbf{y}^{(q)}\right), \quad 1 \le q \le M_g.$$

Loop
Estimate

$$\hat{\mathbf{x}}^{(q,m)} = \left(\mathbf{\Lambda}_{\text{af}}^{(q)}\right)^{-1}\hat{\mathbf{y}}^{(q,m)}, \quad 1 \le q \le M_g.$$

Cancel ISI

$$\hat{\mathbf{y}}^{(q,m+1)} = \hat{\mathbf{y}}^{(q,1)} - \Re\left(\mathbf{H}_{\text{eq-af}}^{(q)\,\mathcal{H}}\sum_{j=1}^{N}\psi_j^{(0)}\left(\mathbf{H}_{jd,\text{nqs}}^{(q)}\mathbf{A}_j\mathbf{H}_{sj,\text{qs}}^{(q)} + \mathbf{H}_{jd,\text{qs}}^{(q)}\mathbf{A}_j\mathbf{H}_{sj,\text{nqs}}^{(q)} + \mathbf{H}_{jd,\text{nqs}}^{(q)}\mathbf{A}_j\mathbf{H}_{sj,\text{nqs}}^{(q)}\right)\right)\hat{\mathbf{x}}^{(q,m)}, \quad 1 \le q \le M_g.$$

Form $\hat{\mathbf{c}}_{rj}^{(q,m)}$ *from*

$$\hat{\mathbf{c}}_{rj}^{(q,m)} = \sqrt{\frac{E_1E_2}{E_1+\sigma^2}}\mathbf{A}_j\mathbf{H}_{sj}^{(q)}\hat{\mathbf{x}}^{(q,m)}, \quad 1 \le q \le M_g,\ 1 \le j \le N.$$

Stack $\hat{\mathbf{c}}_{rj}^{(q,m)}$ and form $\hat{\mathbf{c}}_{rj}^{(m)}$
Cancel ICI

$$\hat{\mathbf{y}}^{(q,m+1)} = \hat{\mathbf{y}}^{(q,m+1)} - \Re\left(\mathbf{H}_{\text{eq-af}}^{(q)\,\mathcal{H}}\sum_{j=1}^{N}\left(\mathbf{\Psi}_j - \psi_j^{(0)}\mathbf{I}\right)^{[q]}\mathbf{H}_{jd}\hat{\mathbf{c}}_{rj}^{(m)}\right), \quad 1 \le q \le M_g.$$

$m = m + 1$ *goto Loop*.

Algorithm 1

Detection in frequency-flat channel in the absence of CFO

For a frequency-flat channel (i.e., $\mathbf{H}_{jd}^{(q)} = H_{jd}^{((q-1)K+1)}\mathbf{I}$) with no carrier frequency offset (i.e., $\epsilon_j = 0\ \forall j$), (14) reduces to

$$\mathbf{y}^{(q)} = \sum_{j=1}^{N} H_{jd}^{((q-1)K+1)}\mathbf{A}_j\tilde{\mathbf{x}}^{(q)} + \mathbf{z}_d^{(q)}. \tag{25}$$

Define $\mathbf{H}_{\text{eq}}^{\prime(q)} = \sum_{j=1}^{N} H_{jd}^{((q-1)K+1)}\mathbf{A}_j$. Then, by the properties of $\mathbf{A}_j$ given in [20], $\Re(\mathbf{H}_{\text{eq}}^{\prime(q)\,\mathcal{H}}\mathbf{H}_{\text{eq}}^{\prime(q)})$ is a block diagonal matrix containing 2×2 matrices as diagonal entries. Hence it is possible to do full-diversity symbol-by-symbol detection with the operation $\Re(\mathbf{H}_{\text{eq}}^{\prime(q)\,\mathcal{H}}\mathbf{y}^{(q)})$. As in AF protocol, when the channel is frequency-selective and CFOs are nonzero, this detection gives rise to ISI and ICI.

3.1. ICI and ISI in DF protocol

Now, we analyze the ICI and ISI at the output of the diversity combining operation when the relays-to-destination channels are frequency-selective and CFOs are nonzero. Define

$$\mathbf{H}_{\text{eq-df}}^{(q)} = \sum_{j=1}^{N}\psi_j^{(0)}H_{jd}^{((q-1)K+1)}\mathbf{A}_j. \tag{26}$$

Since $\psi_j^{(0)}$ is a scalar, $\Re(\mathbf{H}_{\text{eq-df}}^{(q)\,\mathcal{H}}\mathbf{H}_{\text{eq-df}}^{(q)})$ is also a block diagonal matrix. If $\mathbf{H}_{jd}^{(q)}$ matrix is split as in (21), the output of the operation $\Re(\mathbf{H}_{\text{eq-df}}^{(q)\,\mathcal{H}}\mathbf{y}^{(q)})$ on (14) can be written as

$$\begin{aligned}\hat{\mathbf{y}}^{(q)} = {} & \underbrace{\Re(\mathbf{H}_{\text{eq-df}}^{(q)\,\mathcal{H}}\mathcal{H}_{\text{eq-df}}^{(q)})\tilde{\tilde{\mathbf{x}}}^{(q)}}_{\text{Signal part}} \\ & + \underbrace{\Re(\mathbf{H}_{\text{eq-df}}^{(q)\,\mathcal{H}}\textstyle\sum_{j=1}^{N}\psi_j^{(0)}\mathbf{H}_{jd,\text{nqs}}^{(q)}\mathbf{A}_j)\tilde{\tilde{\mathbf{x}}}^{(q)}}_{\text{ISI}} \\ & + \underbrace{\Re(\mathbf{H}_{\text{eq-df}}^{(q)\,\mathcal{H}}\textstyle\sum_{j=1}^{N}\left(\mathbf{\Psi}_j - \psi_j^{(0)}\mathbf{I}\right)^{[q]}\mathbf{H}_{jd}\mathbf{c}_{rj})}_{\text{ICI due to CFOs}} \\ & + \underbrace{\Re(\mathbf{H}_{\text{eq-df}}^{(q)\,\mathcal{H}}\mathbf{z}_d^{(q)})}_{\text{Total noise}}.\end{aligned} \tag{27}$$

As in AF protocol, the optimum detector in this case would be a maximum likelihood detector in PM_g variables, which has prohibitive exponential receiver complexity.

3.2. Proposed ISI-ICI cancelling detector for DF protocol

Similar to the AF protocol, we propose a two-step PIC receiver for the DF protocol that cancels the frequency-selectivity induced ISI, and the CFO induced ICI. As can be seen, (27) identifies the desired signal, ISI, ICI, and noise components present in the output $\hat{\mathbf{y}}^{(q)}$. Based on this received signal model and the knowledge of the matrices $\mathbf{H}_{jd,\text{nqs}}^{(q)}$, $\mathbf{H}_{jd,\text{qs}}^{(q)}$, and $\mathbf{H}_{\text{eq-df}}^{(q)}$, for all q, j, we formulate the proposed interference estimation and cancellation procedure. Let $\mathbf{\Lambda}_{\text{df}}^{(q)} = \Re(\mathbf{H}_{\text{eq-df}}^{(q)\,\mathcal{H}}\mathbf{H}_{\text{eq-df}}^{(q)})$. The cancellation algorithm for the mth stage can be summarized as in Algorithm 2.

Initialization: Set $m = 1$.
Evaluate

$$\hat{\mathbf{y}}^{(q,m)} = \Re\left(\mathbf{H}_{\text{eq-df}}^{(q)\ \mathcal{H}}\mathbf{y}^{(q)}\right), \quad 1 \le q \le M_g.$$

Loop
Estimate

$$\hat{\mathbf{x}}^{(q,m)} = \left(\mathbf{\Lambda}_{\text{df}}^{(q)}\right)^{-1}\hat{\mathbf{y}}^{(q,m)}, \quad 1 \le q \le M_g.$$

Cancel ISI

$$\hat{\mathbf{y}}^{(q,m+1)} = \hat{\mathbf{y}}^{(q,1)} - \Re\left(\mathbf{H}_{\text{eq-af}}^{(q)\ \mathcal{H}}\sum_{j=1}^{N}\psi_j^{(0)}\mathbf{H}_{jd,\text{nqs}}^{(q)}\mathbf{A}_j\right)\hat{\mathbf{x}}^{(q,m)}, \quad 1 \le q \le M_g.$$

Form $\hat{\mathbf{c}}_{rj}^{(q,m)}$ *from*

$$\hat{\mathbf{c}}_{rj}^{(q,m)} = \sqrt{E_2}\mathbf{A}_j\hat{\mathbf{x}}^{(q,m)}, \quad 1 \le q \le M_g,\ 1 \le j \le N.$$

Stack $\hat{\mathbf{c}}_{rj}^{(q,m)}$ and form $\hat{\mathbf{c}}_{rj}^{(m)}$
Cancel ICI

$$\hat{\mathbf{y}}^{(q,m+1)} = \hat{\mathbf{y}}^{(q,m+1)} - \Re\left(\mathbf{H}_{\text{eq-df}}^{(q)\ \mathcal{H}}\sum_{j=1}^{N}\left(\mathbf{\Psi}_j - \psi_j^{(0)}\mathbf{I}\right)^{[q]}\mathbf{H}_{jd}\hat{\mathbf{c}}_{rj}^{(m)}\right), \quad 1 \le q \le M_g.$$

$m = m + 1$ *goto Loop*.

Algorithm 2

The order of complexity for Algorithm 2 is the same as that of the algorithm for AF protocol presented in Section 2.2.

4. SIMULATION RESULTS AND DISCUSSIONS

Simulation results for AF protocol

In this section, we evaluate the BER performance of the proposed interference cancelling receiver through simulations for the AF protocol in CO-SFBC-OFDM. For all the simulations, the total transmit power per symbol is equally divided between broadcast phase and relay phase. The noise variance at the destination is kept at unity and the transmit power per bit is varied. When there is no noise at the relays, then the transmit power per bit will be equal to the SNR per bit. We consider the following codes [23] in our simulations:

$$G_2 = \begin{pmatrix} x_1 & x_2 \\ -x_2^* & x_1^* \end{pmatrix},$$

$$G_4 = \begin{pmatrix} x_1 & x_2 & x_3 & 0 \\ -x_2^* & x_1^* & 0 & x_3 \\ -x_3^* & 0 & x_1 & x_2 \\ 0 & -x_3^* & -x_2^* & x_1^* \end{pmatrix},$$

$$G_8 = \begin{pmatrix} x_1 & x_2 & x_3 & 0 & x_4 & 0 & 0 & 0 \\ -x_2^* & x_1^* & 0 & x_3 & 0 & x_4 & 0 & 0 \\ -x_3^* & 0 & x_1 & x_2 & 0 & 0 & x_4 & 0 \\ 0 & -x_3^* & -x_2^* & x_1^* & 0 & 0 & 0 & x_4 \\ -x_4^* & 0 & 0 & 0 & x_1 & x_2 & x_3 & 0 \\ 0 & -x_4^* & 0 & 0 & -x_2^* & x_1^* & 0 & x_3 \\ 0 & 0 & -x_4^* & 0 & -x_3^* & 0 & x_1 & x_2 \\ 0 & 0 & 0 & -x_4^* & 0 & -x_3^* & -x_2^* & x_1^* \end{pmatrix}. \tag{28}$$

First, in Figure 4, we present the performance of a two-relay CO-SFBC-OFDM scheme using G_2 code. The received SNRs at all the relays are set to 35 dB. Two-ray, equal-power Rayleigh fading channel model is used for all the links. Number of subcarriers used is $M = 64$ and modulation used is 16-QAM. The CFO values at the destination for relays 1 and 2, $[\epsilon_1, \epsilon_2]$, are taken to be $[0.1, -0.08]$. We plot the BER performance of CO-SFBC-OFDM without IC and with 2 and 3 stages ($m = 2, 3$) of IC. The BER performance of noncooperative OFDM (i.e., simple point-to-point OFDM) which has the same power per transmitted bit as that of CO-SFBC-OFDM is also plotted for comparison. For CO-SFBC-OFDM, we also plot the performance of an ideal case when there is no interference, that is, when CFO = $[0, 0]$ and $L = 1$ (frequency-flat fading). From Figure 4, it can be seen that without interference cancellation, the performance of CO-SFBC-OFDM is worse than that of noncooperative OFDM. The performance improves significantly with 2 and 3 stages of cancellation, and it approaches the ideal performance of cooperation without interference. For example, at a BER of 10^{-2}, the performance improves by 12 dB with 3 stages of cancellation compared to no cancellation, and it is 0.5 dB close to the ideal performance. It can be seen that, at low SNRs, the ideal performance with cooperation is worse than that of no cooperation. This is because of the half-power split of CO-SFBC-OFDM between broadcast and relay phases. It can be observed that the slope of the BER curve of the ideal performance is steeper (2nd order diversity) than that of no cooperation (1st order diversity), and the crossover due to this diversity order difference happens at around 24 dB.

Next, in Figure 5, we repeat the same experiment (as in Figure 4) with 3 relays using G_3 code, which is obtained by deleting one column from G_4 code in (38–40). The CFO values at the destination for relays 1, 2, and 3, $[\epsilon_1, \epsilon_2, \epsilon_3]$, are taken to be $[0.1, -0.08, 0.06]$. Similar observations on the performance as in Figure 4 can be made in Figure 5 also.

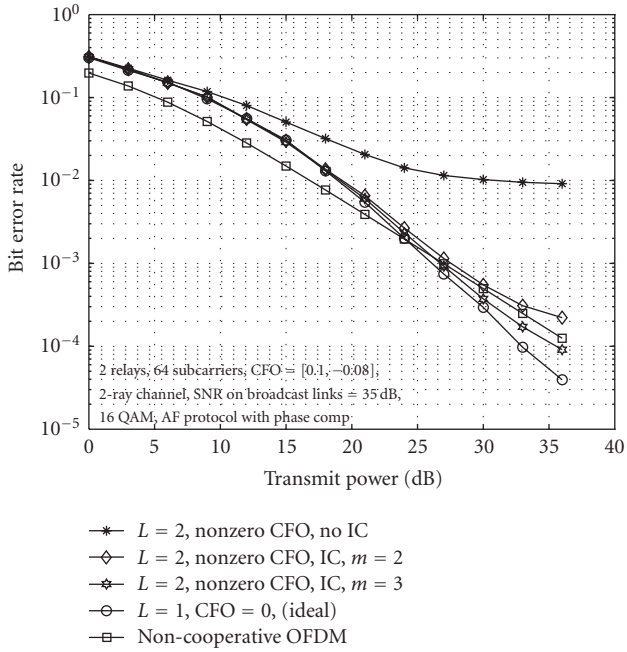


Figure 4: BER performance as a function of SNR for CO-SFBC-OFDM on frequency-selective fading ($L = 2$). $M = 64$, 2 relays ($N = 2$, G_2 code), CFO = [0.1, −0.08], 16-QAM, SNR on broadcast links = 35 dB. AF protocol and phase compensation at the relays.

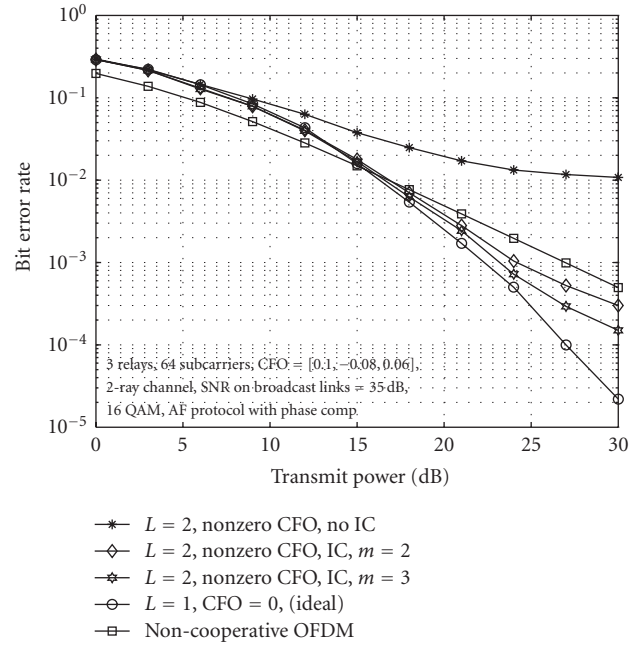


Figure 5: BER performance as a function of SNR for CO-SFBC-OFDM on frequency-selective fading ($L = 2$). $M = 64$, 3 relays ($N = 3$, G_3 code), CFO = [0.1, −0.08, 0.06], 16-QAM, SNR on broadcast links = 35 dB. AF protocol and phase compensation at the relays.

For example, at a BER of 10^{-2}, the performance of CO-SFBC-OFDM improves by over 5 dB because of interference cancellation compared to no cancellation. The difference is less compared to G_2 code because of higher-order diversity (3rd order diversity) in this case of G_3 code.

In Figure 6, we present the effect of number of relays on the performance of the interference cancellation algorithm. Codes G_2, G_3, G_4, and G_8 are used to evaluate the performance with 2, 3, 4 and 8 relays, respectively. The received SNRs at the relays are set to 45 dB. The CFOs for the different relays are [0.1, −0.08, 0.06, 0.12, −0.04, 0.02, 0.01, −0.07] and all the channels are assumed to be 2-ray, equal-power Rayleigh channels. The transmit power is kept at 18 dB per bit. The BER performance of noncooperative OFDM and no interference ($L = 1$, CFO = **0**, ideal) are also plotted. It can be observed that without IC, the performance of CO-SFBC-OFDM is worse than no cooperation and the performance improves with increasing stages of IC and approaches the ideal performance for all the cases considered. It can also be observed that performance improves with increase in number of relays, and the returns are diminishing with increase in number of relays.

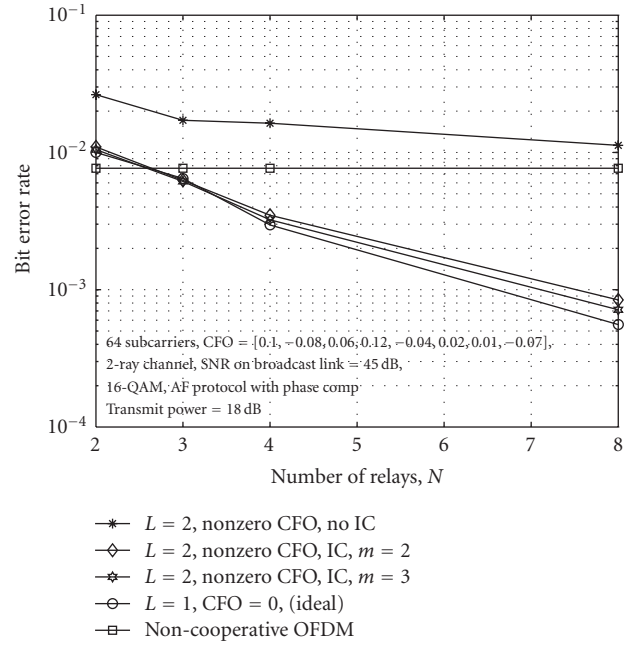


Figure 6: BER performance as a function of number of relays for CO-SFBC-OFDM on frequency-selective fading ($L = 2$). $M = 64$, Transmit power = 18 dB per bit. CFO = [0.1, −0.08, 0.06, 0.12, −0.04, 0.02, 0.01, −0.07], 16-QAM, SNR on broadcast links = 45 dB. G_2, G_3, G_4 and G_8 codes with rates 1, 3/4, 3/4 and 1/2 are used. AF protocol and phase compensation at the relays.

Simulation results for DF protocol

In Figures 7, 8, and 9, we repeat the same experiments as in Figures 4, 5, and 6, respectively, for DF protocol at the relays. For G_2 code, from Figure 7, it can be observed that the performance without IC is worse than no cooperation. The performance improves with increasing number of

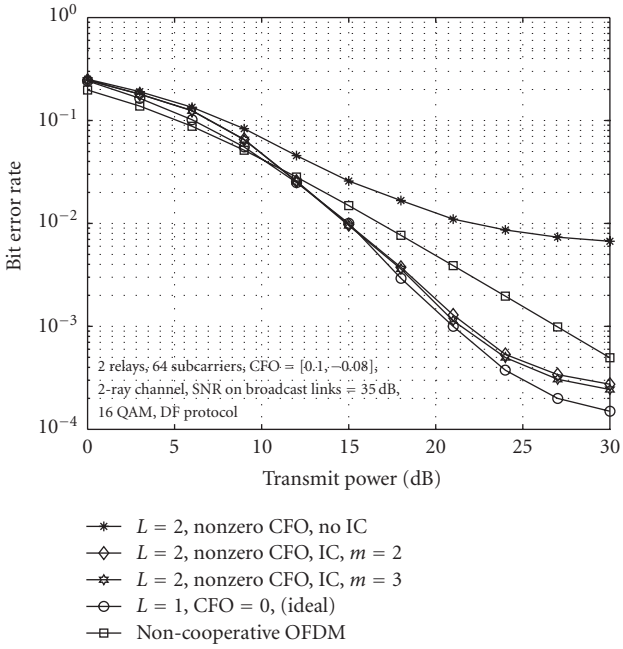


Figure 7: BER performance as a function of SNR for CO-SFBC-OFDM on frequency-selective fading ($L = 2$). $M = 64$, 2 relays ($N = 2$, G_2 code), CFO = [0.1, −0.08], 16-QAM, SNR in broadcast links = 35 dB. DF protocol at the relays.

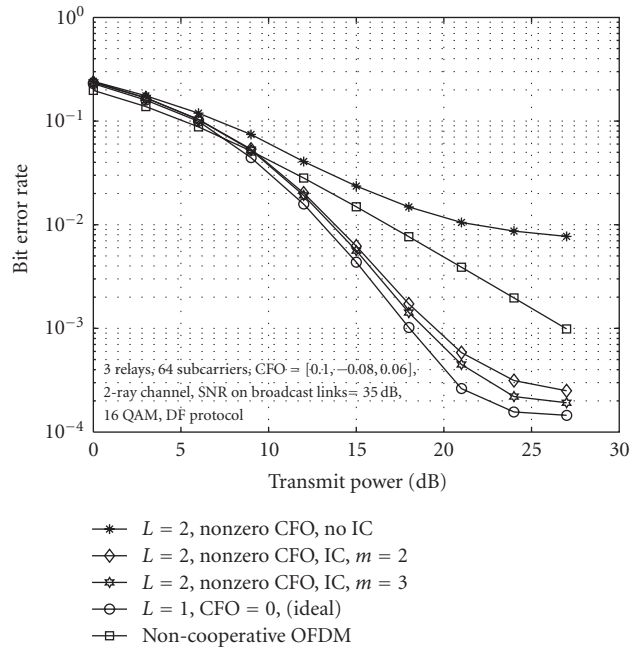


Figure 8: BER performance as a function of SNR for CO-SFBC-OFDM on frequency-selective fading ($L = 2$). $M = 64$, 3 relays ($N = 3$, G_3 code), CFO = [0.1, −0.08, 0.06], 16-QAM, SNR in broadcast links = 35 dB. DF protocol at the relays.

cancellation stages. For example, at a BER of 10^{-2}, there is a 6 dB improvement with 3 stages of cancellation. It can also be observed that crossover between CO-SFBC-OFDM (ideal) and no cooperation happens at a transmit power of 12 dB. For G_3 code also, Figure 8 shows similar performance improvement with IC. Figure 9 shows the performance plots for different number of relays using G_2, G_3, G_4, and G_8 codes. Finally, comparing the performances of AF and DF protocols, that is, Figures 4 with 7, 5 with 8, and 6 with 9, it can be observed that DF protocol has better performance compared to AF protocol for all the cases considered.

5. CONCLUSIONS

In this paper, we addressed the issue of interference (ISI and ICI due to synchronization errors and frequency selectivity of the channel) when SFBC codes are employed in cooperative OFDM systems, and proposed a low-complexity interference mitigation approach. We proposed an interference cancellation algorithm for a CO-SFBC-OFDM system with AF protocol and phase compensation at the relays. We also proposed an interference cancellation algorithm for the same system when DF protocol is used at the relays, instead of AF protocol with phase compensation. Our simulation results showed that, with the proposed algorithms, the performance of the CO-SFBC-OFDM was better than OFDM without cooperation even in the presence of carrier synchronization errors. It is also shown that DF protocol performs better than the AF protocol in these CO-SFBC-OFDM systems. The proposed IC algorithms can be extended to handle the

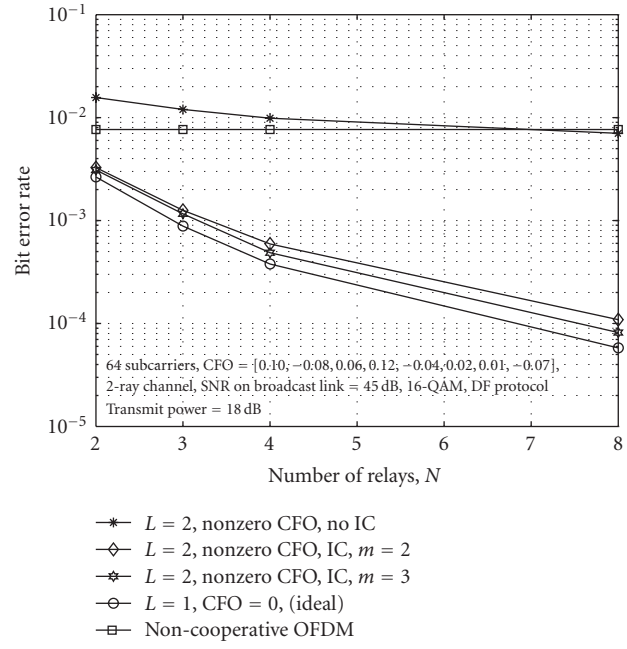


Figure 9: BER performance as a function of number of relays for CO-SFBC-OFDM on frequency-selective fading ($L = 2$). $M = 64$, at a transmit power of 18 dB per bit. CFO = [0.1, −0.08, 0.06, 0.12, −0.04, 0.02, 0.01, −0.07], 16-QAM, SNR in broadcast links = 45 dB. G_2, G_3, G_4 and G_8 codes with rates 1, 3/4, 3/4 and 1/2 are used. DF protocol is employed at the relays.

ISI effects caused due to imperfect timing on the relays-to-destination channels, that is, due to nonzero timing offsets at the destination. In the simulation results presented, the receiver is assumed to know the exact channel state information. The performance is expected to deteriorate when the receiver has only an estimated channel state information. The analysis of this deterioration and possible ways of mitigating this would be an interesting area of future work. Also, it is assumed that the relays are always available for cooperation. Algorithms to "discover" the nodes that could participate in the cooperation could also be an area of future work.

ACKNOWLEDGMENTS

This work in part was presented in the IEEE PIMRC'2007, Athens, September 2007. This work was supported in part by the Swarnajayanti Fellowship, Department of Science and Technology, New Delhi, Government of India, under Project Ref: No.6/3/2002-S.F, and the DRDO-IISc Program on Advanced Research in Mathematical Engineering.

REFERENCES

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

[2] A. Sendonaris, E. Erkip, and B. Aazhang, "User cooperation diversity—part I: system description," *IEEE Transactions on Communications*, vol. 51, no. 11, pp. 1927–1938, 2003.

[3] J. N. Laneman, D. N. C. Tse, and G. W. Wornell, "Cooperative diversity in wireless networks: efficient protocols and outage behavior," *IEEE Transactions on Information Theory*, vol. 50, no. 12, pp. 3062–3080, 2004.

[4] T. E. Hunter and A. Nosratinia, "Diversity through coded cooperation," *IEEE Transactions on Wireless Communications*, vol. 5, no. 2, pp. 283–289, 2006.

[5] J. N. Laneman and G. W. Wornell, "Distributed space-time coded protocols for exploiting cooperative diversity in wireless networks," *IEEE Transactions on Information Theory*, vol. 49, no. 10, pp. 2415–2426, 2003.

[6] S. Yiu, R. Schober, and L. Lampe, "Distributed space-time block coding," *IEEE Transactions on Communications*, vol. 54, no. 7, pp. 1195–1206, 2006.

[7] L. Yu and A. Stefanov, "Cooperative space-time coding for MIMO OFDM systems," in *Proceedings of the IEEE Military Communications Conference (MILCOM '05)*, vol. 2, pp. 990–995, Atlatnic City, NJ, USA, October 2005.

[8] O.-S. Shin, A. M. Chan, H. T. Kung, and V. Tarokh, "Design of an OFDM cooperative space-time diversity system," *IEEE Transactions on Vehicular Technology*, vol. 56, no. 4, part 2, pp. 2203–2215, 2007.

[9] F. Ng and X. Li, "Cooperative STBC-OFDM transmissions with imperfect synchronization in time and frequency," in *Proceedings of the 39th Asilomar Conference on Signals, Systems and Computers*, pp. 524–528, Pacific Grove, Calif, USA, October-November 2005.

[10] Z. Li, D. Qu, and G. Zhu, "An equalization technique for distributed STBC-OFDM system with multiple carrier frequency offsets," in *Proceedings of the IEEE Wireless Communications and Networking Conference (WCNC '06)*, vol. 2, pp. 839–843, Las Vegas, Nev, USA, April 2006.

[11] Y. Mei, Y. Hua, A. Swami, and B. Daneshrad, "Combating synchronization errors in cooperative relays," in *Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP '05)*, vol. 3, pp. 369–372, Philadelphia, Pa, USA, March 2005.

[12] E. G. Larsson and P. Stoica, *Space-Time Block Coding for Wireless Communications*, Cambridge University Press, Cambridge, UK, 2003.

[13] D. Huang and K. B. Letaief, "An interference-cancellation scheme for carrier frequency offsets correction in OFDMA systems," *IEEE Transactions on Communications*, vol. 53, no. 7, pp. 1155–1165, 2005.

[14] S. Manohar, D. Sreedhar, V. Tikiya, and A. Chockalingam, "Cancellation of multiuser interference due to carrier frequency offsets in uplink OFDMA," *IEEE Transactions on Wireless Communications*, vol. 6, no. 7, pp. 2560–2571, 2007.

[15] S. M. Alamouti, "A simple transmit diversity technique for wireless communications," *IEEE Journal on Selected Areas in Communications*, vol. 16, no. 8, pp. 1451–1458, 1998.

[16] H. Bölcskei and A. J. Paulraj, "Space-frequency coded broadband OFDM systems," in *Proceedings of the IEEE Wireless Communications and Networking Conference (WCNC '00)*, vol. 1, pp. 1–6, Chicago, Ill, USA, September 2000.

[17] Z. Liu, Y. Xin, and G. B. Giannakis, "Space-time-frequency coded OFDM over frequency-selective fading channels," *IEEE Transactions on Signal Processing*, vol. 50, no. 10, pp. 2465–2476, 2002.

[18] Y. Gong and K. B. Letaief, "An efficient space-frequency coded wideband OFDM system for wireless communications," *IEEE Transactions on Communications*, vol. 51, no. 12, pp. 2019–2029, 2003.

[19] D. Sreedhar and A. Chockalingam, "Detection of SFBC-OFDM signals in frequency- and time-selective MIMO channels," in *Proceedings of the IEEE Wireless Communications and Networking Conference (WCNC '07)*, pp. 852–857, Kowloon, Hong Kong, March 2007.

[20] D. Sreedhar, A. Chockalingam, and B. S. Rajan, "Single-symbol ML decodable distributed STBCs for partially-coherent cooperative networks," submitted to *IEEE Transactions on Information Theory*.

[21] E. Malkamäki and H. Leib, "Evaluating the performance of convolutional codes over block fading channels," *IEEE Transactions on Information Theory*, vol. 45, no. 5, pp. 1643–1646, 1999.

[22] M. Chiani, A. Conti, and O. Andrisano, "Outage evaluation for slow frequency-hopping mobile radio systems," *IEEE Transactions on Communications*, vol. 47, no. 12, pp. 1865–1874, 1999.

[23] V. Tarokh, H. Jafarkhani, and A. R. Calderbank, "Space-time block codes from orthogonal designs," *IEEE Transactions on Information Theory*, vol. 45, no. 5, pp. 1456–1467, 1999.

Hindawi Publishing Corporation
EURASIP Journal on Advances in Signal Processing
Volume 2008, Article ID 817401, 13 pages
doi:10.1155/2008/817401

Research Article

Persistent RCSMA: A MAC Protocol for a Distributed Cooperative ARQ Scheme in Wireless Networks

J. Alonso-Zárate,[1] E. Kartsakli,[2] Ch. Verikoukis,[1] and L. Alonso[2]

[1] *Centre Tecnològic de Telecomunicacions de Catalunya (CTTC), Avinguda Del Canal Olímpic S/N, Parc Mediterrani de la Tecnologia, 08860 Castelldefels, Barcelona, Spain*
[2] *Department of Signal Theory and Communications, Escola Politècnica Superior de Castelldefels (EPSC), Universitat Politècnica de Catalunya (UPC), Avinguda Del Canal Olímpic S/N, Parc Mediterrani de la Tecnologia, 08860 Castelldefels, Barcelona, Spain*

Correspondence should be addressed to J. Alonso-Zárate, jesus.alonso@cttc.es

Received 15 November 2007; Accepted 2 April 2008

Recommended by J. Wang

The persistent relay carrier sensing multiple access (PRCSMA) protocol is presented in this paper as a novel medium access control (MAC) protocol that allows for the execution of a distributed cooperative automatic retransmission request (ARQ) scheme in IEEE 802.11 wireless networks. The underlying idea of the PRCSMA protocol is to modify the basic rules of the IEEE 802.11 MAC protocol to execute a distributed cooperative ARQ scheme in wireless networks in order to enhance their performance and to extend coverage. A closed formulation of the distributed cooperative ARQ average packet transmission delay in a saturated network is derived in the paper. The analytical equations are then used to evaluate the performance of the protocol under different network configurations. Both the accuracy of the analysis and the performance evaluation of the protocol are supported and validated through computer simulations.

1. INTRODUCTION

One of the unique features of the wireless channel is its inherent broadcast nature. The air interface is a common communication channel that is shared among all the stations in a wireless network. Therefore, all the transmissions can be overheard by any station which receives enough signal strength from the transmitter. This broadcast nature poses severe challenges in the field of security, but on the other hand, opens a wide and interesting line of research targeted at exploiting all the potential benefits of those schemes that promote stations to help each other in the communications. In multiuser environments, these cooperative schemes constitute a potential alternative to overcome the practical implementation drawbacks found out when experimenting with multiple input multiple output (MIMO) techniques using relatively small devices.

The improvement induced by exploiting cooperation in wireless networks can be attained in terms of higher transmission rate, lower transmission delay, more efficient power consumption, or even increased coverage range. In the example illustrated in Figure 1, all the stations located in the transmission range of the source station (idealized in the figure with the solid circle centered at the source station) can collaborate to convey a message to a destination out of the transmission range of the transmitter. These helping stations are typically referred to as the relays.

The fundamental theory behind the concept of cooperation has been deeply studied among researchers during the last years [1–6] and now, it is currently one of the hottest topics in several engineering fields ranging from information theory to computer science. However, there is still a long way ahead in bringing to life all these theoretical concepts and developing efficient protocols that can exploit the inherent broadcast nature of wireless links to improve the performance of networks operating over the air interface. Among other open issues, the design of efficient medium access control (MAC) protocols required to manage the relay retransmissions is yet a topic of great interest.

The focus of this paper is on the design and analysis of an MAC protocol that allows executing a distributed and cooperative automatic retransmission request (ARQ) scheme

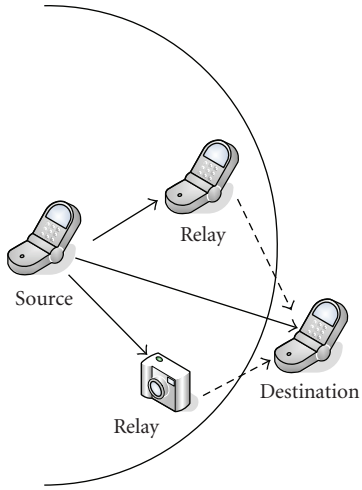


Figure 1: Cooperative scenario.

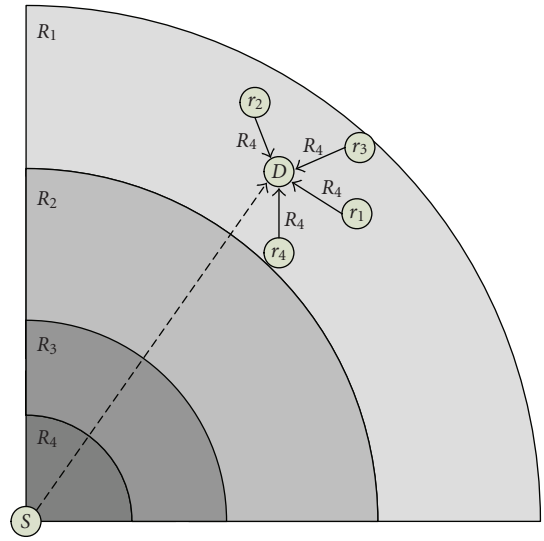


Figure 2: Example of cooperative scenario.

in wireless networks. These schemes exploit the broadcast nature of the wireless channel in the following manner; once a destination station receives a data packet containing errors, it can request a set of retransmissions from any of the relays which overheard the original transmission. retransmissions from the relays might be attained at higher transmission rates and they may allow for the exploitation of either space or time diversity. With such a distributed scheme, it is possible to improve the channel usage as well as to extend the coverage of the transmissions. Consider the example illustrated in Figure 2. It represents a multirate system, such as the IEEE 802.11 or WiMax standards, where the achievable transmission rate between any pair of source and destination stations depends, among other factors, on the signal strength at the receiver. Typically, the higher the distance between transmitter and receiver, the lower the achievable transmission rate is for a given network configuration. This allows for idealizing a scenario whereby it is possible to define different transmission rate areas surrounding any transmitting station, as illustrated in the example of Figure 2. The station S represents a source station attempting to transmit a data packet to the destination station D. There are four available transmission rates $R_4 > R_3 > R_2 > R_1$. The station D lies within the R_1 region of the station S, and thus communication will be performed at the lowest available transmission rate. This means, in turn, that a retransmission from the station S to the station D will have the highest possible cost in terms of channel time use. However, if the station D requests different retransmissions from the set of relay stations $r_1, \ldots, r_4$, with whom communications might be performed at higher data transmission rates, for example, at R_4, then, the total time required for the complete transmission process may be reduced, and thus, the channel usage increased.

Although it would be desirable to be able to tailor near-optimum protocols to get the most of the cooperative-prone nature of wireless communications, technological evolution is somehow constrained by economical drivers and the so-called imperative backwards compatibility. It is not possible to deploy completely novel devices as technology moves forward, and thus, it is of utmost interest to develop novel proposals that can at least coexist with currently available technology. With this idea in mind, a novel MAC protocol that can obtain the benefits of a distributed cooperative ARQ scheme while still using widely deployed commercial devices for wireless local area networks (WLANs) based on the standard IEEE 802.11 for WLANs [7] is presented in this paper. The new proposal is called persistent relay carrier sensing multiple access (PRCSMA) protocol, and it is based on the seminal idea outlined in [8]. In addition, the protocol is analytically modeled and a performance discussion of the protocol is also presented.

It is worth mentioning that according to the specific way, the selected relay stations handle the original signal and the way the different copies are combined to reconstruct the original message at the destination station, it is possible to classify cooperative (relay) techniques as (i) amplify and forward techniques, when the relays send an amplified version of the original message; (ii) compress and forward techniques, when the relays send a compressed version of the original transmitted signal, and (iii) decode and forward techniques, when the relays send recoded copies of the original message. Note that using decode and forward, the recoding process can be done on the basis of repeating the original codification, recoding the original data (or only a relevant part of it), or using more sophisticated space-time codification [9]. The MAC protocol presented in this paper could run on top of any of these schemes, which are, indeed, transparent to the MAC operation.

The remainder of the paper is organized as follows. A review of the current state-of-the-art in MAC protocols for distributed cooperative ARQ schemes is presented in Section 2. The IEEE 802.11 MAC protocol is outlined in Section 3, whilst Section 4 is fully devoted to the description and operational example of the PRCSMA. Section 5 presents an analytical model to calculate the average packet

transmission delay considering the cooperation scheme. System level simulations are presented in Section 6 in order to validate the accuracy of the proposed model and to evaluate the performance of the protocol under different configurations. Finally, Section 7 concludes the paper and gives some final remarks.

2. RELATED WORK

The concept of distributed cooperative ARQ has been already tackled in the past from a fundamental point of view, considering simplified network topologies, and considering ideal scheduling among the relays [10–15]. The gains of a cooperative ARQ scheme analyzed in terms of improved probability of error are discussed in [10]. In [11], the signal-to-noise ratio (SNR) gain and the average number of required retransmissions of a single source cooperative ARQ protocol are studied. In [12], the performance of different cooperative protocols is derived in terms of outage probability and SNR gain, while in [13], the saturation throughput of three double-source cooperative ARQ protocols is presented. Cerutti et al. present in [14] a delay model for single-source and single-relay cooperative ARQ protocols. In [15], Morillo-Pozo et al. propose a collaborative ARQ protocol that exploits diversity through collaboration in wireless networks. They demonstrate that when M neighboring stations collaborate using the proposed algorithm can get the same efficiency as an array of M antennas. Some other works have been focused on the relay selection criteria within the context of distributed cooperative ARQ schemes. For example, in the works presented in [16, 17], an opportunistic forwarding scheme is presented wherein the best candidate to retransmit is selected whenever a communication has failed. On the other hand, in [18, 19], a scenario wherein a set of the best candidates is selected, therein referred to as a cloud of relays, is discussed.

Previous work has put in evidence that distributed cooperative ARQ schemes may yield improved performance, lower energy consumption and interference, as well as increased coverage area by allowing communication at lower SNRs. However, up to the knowledge of the authors, there are no MAC protocols conceived to execute distributed cooperative ARQ schemes in wireless networks and to attain the achievable benefits discussed in the aforementioned research works. This is the main motivation of the work presented in this paper. The focus is on the contention process that takes place in scenarios such as the ones in [18, 19], where the relays should contend for the access to the channel.

It is worth mentioning that there exists in the literature a completely different family of cooperative MAC protocols [20–26] which have not been designed for the execution of distributed cooperative ARQ schemes in wireless networks, but they are aimed at solving other kind of interesting cooperative issues. In particular, in [20] two versions of the CoopMAC protocol are designed in the context of 802.11b WLANs in order to solve the performance anomaly problem induced by the multirate capability of the distributed coordination function (DCF) of the standard [7]. Korakis et al. implemented the protocol in actual WLAN cards, as reported in [21]. The main contribution in [21] is the description of the overall implementation process and the limitations found when attempting to implement the protocol. These limitations were mainly due to the constraints imposed by the time sensitive tasks performed by wireless cards' firmware. In addition, the CoopMAC was adapted to wireless networks using directional antennas in [22]. On the other hand, both the cooperative-MAC (CMAC) and forward error correction CMAC (FCMAC) protocols were presented in [23] within the context of 802.11e networks to improve the performance and to ensure a certain quality of service. In [24], the cooperative diversity medium access with collision avoidance (CD-MACA) protocol is proposed within the context of wireless ad hoc networks operating over the carrier sensing multiple access with collision avoidance (CSMA/CA) protocol. Although the general idea of CD-MACA is rather interesting, the definition in [24] is quite general and several implementation details are not considered. From an energy-efficient perspective, another cooperative MAC protocol is also presented within the context of ad hoc networks in [25]. This proposal integrates cooperative diversity into two different wireless routing protocols by embedding a distributed cooperative MAC. In [26], a cooperative MAC protocol was presented within the context of a mesh network formed by an access point, a number of regular stations, and one fixed wireless router (relay).

Therefore, and as aforementioned, the PRCSMA analyzed in this paper has been designed as a MAC protocol to execute a distributed cooperative ARQ scheme in wireless networks. Since it is based on the IEEE 802.11 MAC protocol, Section 3 is devoted to summarize the fundamental operational rules of the standard.

3. IEEE 802.11 DCF MAC OVERVIEW

The MAC protocol defined in the standard IEEE 802.11 for WLAN is summarized in this section. The focus has been put on the DCF, which is the one considered for ad hoc operation. Further details can be found in [7]. An example of operation of the protocol is illustrated in Figure 3.

Any station with data to transmit executes a clear channel assessment (CCA) by which it listens to the channel for a DCF interframe space (DIFS). If the channel is sensed free during this DIFS period, the station initiates the transmission of data. Otherwise, it executes a binary exponential backoff algorithm by which any station suffering a collision or a failed transmission, upon detection of the failure, sets a backoff counter at a randomized value within the interval [0,CW]. CW is referred to as the contention window, and it is initially set to a predefined value $\mathrm{CW}_{\min}$. As long as the channel is sensed idle, the backoff counter is decreased by one unit, referred to at the PHY layer as slot time and typically denoted by σ. Upon expiration of the timer, the station transmits again. In the case of failure, the CW is doubled up, up to a given maximum value $\mathrm{CW}_m = 2^m \cdot \mathrm{CW}_{\min} = \mathrm{CW}_{\mathrm{MAX}}$, and the backoff counter is reset to a random value within the interval [0,CW]. Note that m is the maximum backoff stage.

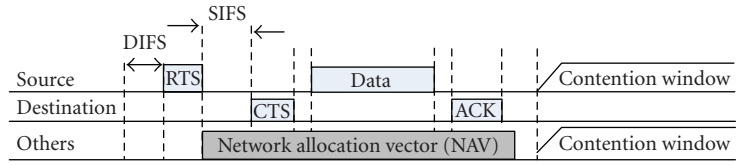


Figure 3: IEEE 802.11 MAC protocol.

Therefore, the CW can be expressed and summarized as

$$\mathrm{CW}_i = \min\{2^i \mathrm{CW}_{\min}, \mathrm{CW}_{\mathrm{MAX}}\}. \quad (1)$$

Any given packet is discarded after m' failed transmission attempts and the CW is reset to the initial value $\mathrm{CW}_{\min}$ in order to process the next packet.

Two transmission modes of operation are defined in the standard, namely, the basic access and the collision avoidance access, which is aimed at combating with the presence of the hidden and exposed terminals. In the former method, data packets are directly transmitted when trying to seize the channel, while in the later method a handshake (request-to-send (RTS)-clear-to-send (CTS)) between source and destination is established before initiating the actual data transmission. Upon the correct reception of a data packet, the destination station sends back an ACK packet after a short interframe space (SIFS). This SIFS is necessary to compensate for propagation delays and radio transceivers turn around times to switch from receiving to transmitting mode. It is worth noting that due to the fact that an SIFS is shorter than a DIFS, acknowledgments are given priority against regular data traffic.

Another relevant feature of the standard is the virtual carrier sensing mechanism, by which stations not involved in an ongoing transmission defer from attempting to transmit during the time the channel is expected to be used for an effective transmission between any pair of source and destination stations. To do so, stations update the network allocation vector (NAV) which counts for the time the channel is expected to be occupied.

4. PERSISTENT RCSMA

4.1. Protocol description

The main design goal of PRCSMA is to enable IEEE 802.11 stations to ask their neighborhood to cooperate upon the erroneous reception of a data packet. This cooperation will result in a distributed cooperative ARQ scheme. Therefore, the key objective is to modify the legacy IEEE 802.11 MAC rules to enable cooperation among the stations in a way that they could be somehow backwards compatible.

When using PRCSMA, all the stations must listen to every ongoing transmission in order to be able to cooperate if required, that is, they should operate in promiscuous mode. In addition, they should keep a copy of any received data packet (regardless of its destination address) until it is acknowledged by the destination station. It is important to note that the term destination station will be used hereafter to denote the next-hop destination of a packet, as specified by the routing protocol, which may not be the final destination station of a packet. On the other hand, the copy retained by the relays might be stored at each station data buffer or in a different dedicated queue.

Whenever a data packet is received with errors at the destination station, a cooperation phase can be initiated. The error-check could be performed by cross-checking a cyclic redundancy code (CRC) attached to the header of the packet or any other equivalent mechanism. This cooperation phase will be initiated by the intended destination station by broadcasting a claim for cooperation (CFC) message in the form of a control packet after sensing the channel idle for an SIFS period. Regular data transmissions in IEEE 802.11 are done after a longer silence period (DIFS), and thus cooperation phases are given priority over regular data traffic.

The CFC packet invites all the stations to become active relays for the communication process as long as they meet some relay selection criteria, not specified in the basic definition of PRCSMA. Different schemes for selecting a nonempty set of the most appropriate relays were discussed by the authors in [27]. It is worth mentioning that although the optimal scheme would consist in selecting the best relay for each cooperation phase, the approach in PRCSMA is to select a set of the most appropriate active relays in order to loosen the requirement of selecting exactly the best candidate in each moment [27]. An interesting open line of research will be focused on assessing the tradeoff between the costs of selecting the best relay against the time required to solve the contention among a set of selected relays.

Upon the reception of the CFC, all the stations which become active relays form the so-called relay set and get ready to forward their cooperative information. Although the specific PHY forwarding strategies applied at the relays and the reconstructing mechanism implemented at the destination station are out of the scope of the basic definition of PRCSMA, it is worth recalling that the retransmitted copy may be simply an amplified version of the original received packet at each relay, a compressed version of the received signal, a recoded version of the information, or any kind of space-time coded packet (see Section 1). For convenience, the packet transmitted by any relay will be referred to as a cooperative packet.

Accordingly, the active relays will try to get access to the channel in order to persistently transmit their cooperative packet. To do so, they will use the MAC rules specified in the IEEE 802.11 standard [7], considering the two following modifications:

(1) there is no expected ACK associated to each transmitted cooperation packet;
(2) since the subnetwork formed by the relay set works in saturation conditions, that is, all the relay stations have a data packet ready to be transmitted, it is necessary to execute a backoff mechanism at the beginning of the cooperation phase in order to avoid a certain initial collision. Therefore, those active relays which do not have an already set backoff counter (from a previous transmission attempt) set it up and initiate a random backoff period before attempting to transmit for the first time. On the other hand, those relays which already have a nonzero backoff countervalue keep the value upon the initialization of a cooperation phase.

A cooperation phase is ended whenever either the destination station is able to decode the original data packet by properly combining the different cooperative packets received from the relay set or a certain maximum cooperation timeout has elapsed. In the former case, that is, a successful cooperation phase, an ACK packet is transmitted by the destination station. In the latter case, that is, if the original packet could not be decoded, a negative ACK (NACK) is transmitted by the destination station. In any case, all the relays popout the cooperative packet from their queue upon the end of a cooperation phase.

According to all this operation, three implementation issues should be considered.

(1) The CFC can be a regular RTS packet, using the empty field for address 4, as done in [21], to distinguish the packet from a normal RTS.
(2) As long as there is at least one active relay, the persistent behavior of PRCSMA eliminates the probability that the destination station does not receive the required amount of cooperation retransmissions [27] by pretending there are infinite stations trying to cooperate.
(3) The active relays could execute either the basic access or the collision avoidance (COLAV) mode during a cooperation phase. On the one hand, as data bit rates become higher, it becomes more critical to reduce the overhead associated to the payload in order to avoid an unnecessary waste of the radio resources; therefore, it would be desirable to use the basic access mode. However, the COLAV mechanism acts as a protection mechanism against the hidden terminal problem, and thus, it will be necessary to consider the use of the RTS-CTS handshake also for the relays retransmissions in multihop networks.

4.2. Operational example

For the sake of understanding of PRCSMA, and before getting into the insights of the proposed analytical model, an example of operation of PRCSMA is presented in this section. A simple network layout with 4 stations is considered, all of them in the transmission range of each other. The basic access mode is considered, and a source station (S) transmits a data packet to destination station (D) with the support of both relays *R1* and *R2*. The cooperation phase is represented in Figure 4, and explained as follows:

(1) at instant t_1, station S sends a data packet to station D;
(2) upon reception, at instant t_2 station D broadcasts a CFC packet asking for cooperation to those stations in its neighborhood (*R1* and *R2* in this example);
(3) stations *R1* and *R2* receive the CFC packet and set up their backoff counters CW_1 and CW_2 at instant t_3;
(4) at instant t_4, the backoff counter of *R1* expires (CW_1), and *R1* attempts a cooperative transmission;
(5) at instant t_5, *R2* resumes the backoff counter while *R1* resets a new value for its backoff counter (CW_1');
(6) at instant t_6, the backoff counter of *R2* expires and *R2* attempts a cooperative transmission;
(7) at instant t_7, *R1* resumes the backoff counter and *R2* resets a new backoff counter;
(8) at instant t_8, the backoff counter of *R2* expires and *R2* attempts a cooperative transmission;
(9) at instant t_9, station D is able to properly decode the original data packet and sends back an ACK packet, indicating the end of the cooperation phase. All the stations then know that the cooperation phase has ended.

5. PRCSMA ANALYTICAL MODEL

5.1. Overview and motivation

It is always interesting to know, or at least to predict, the cost of retransmitting when executing ARQ schemes. In the distributed and cooperative scenario proposed in this paper, any destination station should assess the suitability of initiating a cooperative phase before actually sending the CFC. Therefore, accurate models to estimate the average delay associated to the distributed cooperative ARQ scheme, seen as the expected duration of the whole packet transmission time, including the cooperation phase, are required. In addition, these models may allow optimizing any given figure of merit of the system (such as the network throughput). Accordingly, an analytical model to evaluate the average distributed cooperative ARQ packet transmission delay when using PRCSMA is presented in this section.

As mentioned before, upon the reception of a CFC packet, all those stations that accomplish with some certain relay selection criteria become active relays and they will attempt to transmit their cooperative packet as many times as necessary until the cooperation phase is over. Therefore, the network will be in saturation conditions for the whole cooperative phase until the destination station is able to decode the packet. As a consequence, the relay set during a cooperation phase can be seen as a saturated network, and thus, existing analytical models for saturated IEEE 802.11 networks can be used as the foundations to develop the PRCSMA analytical model.

There exist in the literature different analytical models which develop accurate expressions of both throughput and average data packet transmission delay for IEEE 802.11 networks [28–32]. Most of them model the backoff counter

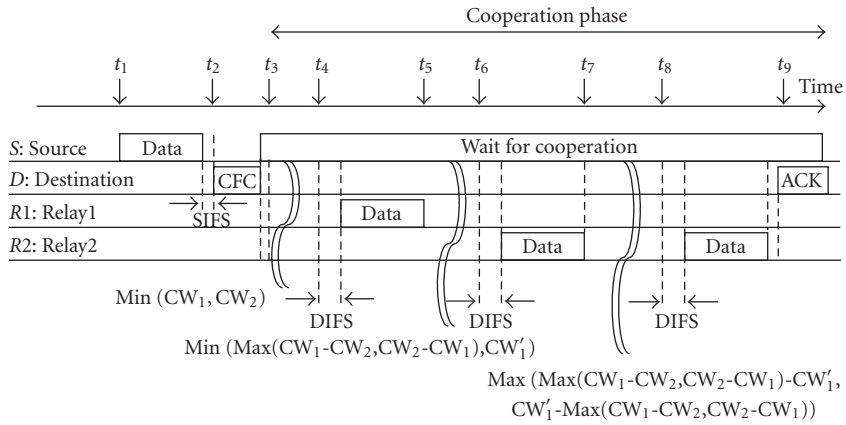


Figure 4: PRCSMA example of operation.

of an individual station with a Markov chain, and then use it to derive the overall network performance metrics. The main interest now is in deriving the average delay required to achieve a given average number of successful consecutive transmissions among all the stations forming the relay set. Therefore, although the individual station approach is useful to model PRCSMA, up to the knowledge of the authors, the derivation of this average distributed cooperative ARQ packet transmission delay, seen as the average duration of a complete data packet transmission time, including the cooperation phase that involves the relays, has not been tackled so far. This paper presents the analysis of this figure of merit for an IEEE 802.11-PRCSMA network.

5.2. Markov chain model

The backoff counter of a single PRCSMA station can be modeled using the embedded Markov chain presented and analyzed by Wu et al. in [29], and which is illustrated in Figure 5 to facilitate the understanding of the presented analysis. Each of the states represents a different value that the backoff counter of a station can take. Any pair (i, j) denotes the current value of the backoff counter i at the backoff stage j. Accordingly, the state $(0, 0)$ represents a transmission attempt. A comprehensive description of the chain can be found in [29].

A time-slotted system is considered where a total of n stations are within the transmission range of each other. A slot is defined as the unit of time between consecutive backoff counter decrements and it has a different duration depending on whether a slot is idle or busy. The main assumption of the model is that the probability of having a collision when attempting to transmit in a given time slot, p, is considered to be constant along time. $W_0 = \mathrm{CW}_{\min}$ is the size of the initial CW, m is the maximum backoff stage, and m' is the maximum number of retransmissions before discarding a packet. It is worth noting that if $m' > m$, then the backoff window will remain at the maximum stage (m) for the last $m' - m$ transmission attempts. Therefore, the probability that one station attempts to transmit in a given slot, denoted by τ, is derived in [29] as

$$\tau = \frac{1 - p^{m+1}}{1 - p} b_{0,0}, \tag{2}$$

where

$$b_{0,0} = \begin{cases} \dfrac{2(1-2p)(1-p)}{W_0(1-(2p)^{m+1})(1-p)+A}, & m \le m' \\ \dfrac{2(1-2p)(1-p)}{W_0(1-(2p)^{m'+1})(1-p)+A+B}, & m > m', \end{cases} \tag{3}$$

and $A = (1-2p)(1-p^{m+1})$ and $B = W_0 2^{m'} p^{m'+1}(1-2p)(1-p^{m-m'})$. Therefore, the probability of collision p in a given slot is equal to

$$p = 1 - (1-\tau)^{n-1}. \tag{4}$$

The probability that at least one of the n stations attempts to transmit in a given slot, P_{tr}, can be expressed as

$$P_{\mathrm{tr}} = 1 - (1-\tau)^n, \tag{5}$$

and the probability of having a successful slot given that a station transmits, p_s, is given by

$$p_s = \frac{n\tau(1-\tau)^{n-1}}{P_{\mathrm{tr}}}. \tag{6}$$

Finally, the probabilities of having an idle (P_i), successful (P_s), or collided (P_c) slot can be then written as

$$\begin{aligned} P_i &= 1 - P_{\mathrm{tr}}, \\ P_s &= P_{\mathrm{tr}} \cdot p_s = n\tau(1-\tau)^{n-1}, \\ P_c &= P_{\mathrm{tr}}(1 - p_s). \end{aligned} \tag{7}$$

Using these expressions, the average distributed cooperative ARQ packet transmission delay is analyzed in the following subsections.

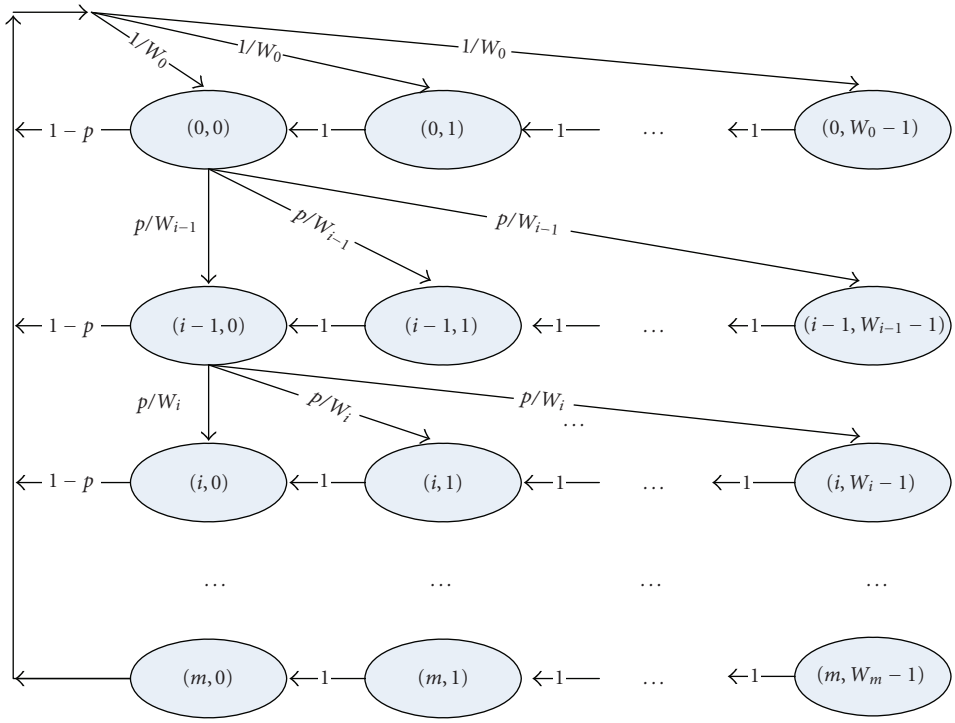


Figure 5: Wu's Markov chain to model the backoff window of the IEEE 802.11 standard.

5.3. Average distributed cooperative ARQ packet transmission delay analysis

The average distributed cooperative ARQ packet transmission delay of PRCSMA is defined as the average duration of the first failed transmission plus the average time required to complete a successful cooperation phase given an average number of retransmissions, $E[r]$, required to properly decode a packet received with errors at destination. This average delay will be denoted by $E[T_{\text{COOP}}]$. It is worth mentioning that the value of $E[r]$ will depend on (i) the channel conditions between the relays and the destination stations; (ii) the specific cooperative scheme applied at the PHY layer, and (iii) the used relay selection criteria [27]. Therefore, the value of $E[T_{\text{COOP}}]$ can be calculated as

$$E[T_{\text{COOP}}] = E[T_{\text{min}}] + E[T_{\text{cont}}], \tag{8}$$

where $E[T_{\text{min}}]$ is the expected minimum distributed cooperation ARQ packet transmission delay, which would be only achievable in the case of attaining a perfect scheduling among all the active relays, that is, avoiding contention. However, the perfect scheduling among the relays required to attain this ideal minimum average delay is impossible to attain without perfect a priori knowledge of the relays. Therefore, a contention process among the relay stations is unavoidable. This contention may lead to silence periods as well as collisions that will increase the average distributed cooperative ARQ packet transmission delay. The term $E[T_{\text{cont}}]$ will be used to denote the expected delay caused by the contention among the relays when accessing to the channel.

The term $E[T_{\text{min}}]$ can be computed as

$$E[T_{\text{min}}] = T_0 + T_{\text{CFC}} + E[r]T_{\text{DR}} + T_{\text{ACK}} + 4T_{\text{SIFS}}, \tag{9}$$

where T_0 is the duration of the first transmission from the source station to the intended destination station. T_{CFC} and T_{ACK} are, respectively, the transmission time of the CFC and the ACK packets. T_{DR} is the time required to retransmit a single packet considering that all the relay stations transmit their cooperative packets at a same common transmission rate. This value depends on whether the basic access mechanism or the collision avoidance handshake RTS/CTS is executed by the relays, and it is equal to $T_{\text{DR|BASIC}}$ or $T_{\text{DR|COLAV}}$, respectively, and calculated as

$$\begin{aligned} T_{\text{DR|BASIC}} &= T_{\text{DIFS}} + T_{\text{DATA}} + T_{\text{SIFS}}, \\ T_{\text{DR|COLAV}} &= T_{\text{DIFS}} + T_{\text{RTS}} + T_{\text{SIFS}} + T_{\text{CTS}} + T_{\text{SIFS}} \\ &\quad + T_{\text{DATA}} + T_{\text{SIFS}}, \end{aligned} \tag{10}$$

where T_{DIFS} and T_{SIFS} are, respectively, the duration of DIFS and SIFS silence periods, and T_{RTS} and T_{CTS} are the transmission times of an RTS and CTS packets. T_{DATA} is the duration of the transmission of a data packet (using the maximum available transmission rate between the relays and the destination).

On the other hand, and as long as the contention time of a packet is independent of the contention time of any other packet, which is true within the context of IEEE 802.11 [7], the value of $E[T_{\text{cont}}]$ can be calculated as

$$E[T_{\text{cont}}] = E[r]E[T_c], \tag{11}$$

where $E[T_c]$ is the average contention time required to transmit a single packet among all the relays. Therefore, the interest now is on calculating the average time elapsed between successful transmissions. This time is composed of a number of idle or collided slots of different durations, and can be derived as follows. According to the model presented in Section 5.2, a successful transmission is carried out in a given slot with a probability P_s. Therefore, the average number of slots before having a successful transmission is denoted by $E[X]$ and it can be calculated as

$$\begin{aligned} E[X] &= \sum_{k=0}^{\infty}(k+1)(1-P_s)^k P_s \\ &= P_s\left[-\frac{\partial}{\partial P_s}\sum_{k=0}^{\infty}(1-P_s)^{k+1}\right] \\ &= \frac{1}{P_s}. \end{aligned} \tag{12}$$

According to this, the average number of nonsuccessful slots before having a successful transmission is equal to $E[X]-1$. Therefore, the total contention time will be equal to

$$E[T_c] = (E[X]-1)E\left[T_{\text{slot}}\,|_{\text{non\_successful\_slot}}\right], \tag{13}$$

where $E[T_{\text{slot}}|_{\text{non\_successful\_slot}}]$ is the average duration of a slot given that the slot is not successful. A slot is not successful if it is idle or collided. As previously discussed, a given slot will be idle with probability P_i, and its duration will be equal to the basic slot time, denoted by σ. On the other hand, a given slot will suffer a collision among stations with probability P_c. As for the case of the duration of a successful transmission expressed in (10), the duration of a collision depends on whether collision avoidance is used or not, and is given in (14) as

$$\begin{aligned} T_{\text{col|BASIC}} &= T_{\text{DIFS}} + T_{\text{DATA}} + T_{\text{SIFS}}, \\ T_{\text{col|COLAV}} &= T_{\text{DIFS}} + T_{\text{RTS}} + T_{\text{SIFS}} + T_{\text{CTS\_TIMEOUT}}. \end{aligned} \tag{14}$$

The term $T_{\text{CTS\_TIMEOUT}}$ is the duration of the CTS time-out period after with a collision is considered to have occurred if no CTS packet is received by the station transmitting the corresponding RTS [7].

Applying Bayes' theorem, the average duration of any slot given that the slot is either idle or collided can be expressed as

$$E\left[T_{\text{slot}}\,|_{\text{non\_successful\_slot}}\right] = \left(\frac{P_i}{1-P_s}\right)\sigma + \left(\frac{P_c}{1-P_s}\right)T_{\text{collision}}. \tag{15}$$

Finally, the average total contention time can be rewritten as

$$\begin{aligned} &E[T_{\text{cont}}] \\ &\quad = E[r]\left(\frac{1}{P_s}-1\right)\left[\left(\frac{P_i}{1-P_s}\right)\sigma + \left(\frac{P_c}{1-P_s}\right)T_{\text{collision}}\right]. \end{aligned} \tag{16}$$

It is worth recalling that probabilities P_s, P_c, and P_i, calculated with (7), depend on the number of active relays n, the initial backoff window W_0, the maximum backoff stage m, and finally the maximum number of transmission attempts before discarding a packet m'.

6. MODEL VALIDATION AND PERFORMANCE EVALUATION

6.1. Introduction and system model

The aim of this section is twofold: first, to validate the accuracy of the model presented in Section 5 through computer simulations and, second, to evaluate the performance of the PRCSMA under different network configurations. To this end, a custom-made C++ simulator has been implemented to simulate a network formed by a total of N stations, all within the transmission range of each other, and wherein all the stations have always a packet ready to be transmitted. Note that under these saturated conditions, all the stations will always have a nonzero value of the backoff counter unless they are actually transmitting.

In order to focus on the analysis the contention problem among the relays and to avoid obscuring the performance evaluation with other system parameters, the following assumptions have been made.

(i) Original transmissions from a source station to any other destination station are always received with errors, and thus, a cooperation phase is always initiated upon the reception of an original packet. In this way, only the cooperative behavior is studied. These transmissions are performed at two constant common transmission rates, referred to as the *main control_rate* and *main data_rate*, indicating the bit rate for both the control and data plane transmissions, respectively.

(ii) Relay retransmissions are assumed to be error-free. Although this assumption may seem too restrictive, the objective is to focus on the role that the MAC plays on the performance, irrespectively of the channel conditions, assuming that they will be similar for relays close to the destination station. The parameter considered in this paper for the performance evaluation will be the average number of required retransmissions by the destination station in order to properly decode a packet originally received with errors ($E[r]$). Note that in a realistic scenario, this value will be determined by the specific cooperative scheme applied at the PHY layer, together with the actual channel conditions between the relays and the destination station. These transmissions are performed at two constant common transmission rates, referred to as the *relay control_rate* and *relay data_rate*, indicating the bit rate for both the control and data plane transmissions, respectively.

The configuration parameters of the stations in the network are summarized in Table 1, and they have been set in accordance to the orthogonal frequency division multiplex/direct sequence spread spectrum (OFDM/DSSS) PHY layer of the

Table 1: System parameters.

Parameter	Value	Parameter	Value
MAC header	34 bytes	DATA packets	1500 bytes
PHY header	96 μs	SlotTime, SIFS	10 μs
ACK, CFC	14 bytes	DIFS	50 μs
RTS	20 bytes	CTS	14 bytes

standard IEEE 802.11g [33], which allows for backwards compatibility with IEEE 802.11b stations.

6.2. Evaluation procedure

The performance evaluation presented in this paper is focused on the average distributed cooperative ARQ packet transmission delay, as defined in Section 5.3. This value has been computed in different evaluation cases by varying the following parameters:

(1) the number of active relays upon cooperation request;
(2) the transmission rates of both the main link (source-destination) and the relay transmissions (relays-destination), using the sets of rates specified in Table 2;
(3) the average number of required retransmissions upon cooperation request, $E[r]$. It is worth recalling that although this value is not a tunable parameter, and it is fixed by the network topology and conditions, it may be selected to a certain extent by appropriately selecting both the PHY cooperative scheme and the relay selection criteria, taking into account the network configuration;
(4) the access method of the relays: basic access or collision avoidance access with RTS/CTS exchange;
(5) the size of the contention windows used by the relays.

In order to study the influence of these parameters, several evaluation cases have been considered. In each case, the parameter under evaluation has been modified whereas the rest of the parameters have been kept constant and will be specified in the following subsections. They are also summarized in Table 3.

6.3. Evaluation case 1: data and control transmission rates

In order to evaluate the impact of the transmission rates on the performance of the PRCSMA, the initial CW has been set to 32 and the number of active relays (stations contending for the channel) in each cooperation phase has been set to 10. All the relay stations use the basic access method to get access to the channel.

The average distributed cooperative ARQ packet transmission delay is illustrated in both Figures 6 and 7 as a function of $E[r]$ and for different sets of transmission rates. First, it should be emphasized the almost perfect match between the analytical model and the simulations. This accuracy will be also contrasted along the other subsections of this performance evaluation.

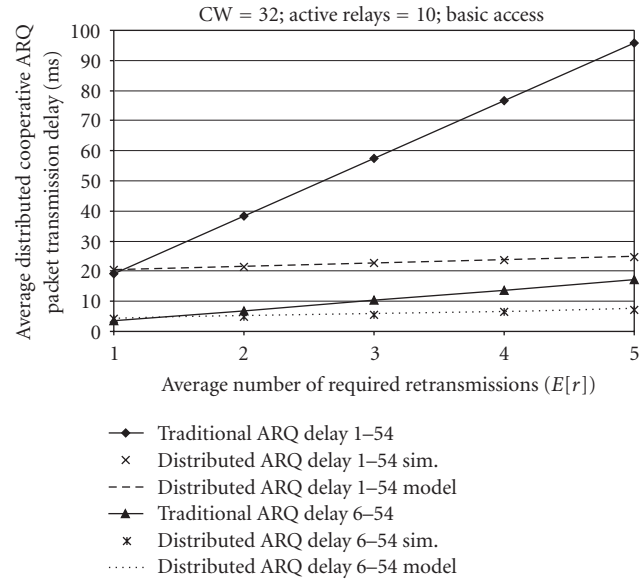


Figure 6: Average distributed cooperative ARQ packet transmission delay as a function of the transmission rate (relay low-rate regime).

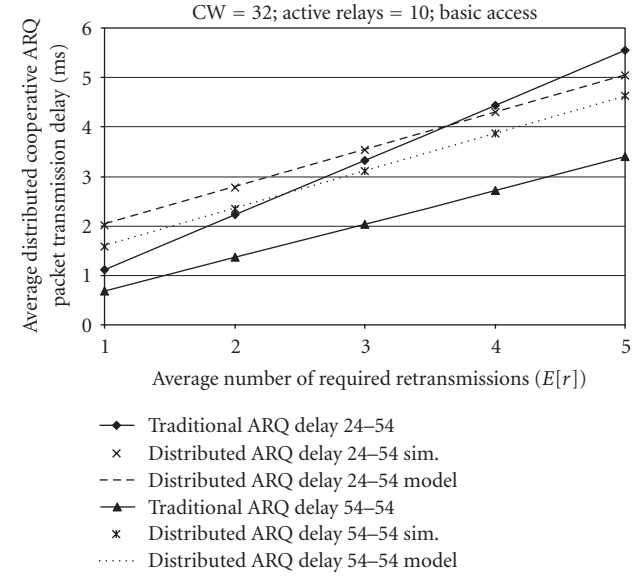


Figure 7: Average distributed cooperative ARQ packet transmission delay as a function of the transmission rates (relay high-rate regime).

As it could be expected, the ratio between the main transmission rates and the relays transmission rates determines how efficient the distributed ARQ mechanism is in comparison to the traditional noncooperative ARQ approach, where the retransmissions are only requested from the source at the best available transmission rate between the source and the intended destination station and without contention between consecutive retransmissions.

For example, in the case of using the transmission rate set 1–54 (faster relays compared to the data transmission rate

Table 2: Sets of transmission rates (Mbps).

Name	Main control_rate	Main data_rate	Relay control_rate	Relay data_rate
1–54	1	1	6	54
6–54	6	6	6	54
24–54	6	24	6	54
54–54	6	54	6	54

Table 3: Simulation parameters for each evaluation case.

Evaluation	Data/Ctr transmission rates (Mbps)	$E[r]$	Relays access method	Size of the initial CW (slots)	Number of active relays in each cooperation phase
Case 1	1–54, 6–54, 24–54, 54–54	1, 2, 3, 4, and 5	BASIC	32	10
Case 2	1–54, 6–54, 24–54, 54–54	1, 2, 3, 4, and 5	BASIC	32	10
Case 3	24–54	3	BASIC/COLAV	16	1 to 10
Case 4	24–54	3	BASIC	16, 32, 64, 128, 256, and 512	1, 5, and 10

of the main link), when $E[r]$ is 5, the distributed approach reduces the average packet transmission delay in a factor 4 compared to the traditional ARQ scheme. On the other hand, at the limit where the relay stations transmit at the same rate that the source station, the total delay in the distributed scheme is higher due to the cost of coordinating the set of relays.

It is worth mentioning that, as it could be expected, if $E[r]$ is very low, then the efficiency of the distributed ARQ scheme becomes similar to that of a traditional ARQ scheme. This is due to the fact that, despite the faster relay retransmissions, the overhead associated to the protocol does not payoff the reduction of the actual data retransmission time.

In the case of networks where the data transmission rate of each station is selected as a function of the channel state between source and destination stations, as in IEEE 802.11 WLANs, the behavior of PRCSMA shows that distributed cooperative ARQ schemes would be especially beneficial for those stations located far away, in radio-electric terms, from a transmitting station. Note that these stations will be prone to transmit at low transmission rates, and therefore, they could benefit from the faster retransmissions performed by relay stations halfway from the source station. In addition, the whole network, that is, the rest of the stations, will benefit from this scheme in the sense that faster transmissions will occupy the channel for shorter periods of time.

6.4. Evaluation case 2: average number of required retransmissions (E[r])

The same scenario as the one in Section 6.3 has been considered in this subsection.

It can be inferred from Figures 6 and 7 that the cooperative distributed ARQ packet transmission delay grows linearly with $E[r]$ in PRCSMA.

Consider a network where the relays can transmit at very high transmission rates in comparison to the main transmission link. In this scenario, the cost of increasing in one unit the value of $E[r]$ is very low in terms of delay. Therefore, it may be concluded that in this situation, it would be possible to employ simpler cooperative schemes at the PHY layer even if they may require higher values of $E[r]$ in order to properly decode an erroneous message.

However, if the transmission rates of relays are comparable to that of the main link (source-destination), then the cost of a retransmission could spoil the benefits of the distributed cooperative ARQ scheme. Therefore, the use of cooperative schemes that can reduce the value of $E[r]$ should be employed, for example, by executing more efficient cooperative schemes at the PHY layer.

6.5. Evaluation case 3: the relays access method

In this case, all the relays use an initial CW set to 16. The selected transmission rate set has been 24–54 Mbps (main-relays).

The average distributed cooperative ARQ packet transmission delay as a function of the number of active relays and for different values of $E[r]$ is depicted in Figure 8. The depicted curves represent situations where the relays use either the basic access method or the collision avoidance access. Taking into account the absence of hidden terminals in the considered scenario, it can be observed that the basic access method is always the best configuration scheme. It has to be noted that this is not an immediate conclusion; the RTS/CTS handshake mechanism does not only act as a protection mechanism against hidden terminals, but it also avoids collisions of data packets, and confines them to the control plane. However, in spite of the use of a relatively small size of the contention window compared to the number of stations contending for the channel,

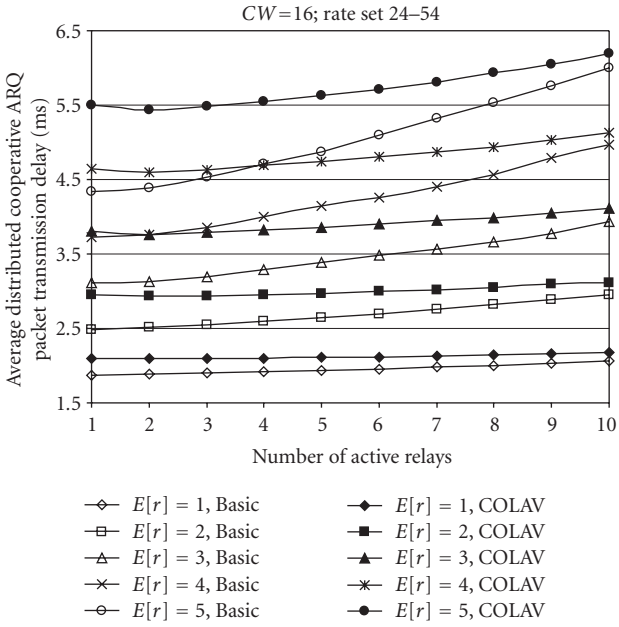


Figure 8: Performance of PRCSMA with different access methods (BASIC versus COLAV).

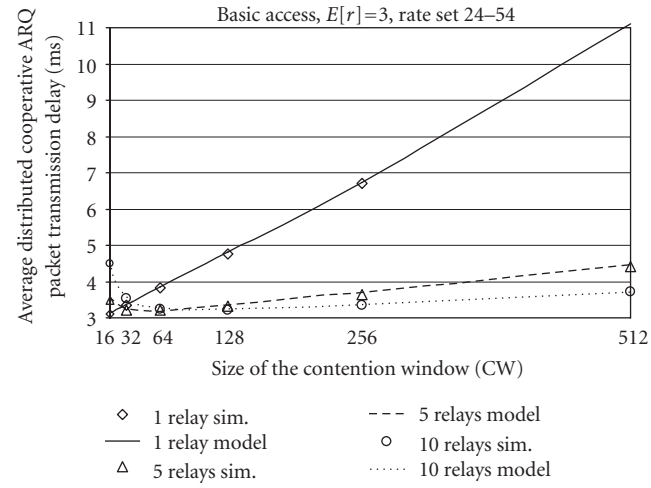


Figure 9: Average distributed cooperative ARQ packet transmission delay as a function of the size of the contention window.

which yields a high probability of collisions, the basic access method outperforms the collision avoidance in all cases. This is mainly due to the fact that the collisions in the control plane (at lower transmission rates) have a bigger cost in terms of transmission time than those in the data transmission plane (at much higher transmission rates) despite the fact that the RTS and CTS packets are shorter than data packets. Therefore, it is possible to conclude that the COLAV mechanism adds significant overhead to the communication process and compromises the benefits of the distributed cooperative ARQ scheme.

6.6. Evaluation case 4: the size of the contention window (CW)

In this case, the relay stations use the basic access mode during a cooperation phase. The average number of required retransmissions has been set to 3 and three curves represent the delay with 1, 5, or 10 active relays in each case. The transmission rate set used in these simulations is 24–54 Mbps (main-relays).

The average distributed cooperative ARQ packet transmission delay as a function of the size of the CW is illustrated in Figure 9. For the single-relay case, the average delay grows linearly with the size of the CW. Note that, the average time wasted due to the backoff will be equal to half the value of the CW, which corresponds to the expectation of the selected backoff counter. The most interesting deduction can be extracted for low values of the CW. When the size of the CW is comparable to the number of active relays, the probability of collision grows remarkably, and thus, the cooperation delay is also increased. As an example, we can see that when the size of the CW is set to 16 and the number of relays is 10, the delay is higher than when only 5 active relays are required. Therefore, the size of the CW should be properly selected as a function of the number of active relays. Higher values of the CW will lead to too much time wasted in backoff periods, while lower values of the CW will lead to increase the number of collisions. It is worth mentioning that in the case of not being able to operate at the optimum value of the CW, it would be more convenient to use higher values of the CW, since in both basic and collision avoidance access method, the cost of a collision is much higher than the cost of some extra backoff slots.

7. CONCLUSIONS

The PRCSMA protocol and its analytical performance model have been presented in this paper as an innovative solution to allow cooperative behavior in standardized IEEE 802.11 networks. By using PRCSMA, it would be possible to exploit the broadcast nature of wireless communications to save energy, to reduce interference to other systems, to increase performance and reliability of wireless communications, and to increase the range of the transmissions.

An analytical model has been derived in order to compute the delay added by the distributed cooperative ARQ scheme, which in turn, allows evaluating the overall performance of the network when using PRCSMA at the MAC layer. System level simulations have been developed to validate the accuracy of the model. In addition, a performance evaluation of the protocol has been presented in this paper, both with computer simulations and by using the model presented in the paper. The main conclusions of the presented work are that those networks where the main link between any pair of source and destination stations can use relatively lower data rates compared to those available between the active relays and the destination station constitute the best scenario where the benefits of the distributed cooperative ARQ scheme based on the IEEE 802.11 MAC protocol can be more remarkable. Moreover, the size of the contention window should be properly tuned

as a function of the number of activated relays for each cooperation phase in order to avoid either wasted time due to referral periods or existence of a high probability of collision. In any case, since collisions have a higher cost in terms of channel usage than idle periods due to unnecessary backoff deferral periods, a PRCSMA-based network should be configured with relatively high values of the contention windows compared to the average number of active relays in a cooperation phase.

Future work will be aimed at extending the analysis herein presented to multihop scenarios where the presence of hidden terminals may hamper regular communications. Another line of research will be aimed at analyzing the benefits of the proposed distributed ARQ scheme in terms of energy consumption and coverage extension.

ACKNOWLEDGMENTS

This work has been funded by the Research Projects LOOP (FIT-330215-2007-8), ETECLAS (TEC2005-07326-C02-01/TCM), PERSEO (TEC2006-10459/TCM), NEWCOM++ (IST-216715), and COOLNESS (218163-FP7-PEOPLE-2007-3-1-IAPP).

REFERENCES

[1] T. M. Cover and A. E. Gamal, "Capacity theorems for the relay channel," *IEEE Transactions on Information Theory*, vol. 25, no. 5, pp. 572–584, 1979.

[2] A. Sendonaris, E. Erkip, and B. Aazhang, "Increasing uplink capacity via user cooperation diversity," in *Proceedings of the IEEE International Symposium on Information Theory (ISIT '98)*, p. 156, Cambridge, Mass, USA, August 1998.

[3] A. Sendonaris, E. Erkip, and B. Aazhang, "User cooperation diversity—part I: system description," *IEEE Transactions on Communications*, vol. 51, no. 11, pp. 1927–1938, 2003.

[4] A. Sendonaris, E. Erkip, and B. Aazhang, "User cooperation diversity—part II: implementation aspects and performance analysis," *IEEE Transactions on Communications*, vol. 51, no. 11, pp. 1939–1948, 2003.

[5] J. N. Laneman, D. N.C. Tse, and G. W. Wornell, "Cooperative diversity in wireless networks: efficient protocols and outage behavior," *IEEE Transactions on Information Theory*, vol. 50, no. 12, pp. 3062–3080, 2004.

[6] J. N. Laneman and G. W. Wornell, "Distributed space-time-coded protocols for exploiting cooperative diversity in wireless networks," *IEEE Transactions on Information Theory*, vol. 49, no. 10, pp. 2415–2425, 2003.

[7] IEEE, Part 11: Wireless LAN Medium Access Control (MAC) and Physical Layer (PHY) Specifications, IEEE Std 802.-11-99, August 1999.

[8] J. Alonso-Zárate, J. Gómez, Ch. Verikoukis, L. Alonso, and A. I. Pérez-Neira, "Performance evaluation of a cooperative scheme for wireless networks," in *Proceedings of the 17th IEEE International Symposium on Personal, Indoor, Mobile Radio Communications Sysmposium (PIMRC '06)*, pp. 1–5, Helsinki, Finland, September 2006.

[9] F. H. P. Fitzek and M. D. Katz, Eds., *Cooperation in Wireless Networks: Principles and Applications*, Springer, New York, NY, USA, 2006.

[10] M. Dianati, X. Ling, K. Naik, and X. Shen, "A node-cooperative ARQ scheme for wireless ad hoc networks," *IEEE Transactions on Vehicular Technology*, vol. 55, no. 3, pp. 1032–1044, 2006.

[11] E. Zimmermann, P. Herhold, and G. Fettweis, "The impact of cooperation on diversity-exploiting protocols," in *Proceedings of the 59th IEEE Vehicular Technology Conference (VTC '04)*, vol. 1, pp. 410–414, Milan, Italy, May 2004.

[12] E. Zimmermann, P. Herhold, and G. Fettweis, "On the performance of cooperative relaying protocols in wireless networks," *European Transactions on Telecommunications*, vol. 16, no. 1, pp. 5–16, 2005.

[13] P. Gupta, I. Cerruti, and A. Fumagalli, "Three transmission scheduling policies for a cooperative ARQ protocol in radio networks," in *Proceedings of Wireless Networking Symposium (WNCG '04)*, Austin, Tex, USA, October 2004.

[14] I. Cerutti, A. Fumagalli, and P. Gupta, "Delay model of single-relay cooperative ARQ protocols in slotted radio networks with non-instantaneous feedback and poisson frame arrivals," in *Proceedings of the 26th IEEE International Conference on Computer Communications (INFOCOM '07)*, pp. 2276–2280, Anchorage, Alaska, USA, May 2007.

[15] J. Morillo-Pozo, J. García-Vidal, and A. I. Pérez-Neira, "Collaborative ARQ in wireless energy-constrained networks," in *Proceedings of the Joint Workshop on Foundations of Mobile Computing (DIALM-POMC '05)*, pp. 2–7, Cologne, Germany, September 2005.

[16] S. Biswas and R. Morris, "ExOR: opportunistic multi-hop routing for wireless networks," in *Proceedings of the ACM SIGCOMM Conference on Applications, Technologies, Architectures, and Protocols for Computer Communications (SIGCOMM '05)*, pp. 133–144, Philadelphia, Pa, USA, August 2005.

[17] P. Larsson and N. Johansson, "Multiuser diversity forwarding in multihop packet radio networks," in *Proceedings of the IEEE Wireless Communications and Networking Conference (WCNC '05)*, vol. 4, pp. 2188–2194, New Orleans, La, USA, March 2005.

[18] J. García-Vidal, "Addressing and Forwarding in Cooperative Wireless Networks," Tech. Rep. UPC-DAC-RR-XCSD-2005-8, Technical University of Catalonia, Barcelona, Spain, December 2005.

[19] J. García-Vidal, M. Guerrero-Zapata, J. Morillo-Pozo, and D. Fusté-Vilella, "A protocol stack for cooperative wireless networks," in *Proceedings of the 3rd International Workshop of the EURO-NGI Network of Excellence*, vol. 4396 of *Lecture Notes in Computer Science*, pp. 62–72, Sitges, Spain, June 2007.

[20] P. Liu, Z. Tao, S. Narayanan, T. Korakis, and S. Panwar, "CoopMAC: a cooperative MAC for wireless LANs," *IEEE Journal on Selected Areas in Communications*, vol. 25, no. 2, pp. 340–353, 2007.

[21] T. Korakis, S. Natayanan, A. Bagri, and S. Panwar, "Implementing a cooperative MAC protocol for wireless LANs," in *Proceedings of the IEEE International Conference on Communications (ICC '06)*, vol. 10, pp. 4805–4810, Istanbul, Turkey, June 2006.

[22] Z. Tao, T. Korakis, Y. Slutskiy, S. Panwar, and L. Tassiulas, "Cooperation and directionality: a co-opdirectional MAC for wireless ad hoc networks," in *Proceedings of the 5th International Symposium on Modeling and Optimization in Mobile, Ad Hoc, and Wireless Networks (WiOpt '07)*, Limassol, Cyprus, April 2007.

[23] N. Sai Shankar, C.-T. Chou, and M. Ghosh, "Cooperative communication MAC (CMAC)—a new MAC protocol for next generation wireless LANs," in *Proceedings of the International Conference on Wireless Networks, Communications and Mobile Computing*, vol. 1, pp. 1–6, Maui, Hawaii, USA, June 2005.

[24] X. Wang and C. Yang, "A MAC protocol supporting cooperative diversity for distributed wireless ad hoc networks," in *Proceedings of the 16th IEEE International Symposium on Personal, Indoor and Mobile Radio Communications (PIMRC '05)*, vol. 2, pp. 1396–1400, Berlin, Germany, September 2005.

[25] A. Azgin, Y. Altunbasak, and G. Alregib, "Cooperative MAC and routing protocols for wireless ad hoc networks," in *Proceedings of the IEEE Global Telecommunications Conference (GLOBECOM '05)*, vol. 5, pp. 2854–2859, St. Louis, Mo, USA, November-December 2005.

[26] A. Sadek, K. J. Ray Liu, and A. Ephremides, "Collaborative multiple-access protocols for wireless networks," in *Proceedings of the IEEE International Conference on Communications (ICC '06)*, vol. 10, pp. 4495–4500, Istanbul, Turkey, June 2006.

[27] J. Gómez, J. Alonso-Zárate, Ch. Verikoukis, A. I. Pérez-Neira, and L. Alonso, "Cooperation on demand protocols for wireless networks," in *Proceedings of the 18th IEEE International Symposium on Personal, Indoor and Mobile Radio Communications*, pp. 1–5, Athens, Greece, September 2007.

[28] G. Bianchi, "Performance analysis of the IEEE 802.11 distributed coordination function," *IEEE Journal on Selected Areas in Communications*, vol. 18, no. 3, pp. 535–547, 2000.

[29] H. Wu, Y. Peng, K. Long, S. Cheng, and J. Ma, "Performance of reliable transport protocol over IEEE 802.11 wireless LAN: analysis and enhancement," in *Proceedingsof the 21st Annual Joint Conference of the IEEE Computer and Communications Societies (INFOCOM '02)*, vol. 2, pp. 599–607, New York, NY, USA, June 2002.

[30] J. Choi, J. Yoo, and C. Kim, "A novel performance analysis model for an IEEE 802.11 wireless LAN," *IEEE Communications Letters*, vol. 10, no. 5, pp. 335–337, 2006.

[31] F. Alizadeh-Shabdiz and S. Subramaniam, "Analytical models for single-hop and multi-hop ad hoc networks," *Mobile Networks and Applications*, vol. 11, no. 1, pp. 75–90, 2006.

[32] P. Chatzimisios, A. C. Boucouvalas, and V. Vitsas, "IEEE 802.11 packet delay-a finite retry limit analysis," in *Proceedings of the IEEE Global Telecommunications Conference (GLOBECOM '03)*, vol. 2, pp. 950–954, San Francisco, Calif, USA, December 2003.

[33] IEEE Std. 802.11g, Supplement to Part 11: Wireless LAN Medium Access Control (MAC) and Physical Layer (PHY) Specifications; Further High-Speed Physical Layer Extension in the 2.4 GHz Band.

Hindawi Publishing Corporation
EURASIP Journal on Advances in Signal Processing
Volume 2008, Article ID 241378, 8 pages
doi:10.1155/2008/241378

Research Article

Optimally Joint Subcarrier Matching and Power Allocation in OFDM Multihop System

Wenyi Wang, Shefeng Yan, and Shuyuan Yang

Department of Integrated Digital System, Institute of Acoustics, Chinese Academy of Science, Beijing 100080, China

Correspondence should be addressed to Wenyi Wang, wenyi_wang@126.com

Received 29 October 2007; Accepted 6 February 2008

Recommended by Ramesh Annavajjala

Orthogonal frequency division multiplexing (OFDM) multihop system is a promising way to increase capacity and coverage. In this paper, we propose an optimally joint subcarrier matching and power allocation scheme to further maximize the total channel capacity with the constrained total system power. First, the problem is formulated as a mixed binary integer programming problem, which is prohibitive to find the global optimum in terms of complexity. Second, by making use of the equivalent channel power gain for any matched subcarrier pair, a low-complexity scheme is proposed. The optimal subcarrier matching is to match subcarriers by the order of the channel power gains. The optimal power allocation among the matched subcarrier pairs is water-filling. An analytical argument is given to prove that the two steps achieve the optimally joint subcarrier matching and power allocation. The simulation results show that the proposed scheme achieves the largest total channel capacity as compared to the other schemes, where there is no subcarrier matching or power allocation.

1. INTRODUCTION

Multihop networks have gained recently a lot of interests in the research community. By introducing relay that forwards the signal from the source to far distant destination, channel capacity can be improved and coverage area can be extended. Two main relay strategies have been identified to be usable in such scenarios: amplify-and-forward (AF) and decode-and-forward (DF). AF means that the received signal is multiplied by a parameter and then retransmitted by the relay without performing any decoding. In contrast to this, the signal is decoded at the relay and re-encoded for retransmission in the DF strategy. This has the main advantage that the transmission can be optimized for both links, separately. Furthermore, the signal is regenerated at the relay, which will not amplify the noise including the received signal. In this paper, the relay strategy is DF.

Orthogonal frequency division multiplexing (OFDM) is a mature technique to mitigate the problems of frequency of selectivity and intersymbol interference. Therefore, for the wide bandwidth multihop system, the combination of multihop system and OFDM modulation is an even more promising way to increase capacity and coverage. However, as the fading gains of different channels are mutually independent, the subcarriers which experience deep fading over the source-relay channel may not be in deep fading over the relay-destination channel. Thus, the channel capacity of a matched subcarrier pair is limited by the worse subcarrier, which will reduce the total channel capacity if the subcarriers are not matched correctly. Here, the matched subcarrier pair means that the bits transmitted on a subcarrier over the source-relay channel will be retransmitted on the other subcarrier over relay-destination channel. This motivates us to consider an adaptive subcarrier matching and power allocation scheme, where the bits transmitted on a subcarrier over the source-relay channel are possibly reallocated to another different subcarrier over the relay-destination channel.

There exist already a large number of publications on different aspects of multihop system. A fundamental analysis of cooperative relay systems was done by Kramer et al. [1], who gave channel capacities of several schemes. The system performance analysis in terms of diversity gain was done by Laneman et al. [2]. Also Sendonaris et al. [3, 4] considered the advantages in code division multiplexing access (CDMA) system using relay. Other issues that were investigated in the past were distributed space-time coding [5], selective

cooperative diversity system [6], cooperative diversity in sensor network [7, 8], and the references therein.

Relaying for OFDM systems was considered theoretically in [9]. In [10], the power allocation problem for nonregenerative OFDM relay links was investigated; in this work, the instantaneous rate is maximized for given source and relay power constraints. Multiuser OFDM relay networks were studied by Han et al. [11]. Relay selectivity in OFDM multihop system was considered by Dai et al. [12]. Bit loading algorithms for cooperative OFDM systems to minimize the system power were considered by Gui et al., where the greedy algorithm and suboptimal algorithm were proposed [13]. Kaneko et al. considered resource allocation for OFDMA system [14]. Adaptive relaying scheme for OFDM that taking channel state information at the relay node into account has been proposed in [15], where subcarrier matching was considered for OFDM amplify-and-forward scheme and power allocation was not considered. To the best of our knowledge, the optimally joint subcarrier and power allocation scheme in OFDM multihop system has not been proposed.

In this paper, we formulate the optimally joint subcarrier matching and power allocation problem as a mixed binary integer programming problem, which is NP-hard and very difficult to find global optimum. Then, by making use of the equivalent channel power gain for any matched subcarrier pair, we propose a low-complexity and optimally joint subcarrier matching and power allocation scheme, where the subcarrier matching is to match the subcarriers by the order of the channel power gains and power allocation among the matched subcarrier pairs is water-filling.

The rest of the paper is organized as follows. Section 2 presents system model used throughout the paper and formulates the problem as a mixed binary integer programming problem. Section 3 provides the optimally joint subcarrier matching and power allocation scheme for the system including only two subcarriers. The scheme is extended to the system including unlimited number of subcarriers in Section 4. Section 5 compares the capacity of the proposed scheme with those of several other schemes by simulations. Section 6 concludes the paper.

2. SYSTEM MODEL AND PROBLEM FORMULATION

2.1. System model

An OFDM multihop system is considered where the source communicates with the destination using a single relay. The relay strategy is decode-and-forward. All nodes hold one antenna. It is assumed that the destination can receive signal from the relay but not from the source because of distance or obstacle. A two-stage transmission protocol is adopted. This means that the communication between the source and the destination covers two equal time slots. The source transmits an OFDM symbol over the source-relay channel during the first time slot. At the same time, the relay receives and decodes the symbol. During the second time slot, the relay re-encodes the signal with the same codebook as the one used at the source, and transmits it towards the destination over the relay-destination channel. The destination decodes the signal based on the received signal only from the relay. The system architecture researched in this paper is shown as Figure 1. Full channel state information (CSI) is assumed. The source transmits the signal to the relay with power allocation among the subcarriers based on the algorithm of joint subcarrier matching and power allocation. The relay receives the signal and decodes the signal. Then, the relay reorders the subcarrier to match subcarrier, and allocates power among the subcarriers according to the algorithm of joint subcarrier matching and power allocation. At last, the destination decodes the signal by using the CSI over the relay-destination channel.

Throughout this paper, we assume that the different channels experience independent fading. The system consists of N subcarriers with total system power constraint. The power spectral densities of additive white Gaussian noise (AWGN) are equal at the source and the relay. The channel capacity of the subcarrier i over the source-relay channel is given:

$$R_{s,i}(P_{s,i}) = \frac{B}{2N}\log_2\left(1 + \frac{P_{s,i}h_{s,i}}{N_0 B/N}\right), \tag{1}$$

where $P_{s,i}$ is the power allocated to the subcarrier i ($1 \le i \le N$) at the source, $h_{s,i}$ is the corresponding channel power gain, N_0 is the power spectral density of AWGN, B is the total available bandwidth. Similarly, the channel capacity of the subcarrier j over the relay-destination channel is given:

$$R_{r,j}(P_{r,j}) = \frac{B}{2N}\log_2\left(1 + \frac{P_{r,j}h_{r,j}}{N_0 B/N}\right), \tag{2}$$

where $P_{r,j}$ is the power allocated to the subcarrier j ($1 \le j \le N$) at the relay, $h_{r,j}$ is the corresponding channel power gain.

When the subcarrier i over the source-relay channel is matched to the subcarrier j over the relay-destination channel, the channel capacity of this subcarrier pair is given:

$$R_{ij} = \min\{R_{s,i}(P_{s,i}), R_{r,j}(P_{r,j})\}. \tag{3}$$

2.2. Problem formulation

Theoretically, the bits transmitted at the source can be reallocated to the subcarriers at the relay in arbitrary way. But for simplification in this paper, an additional constraint is that the bits transported on a subcarrier over the source-relay channel can be reallocated to only one subcarrier over the relay-destination channel, that is, only one-to-one subcarrier matching is permitted. This means that the bits on different subcarriers over the source-relay channel will not be reallocated to the same subcarrier at the relay.

For the optimally joint subcarrier matching and power allocation problem, we can formulate it as an optimization

problem. The optimization problem is given as

$$\begin{aligned}\max_{P_{s,i},P_{r,j},\rho_{ij}} &\sum_{i=1}^{N}\min\left\{R_{s,i}(P_{s,i}),\sum_{j=1}^{N}\rho_{ij}R_{r,j}(P_{r,j})\right\}\\ &\text{subject to } \sum_{j=1}^{N}P_{s,i}+\sum_{j=1}^{N}P_{r,j}\le P_{\text{tot}},\\ &P_{s,i},P_{r,j}\ge 0\ \forall\, i,j,\\ &\rho_{ij}=\{0,1\}\ \forall\, i,j,\\ &\sum_{j=1}^{N}\rho_{ij}=1,\end{aligned}\tag{4}$$

where P_{tot} is the total system power constraint, ρ_{ij} can only be either 1 or 0, indicating whether the bits transmitted on the subcarrier i at the source are retransmitted on the subcarrier j at the relay. The last constraint shows that only one-to-one subcarrier matching is permitted. By introducing the parameter C_i, the optimization problem can be transformed into

$$\begin{aligned}\max_{P_{s,i},P_{r,j},\rho_{ij},C_i} &\sum_{i=1}^{N}C_i \quad \text{subject to } R_{s,i}(P_{s,i})\ge C_i,\\ &\sum_{j=1}^{N}\rho_{ij}R_{r,j}(P_{r,j})\ge C_i,\\ &\sum_{j=1}^{N}P_{s,i}+\sum_{j=1}^{N}P_{r,j}\le P_{\text{tot}},\\ &P_{s,i},P_{r,j}\ge 0\ \forall\, i,j,\\ &\rho_{ij}=\{0,1\}\ \forall\, i,j,\\ &\sum_{j=1}^{N}\rho_{ij}=1.\end{aligned}\tag{5}$$

That is, the original maximization problem is transformed into a mixed binary integer programming problem. It is prohibitive to find the global optimum in terms of computational complexity. However, when ρ_{ij} is given, the objective function and all constraint functions are convex, so the optimization problem is a convex optimization problem. Then the optimal power allocation can be achieved by interior-point algorithm. Therefore, the optimally joint subcarrier matching and power allocation can be found by finding the largest objective function among all subcarrier matching possibilities, and the corresponding subcarrier matching and power allocation are jointly optimal. But, it has been proved to be NP-hard and is fundamentally difficult [16]. In next section, with analytical argument, a low-complexity and optimally joint subcarrier matching and power allocation scheme is given, where the optimal subcarrier matching is to match subcarriers by the order of the channel power gains and the optimal power allocation among the subcarrier pairs is water-filling.

3. OPTIMALLY JOINT SUBCARRIER MATCHING AND POWER ALLOCATION FOR THE SYSTEM INCLUDING TWO SUBCARRIERS

Supposing that the system includes only two subcarriers ($N = 2$): the channel power gains over the source-relay channel are $h_{s,1}$ and $h_{s,2}$, and the channel power gains over the relay-destination channel are $h_{r,1}$ and $h_{r,2}$. Without loss of generality, we assume that $h_{s,1} \le h_{s,2}$ and $h_{r,1} \le h_{r,2}$. The total system power constraint is also P_{tot}. From Section 2, the optimally joint subcarrier matching and power allocation can be found by two steps: (1) for every matching possibility (i.e., ρ_{ij} is given), find the optimal power allocation and the total channel capacity; (2) compare all the total channel capacities, the largest one is the largest total channel capacity whose subcarrier matching and power allocation are joint optimally. But this process is prohibitive in terms of complexity. In this section, an analytical argument is given to prove that the optimal subcarrier is to match subcarrier by the order of the channel power gains and the optimal power allocation between the matched subcarrier pairs is water-filling. More important is that they are joint optimally.

Before giving the scheme, the equivalent channel power gain is given for any matched subcarrier pair. For any given matched subcarrier pair, with the total power constraint, an equivalent channel power gain can be given by the following proposition, whose channel capacity is equivalent to the channel capacity of this subcarrier pair.

Proposition 1. *For any given matched subcarrier pair, with total power constraint, an equivalent subcarrier channel power gain (e.g., h_i') can be given, which is related to the channel power gains (e.g., $h_{s,i}$ and $h_{r,j}$) of the subcarrier pair as follows:*

$$\frac{1}{h_i'}=\frac{1}{h_{s,i}}+\frac{1}{h_{r,j}}.\tag{6}$$

Proof. With the total power constraint P_i', the channel capacity of this subcarrier pair is

$$R_i'=\max_{P_{s,i}}\min\left\{\frac{B}{4}\log_2\left(1+\frac{P_{s,i}h_{s,i}}{N_0B/2}\right),\frac{B}{4}\log_2\left(1+\frac{(P_i'-P_{s,i})h_{r,i}}{N_0B/2}\right)\right\},\tag{7}$$

where $P_{s,i}$ is the power allocated to the subcarrier i at the source, $P_i' - P_{s,i}$ is the remainder power allocated to the subcarrier j at the relay.

The first term is a monotonically increasing function of $P_{s,i}$ and the second term is a monotonically decreasing function of $P_{s,i}$. Therefore, the optimal power allocation between the corresponding subcarriers can be gotten easily so that

$$\frac{B}{4}\log_2\left(1+\frac{P_{s,i}h_{s,i}}{N_0B/2}\right)=\frac{B}{4}\log_2\left(1+\frac{(P_i'-P_{s,i})h_{r,j}}{N_0B/2}\right),\tag{8}$$

which means that $h_{s,i}P_{s,i} = h_{r,j}(P_i' - P_{s,i})$. As a result, the channel capacity of this subcarrier pair is

$$R_i'=\frac{B}{4}\log_2\left(1+\frac{h_{s,i}h_{r,j}P_i'}{(h_{s,i}+h_{r,j})N_0B/2}\right).\tag{9}$$

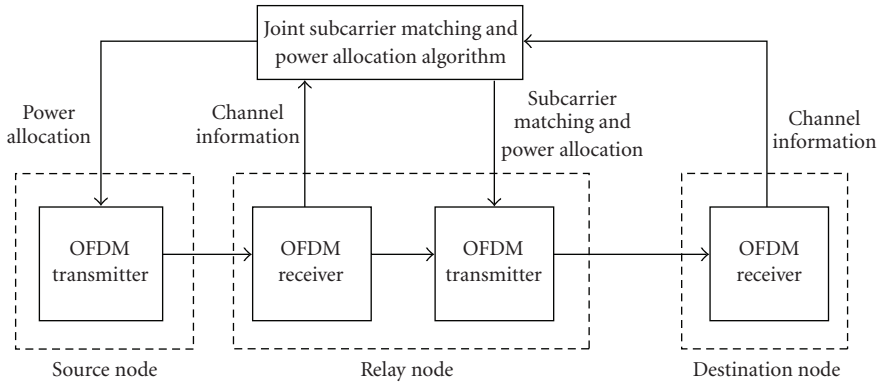


Figure 1: Block diagram of joint subcarrier matching and power allocation.

It can be seen that, by the expression of the channel capacity, the subcarrier pair can be equivalent to a single subcarrier channel with the same total power constraint. The equivalent channel power gain h_i' can be expressed:

$$h_i' = \frac{h_{s,i}h_{r,j}}{h_{s,i} + h_{r,j}}, \tag{10}$$

which can be expressed in another way:

$$\frac{1}{h_i'} = \frac{1}{h_{s,i}} + \frac{1}{h_{r,j}}. \tag{11}$$

□

Here, there are two ways to match the subcarriers: (i) the subcarrier 1 over the source-relay channel is matched to the subcarrier 1 over the relay-destination channel, and the subcarrier 2 over the source-relay channel is matched to the subcarrier 2 over the relay-destination channel (i.e., $h_{s,1}\sim h_{r,1}$ and $h_{s,2}\sim h_{r,2}$); (ii) the subcarrier 1 over the source-relay channel is matched to the subcarrier 2 over the relay-destination channel, and the subcarrier 2 over the source-relay channel is matched to the subcarrier 1 over the relay-destination channel (i.e., $h_{s,1}\sim h_{r,2}$ and $h_{s,2}\sim h_{r,1}$).

For the two ways of matching subcarriers, the equivalent channel power gains are denoted as $h_{k,i}'$ which can be gotten easily based on Proposition 1. Here, the k implies the method of matching subcarrier and the i is the index of the equivalent subcarrier. Then, the power allocation between the subcarrier pairs can be reformulation as follows:

$$\max_{P_i'} \sum_{i=1}^{2} \frac{B}{4} \log_2\left(1 + \frac{h_{k,i}'P_i'}{N_0B/2}\right) \quad \text{subject to } \sum_{i=1}^{2} P_i' \le P_{\text{tot}}, \tag{12}$$

where P_i' is the power allocated to the equivalent subcarrier i.

It is clear that the optimal power allocation is water-filling [17]. Therefore, once the subcarrier matching is provided, the optimal power allocation is given. The remainder task is to decide which way of subcarrier matching is better. The better method can be found by getting the channel capacities of the two ways and comparing them. But, here, we give an analytical argument to prove that the optimal subcarrier matching way is the first way.

Before giving the optimal subcarrier matching way, based on Proposition 1, we can get following lemma.

Lemma 1. *For the two ways of matching subcarrier, the relationship between the equivalent channel power gains can be expressed:*

$$\frac{1}{h_{1,1}'} + \frac{1}{h_{1,2}'} = \frac{1}{h_{2,1}'} + \frac{1}{h_{2,2}'}. \tag{13}$$

Proof. Based on Proposition 1, the equivalent channel power gains of the two ways can be expressed $1/h_{1,1}' = 1/h_{s,1} + 1/h_{r,1}$, $1/h_{1,2}' = 1/h_{s,2} + 1/h_{r,2}$ and $1/h_{2,1}' = 1/h_{s,1} + 1/h_{r,2}$, $1/h_{2,2}' = 1/h_{s,2} + 1/h_{r,1}$. By summing up the corresponding terms, it is clear that the relationship can be gotten. □

By making use of Lemma 1, the following proposition can be proved, which states the optimal subcarrier matching way.

Proposition 2. *For the system including two subcarriers, the optimal subcarrier matching is to match the subcarriers by the order of the channel power gains. Together with the optimal power allocation for this subcarrier matching, they are optimally joint subcarrier matching and power allocation. In this system, the optimal subcarrier matching is as $h_{s,1}\sim h_{r,1}$ and $h_{s,2}\sim h_{r,2}$.*

Proof. For the two ways of matching subcarrier, based on Lemma 1, the equivalent channel power gains satisfy the following constraint: $1/h_{k,1}' + 1/h_{k,2}' = H$ $(H \ge 0)$, where the parameter H is a constant. For the first way, we can get $1/h_{1,1}' - 1/h_{1,2}' = x_1$ $(H \ge x_1 \ge 0)$. For the second way, without loss of generality, it is assumed that $1/h_{2,1}' - 1/h_{2,2}' = x_2$ $(H \ge x_2 \ge 0)$. Therefore, the $h_{k,i}'$ can be expressed as $h_{k,1}' = 2/(H + x_k)$ and $h_{k,2}' = 2/(H - x_k)$. The corresponding total channel capacity is

$$\begin{aligned} R_{\text{tot},k}(P_1', P_2') = {} & \frac{B}{4}\log_2\left(1 + \frac{P_1'}{(H + x_k)N_0B/2}\right) \\ & + \frac{B}{4}\log_2\left(1 + \frac{P_2'}{(H - x_k)N_0B/2}\right). \end{aligned} \tag{14}$$

For denotation simplicity, we denote $N_0B/2$ as σ_2^2. The partial derivative of the channel capacity with respect to x_k can be gotten by making use of $P_2' = P_{\text{tot}} - P_1'$:

$$\frac{\partial R_{\text{tot},k}(P_1', P_2')}{\partial x_k} = \frac{B}{4\ln 2} \times \frac{(H^2\sigma_2^2 + 2H^2\sigma_2^2 x_k)(P_{\text{tot}} - P_1') + 2\sigma_2^2(P_{\text{tot}}H - P_1'x_k) + P_{\text{tot}}x_k^2\sigma_2^2}{(H^2 - x_k^2)[(H + x_k)\sigma_2^2 + P_1'][(H - x_k)\sigma_2^2 + (P_{\text{tot}} - P_1')]}. \quad (15)$$

It is clear that $\partial R_{\text{tot},k}/\partial x_k$ is greater than 0. Therefore, the total channel capacity is a monotonically increasing function of x_k for the given power allocation. This means that, for the given power allocation, the larger the difference between the equivalent channel power gains, the larger the total channel capacity. At the same time, it is clear that the difference between the equivalent channel power gains of the first way is larger than the one of the second way. Therefore, for the same power allocation, the relationship of the total channel capacities of the two ways can be expressed:

$$R_{\text{tot},2}(P_1', P_2') \le R_{\text{tot},1}(P_1', P_2'). \quad (16)$$

Therefore, we can get the following relationship:

$$\max_{P_i'} R_{\text{tot},2}(P_1', P_2') = R_{\text{tot},2}(\overline{P}_1', \overline{P}_2') \le R_{\text{tot},1}(\overline{P}_1', \overline{P}_2') \le \max_{P_i'} R_{\text{tot},1}(P_1', P_2'), \quad (17)$$

where $\overline{P}_1'$ and $\overline{P}_2'$ are the optimal power allocation for the first term. Note that the first term is the total channel capacity of the first way and the last term is the one of the second way. It proves that the first way, whose difference between the equivalent channel power gains is larger, is optimal subcarrier matching way. The more important is that, as the total channel capacity of the fisrt way is the largest one, this subcarrier matching and the corresponding power allocation are the optimally joint subcarrier matching and power allocation. Specially, the optimal subcarrier matching is to match subcarriers by the order of the channel power gains. □

The optimally joint subcarrier matching and power allocation scheme have been given by now. Specially, the optimal subcarrier matching is to match the subcarriers by the order of the channel power gains and the optimal power allocation between the matched subcarrier pairs is according to the water-filling. The power allocation between the matched subcarrier pair is to make the channel capacities of the two subcarriers equivalent.

4. EXTEND TO THE SYSTEM INCLUDING UNLIMITED NUMBER OF SUBCARRIERS

This section extends the method in Section 2 to the system including unlimited number of the subcarriers. The number of the subcarriers is finite, where the subcarrier channel power gains are $h_{s,i}$ $(i \ge 2)$ and $h_{r,j}$ $(j \ge 2)$. First, the optimal power allocation among the matched subcarrier pair is proposed for given subcarrier matching. Second, we prove that the subcarrier matching by the order of the channel power gains is optimal.

When the subcarrier matching is given, the equivalent channel gains of the subcarrier pairs can be gotten based on Proposition 1, for example, h_i' $(1 \le i \le N)$. The power allocation can be formulated as

$$\max_{P_i'} \sum_{i=1}^{N} \frac{B}{2N} \log_2\left(1 + \frac{h_i' P_i'}{\sigma_N^2}\right) \quad \text{subject to } \sum_{i=1}^{N} h_i' \le P_{\text{tot}}, \quad (18)$$

where $\sigma_N^2 = N_0B/N$. It is clear that the power allocation is also water-filling. Therefore, the optimal power allocation among the matched subcarrier pairs is according to the water-filling.

Here, without loss of generality, the channel power gains are assumed $h_{s,i} \le h_{s,i+1}$ and $h_{r,j} \le h_{r,j+1}$. The following proposition gives the optimal subcarrier matching.

Proposition 3. *For the system including unlimited number of the subcarriers, the optimal subcarrier matching is*

$$h_{s,i} \sim h_{r,i}. \quad (19)$$

Together with the optimal power allocation for this subcarrier matching, they are optimally joint subcarrier matching and power allocation

Proof. This proposition will be proved in the contrapositive form. Assuming that there is a subcarrier matching method whose matching result includes two matched subcarrier pairs $h_{s,i} \sim h_{r,i+n}$ and $h_{s,i+n} \sim h_{r,i}$ $(n > 0)$, which means that $h_{s,i} \le h_{s,i+n}, h_{r,i} \le h_{r,i+n}$, and the total capacity is larger than that of the matching method in Proposition 3.

When the power allocated to other subcarrier pairs and the other subcarrier matching are constant, the total channel capacity of this two-subcarrier pair can be improve based on Proposition 2, which implies that the channel capacity can be improved by rematching the subcarriers to $h_{s,i} \sim h_{r,i}$ and $h_{s,i+n} \sim h_{r,i+n}$. It is contrary to the assumption. Therefore, there is no subcarrier matching way better than the way in Proposition 3. At the same time, as the total capacity of this subcarrier matching and the corresponding optimal power allocation scheme is the largest, this subcarrier matching together with the corresponding optimal power allocation are the optimally joint subcarrier matching and power allocation. □

For the system including unlimited number of the subcarriers, the optimally joint subcarrier matching and power allocation scheme has been given by now. Here, the steps are summarized as follow

Step 1. Sort the subcarriers at the source and the relay in ascending order by the permutations π and π', respectively. The process is according to the channel power gains, that is, $h_{s,\pi(i)} \le h_{s,\pi(i+1)}$, $h_{r,\pi'(i)} \le h_{r,\pi'(i+1)}$.

Step 2. Match the subcarriers into pairs by the order of the channel power gains (i.e., $h_{s,\pi(i)} \sim h_{r,\pi'(i)}$), which means that the bits transported on the subcarrier $\pi(i)$ over the source-relay channel will be retransmitted on the subcarrier $\pi'(i)$ over the relay-destination channel.

Step 3. Based on Proposition 1, get the equivalent channel power gain $h'_{\pi(i)}$ according to the matched subcarrier pair, that is, $h'_{\pi(i)} = (h_{s,\pi(i)} h_{r,\pi'(i)})/(h_{s,\pi(i)} + h_{r,\pi'(i)})$.

Step 4. For the equivalent channel power gains, the power allocation is water-filling as follows:

$$P'_{\pi(i)} = \left(\frac{h'_{\pi(i)} B}{2N\lambda \ln 2} - \frac{h'_{\pi(i)}}{\sigma_N^2} \right)^+, \tag{20}$$

where $(a)^+ = \max(a, 0)$ and λ can be found by the following equation:

$$\sum_{i=1}^{N} P'_{\pi(i)} = P_{\text{tot}}. \tag{21}$$

The power allocation between the subcarriers in the given matched subcarrier pair is as follows:

$$\begin{aligned} P_{s,\pi(i)} &= \frac{h_{r,\pi'(i)} P'_{\pi(i)}}{h_{s,\pi(i)} + h_{r,\pi'(i)}}, \\ P_{r,\pi'(i)} &= \frac{h_{s,\pi(i)} P'_{\pi(i)}}{h_{s,\pi(i)} + h_{r,\pi'(i)}}. \end{aligned} \tag{22}$$

Step 5. The total system channel capacity is

$$R_{\text{tot}} = \frac{B}{2N} \sum_{i=1}^{N} \log_2 \left(1 + \frac{h'_{\pi(i)} P'_{\pi(i)}}{\sigma_N^2} \right). \tag{23}$$

5. NUMERICAL RESULTS

In this section, we compare the total channel capacity of the proposed scheme with those of several other schemes by computer simulations.

These schemes include:

(i) no subcarrier matching and no power allocation: the bits transmitted on the subcarrier i at the source will be retransmitted on the subcarrier i at the relay; the power is allocated equally over all subcarriers at the source and the relay (i.e., $P_{s,i} = P_{r,j} = P_{\text{tot}}/2N$);

(ii) optimal power allocation and no subcarrier matching: the bits transmitted on the subcarrier i at the source will be retransmitted on the subcarrier i at the relay; the power allocation is according to water-filling among the subcarrier pairs;

(iii) subcarrier matching and no power allocation: the bits transmitted on the subcarrier $\pi(i)$ at the source will be retransmitted on the subcarrier $\pi'(i)$ at the relay; the power is allocated equally over all subcarriers at the source and the relay (i.e., $P_{s,i} = P_{r,j} = P_{\text{tot}}/2N$).

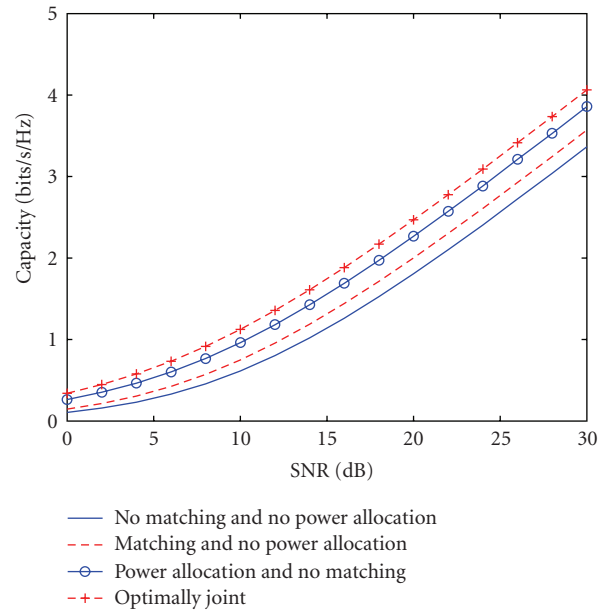


FIGURE 2: Channel capacity versus SNR ($N = 32$).

Here, the subcarrier matching is the same as in Steps 1 and 2 in Section 4 and the power allocation means that the water-filling algorithm is performed among the subcarrier pairs. In the computer simulations, we assume that each subcarrier undergoes Rayleigh fading independently. The total bandwidth is $B = 1$ MHz. The SNR is defined as SNR $= P_{\text{tot}}/N_0 B$. To obtain the average data rate, we have simulated 10 000 independent trials.

Figure 2 shows the total channel capacity versus SNR, where the number of the subcarriers is constant (e.g., $N = 32$) and the average channel power gains are assumed to be one, that is, $E(h_{s,i}) = 1$ and $E(h_{r,j}) = 1$ for all i and j. The capacity of the scheme (i), where there is no subcarrier matching and no power allocation, is the least one compared with that of the other schemes. The capacity of the optimally joint subcarrier and power allocation scheme is the largest one than those of all other schemes. If other conditions remain unchanged, both subcarrier matching and power allocation can improve the total channel capacity. Specially, subcarrier matching can improve the capacity when comparing the capacity of the scheme (i) to that of the scheme (iii). The system capacity can be improved by power allocation when comparing the capacity of the scheme (i) to that of scheme (ii). Another important result is that power allocation is more effective than subcarrier matching, when only one of the two ways can be applied.

The relationship between the total channel capacity and the number of the subcarriers is shown in Figure 3, where the SNR is constant, for example, SNR $= 20$ dB. The average channel power gains are also assumed to be one, that is, $E(h_{s,i}) = 1$ and $E(h_{r,j}) = 1$ for all i and j. Almost the same conclusions about the comparison among all the schemes can be gotten from Figure 3 as those from Figure 2. It is noted that the total channel capacity is almost constant with the

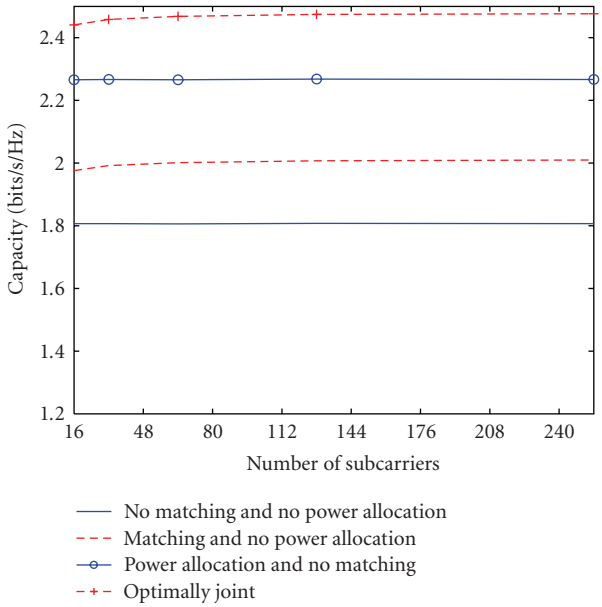


FIGURE 3: Channel capacity versus the number of the subcarriers (SNR = 20 dB).

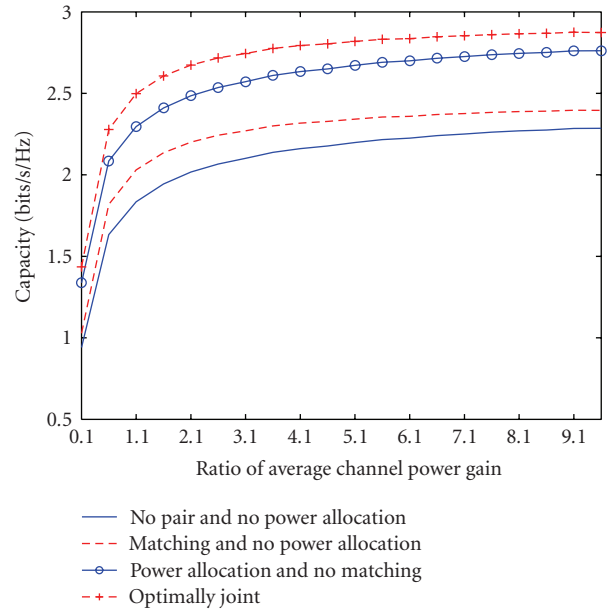


FIGURE 4: Channel capacity versus the ratio of the channel power gains ($E(h_{s,i}) = 1$, SNR = 20 dB).

growth of the number of the subcarriers because the total power constraint is constant.

Figure 4 shows the total channel capacity versus the ratio of the average channel power gains over the relay-destination channel to the ones over the source-relay channel, that is, $E(h_{r,i})/E(h_{s,i})$, where the average channel power gains over the source-relay channel are assumed to be one, that is, $E(h_{s,i}) = 1$. Again the same conclusions about the comparison among all the schemes can be gotten from Figure 4 as those from Figure 2. The total channel capacities increase very quickly with the ratio increasing from 0.1 to 1, this is because of the total channel capacities are limited by the channel capacities over the relay-destination channels in this interval. The total channel capacities increase slowly with the ratio increasing from 1 to 10 because the total channel capacities are limited by the channel capacities over the source-relay channels in this interval.

6. CONCLUSION

For the OFDM multihop system, as the fading gains of different channels are independent, subcarrier matching is a promising way to further improve capacity. Here, subcarrier matching means that the bits on a subcarrier over the source-relay channel are possibly reallocated to another different subcarrier over the relay-destination channel. In this paper, we propose an optimally joint subcarrier matching and power allocation scheme to maximize channel capacity, where the relay is based on the decode-and-forward and the total system power is constrained. Though the problem can be formulated as a mixed binary integer programming problem, it is NP-hard and prohibitive to find the global optimum. A low-complexity scheme is proposed, which is still jointly optimal. First, for any matched subcarrier pair, an equivalent channel power gain is proposed. Then, for the system including only two subcarriers, the optimally joint subcarrier matching and power allocation can be gotten by matching the subcarriers by the order of the channel power gains and allocating power according to water-filling between the two subcarrier pairs. Second, the scheme of optimally joint subcarrier matching and power allocation is extended to the system including unlimited number of the subcarriers. The analytical argument proves that the scheme also gives optimally joint subcarrier matching and power allocation. The simulation results prove that the proposed scheme achieves the largest total channel capacity as compared to the other schemes, where there is no subcarrier or no power allocation.

REFERENCES

[1] G. Kramer, M. Gastpar, and P. Gupta, "Cooperative strategies and capacity theorems for relay networks," *IEEE Transactions on Information Theory*, vol. 51, no. 9, pp. 3037–3063, 2005.

[2] J. N. Laneman, D. N. C. Tse, and G. W. Wornell, "Cooperative diversity in wireless networks: efficient protocols and outage behavior," *IEEE Transactions on Information Theory*, vol. 50, no. 12, pp. 3062–3080, 2004.

[3] A. Sendonaris, E. Erkip, and B. Aazhang, "User cooperation diversity. Part I. System description," *IEEE Transactions on Communications*, vol. 51, no. 11, pp. 1927–1938, 2003.

[4] A. Sendonaris, E. Erkip, and B. Aazhang, "User cooperation diversity. Part II. Implementation aspects and performance analysis," *IEEE Transactions on Communications*, vol. 51, no. 11, pp. 1939–1948, 2003.

[5] J. N. Laneman and G. W. Wornell, "Distributed space-time-coded protocols for exploiting cooperative diversity in wireless networks," *IEEE Transactions on Information Theory*, vol. 49, no. 10, pp. 2415–2425, 2003.

[6] A. Bletsas, A. Khisti, D. P. Reed, and A. Lippman, "A simple cooperative diversity method based on network path selection," *IEEE Journal on Selected Areas in Communications*, vol. 24, no. 3, pp. 659–672, 2006.

[7] N. Shastry, J. Bhatia, and R. S. Adve, "Theoretical analysis of cooperative diversity in wireless sensor networks," in *Proceedings of the IEEE Global Telecommunications Conference (GLOBECOM '05)*, vol. 6, pp. 3269–3273, St. Louis, Mo, USA, November-December 2005.

[8] S. Serbetli and A. Yener, "Power allocation and hybrid relaying strategies for F/TDMA ad hoc networks," in *Proceedings of the IEEE International Conference on Communications (ICC '06)*, vol. 4, pp. 1562–1567, Istanbul, Turkey, June 2006.

[9] G. Li and H. Liu, "On the capacity of broadband relay networks," in *Proceedings of the 38th Asilomar Conference on Signals, Systems and Computers (ACSSC '04)*, vol. 2, pp. 1318–1322, Pacific Grove, Calif, USA, November 2004.

[10] I. Hammerstrom and A. Wittneben, "On the optimal power allocation for nonregenerative OFDM relay links," in *Proceedings of the IEEE International Conference on Communications (ICC '06)*, vol. 10, pp. 4463–4468, Istanbul, Turkey, June 2006.

[11] Z. Han, T. Himsoon, W. P. Siriwongpairat, and K. J. R. Liu, "Energy-efficient cooperative transmission over multiuser OFDM networks: who helps whom and how to cooperate," in *Proceedings of the IEEE Wireless Communications and Networking Conference (WCNC '05)*, vol. 2, pp. 1030–1035, New Orleans, La, USA, March 2005.

[12] L. Dai, B. Gui, and L. J. Cimini Jr., "Selective relaying in OFDM multihop cooperative networks," in *Proceedings of the IEEE Wireless Communications and Networking Conference (WCNC '07)*, pp. 963–968, Kowloon, China, March 2007.

[13] B. Gui and L. J. Cimini Jr., "Bit loading algorithms for cooperative OFDM systems," *EURASIP Journal on Wireless Communications and Networking*, vol. 2008, Article ID 476797, 9 pages, 2008.

[14] M. Kaneko and P. Popovski, "Radio resource allocation algorithm for relay-aided cellular OFDMA system," in *Proceedings of the IEEE International Conference on Communications (ICC '07)*, pp. 4831–4836, Glasgow, Scotland, UK, June 2007.

[15] M. Herdin, "A chunk based OFDM amplify-and-forward relaying scheme for 4G mobile radio systems," in *Proceedings of the IEEE International Conference on Communications (ICC '06)*, vol. 10, pp. 4507–4512, Istanbul, Turkey, June 2006.

[16] B. Korte and J. Vygen, *Combinatorial Optimization: Theory and Algorithms*, Springer, New York, NY, USA, 3rd edition, 2002.

[17] T. Cover and J. Thomas, *Elements of Information Theory*, John Wiley & Sons, New York, NY, USA, 1991.

Hindawi Publishing Corporation
EURASIP Journal on Advances in Signal Processing
Volume 2008, Article ID 476125, 15 pages
doi:10.1155/2008/476125

Research Article

Power and Resource Allocation for Orthogonal Multiple Access Relay Systems

Wessam Mesbah and Timothy N. Davidson

Department of Electrical and Computer Engineering, McMaster University, Hamilton, ON, Canada L8S 4K1

Correspondence should be addressed to Timothy N. Davidson, davidson@mcmaster.ca

Received 1 November 2007; Revised 19 April 2008; Accepted 6 May 2008

Recommended by J. Wang

We study the problem of joint power and channel resource allocation for orthogonal multiple access relay (MAR) systems in order to maximize the achievable rate region. Four relaying strategies are considered; namely, regenerative decode-and-forward (RDF), nonregenerative decode-and-forward (NDF), amplify-and-forward (AF), and compress-and-forward (CF). For RDF and NDF we show that the problem can be formulated as a quasiconvex problem, while for AF and CF we show that the problem can be made quasiconvex if the signal-to-noise ratios of the direct channels are at least -3 dB. Therefore, efficient algorithms can be used to obtain the jointly optimal power and channel resource allocation. Furthermore, we show that the convex subproblems in those algorithms admit a closed-form solution. Our numerical results show that the joint allocation of power and the channel resource achieves significantly larger achievable rate regions than those achieved by power allocation alone with fixed channel resource allocation. We also demonstrate that assigning different relaying strategies to different users together with the joint allocation of power and the channel resources can further enlarge the achievable rate region.

1. INTRODUCTION

In multiple access relay (MAR) systems, several source nodes send independent messages to a destination node with the assistance of a relay node [1–4]. These systems are of interest because they offer the potential for reliable communication at rates higher than those provided by conventional [5] and cooperative multiple access schemes [6–8] (in which source nodes essentially work as relays for each other.) For example, in [4] a comparison was made between the MAR system and the system that employs user cooperation, and the MAR system was shown to outperform the user cooperation system. Full-duplex MAR systems, in which the relay is able to transmit and receive simultaneously over the same channel, were studied in [1–3, 9], where inner and outer bounds for the capacity region were provided. However, full-duplex relays can be difficult to implement due to the electrical isolation required between the transmitter and receiver circuits. As a result, half-duplex relays, which do not simultaneously transmit and receive on the same channel, are often preferred in practice. The receiver at the relay and destination nodes can be further simplified if the source nodes (and the relay) transmit their messages on orthogonal channels, as this enables "per-user" decoding rather than joint decoding.

In this paper, we will consider the design of orthogonal multiple access systems with a half-duplex relay. In particular, we will consider the joint allocation of power and the channel resource in order to maximize the achievable rate region. Four relaying strategies will be considered; namely, regenerative (RDF) and nonregenerative (NDF) decode-and-forward [6, 8], amplify-and-forward (AF) [8, 10], and compress-and-forward (CF) [11, 12]. The orthogonal half-duplex MAR system that we will consider is similar to that considered in [13]. However, the focus of that paper is on the maximization of the sum rate, and, more importantly, it is assumed therein that the source nodes will each be allocated an equal fraction of the channel resources (e.g., time or bandwidth).(This equal allocation of resources is only optimal in the sum rate sense when the source nodes experience equal effective channel gains towards the destination and equal effective channel gains towards the

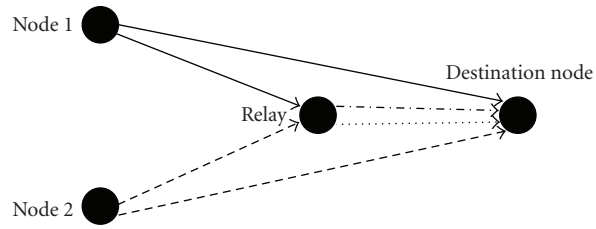


Figure 1: A simple multiple access relay channel with two source nodes.

relay.) In this paper, we will provide an efficiently solvable formulation for finding the jointly optimal allocation of power and the channel resources that enables the system to operate at each point on the boundary of the achievable rate region.

Although the problem of power allocation for an equal allocation of the channel resource was shown to be convex in [13], the joint allocation of power and the channel resource is not convex, which renders the problem harder to solve. In this paper, we show that the joint allocation problem can be formulated in a quasiconvex form, and hence, that the optimal solution can be obtained efficiently using standard quasiconvex algorithms, for example, bisection-based methods [14]. Furthermore, for a given channel resource allocation, we obtain closed-form expressions for the optimal power allocation, which further reduces the complexity of the algorithm used to obtain the jointly optimal allocation.

The practical importance of solving the problem of the joint allocation of power and channel resources is that it typically provides a substantially larger achievable rate region than that provided by allocating only the power for equal (or fixed) channel resource allocation, as will be demonstrated in the numerical results. Those results will also demonstrate the superiority of the NDF and CF relaying strategies over the RDF and AF strategies, respectively, which is an observation that is consistent with an observation in [13] for the case of power allocation with equal resource allocation. We will also demonstrate that joint allocation of the relaying strategy together with the power and channel resources, rather assigning the same relaying strategy to all users, can further enlarge the achievable rate region.

2. SYSTEM MODEL

We consider an orthogonal multiple access relay (MAR) system with N source nodes (nodes $1, 2, \ldots, N$), one destination node (node 0), and one relay (node R) that assists the source nodes in the transmission of their messages to the destination node. (The generalization of our model to different destination nodes is direct.) Figure 1 shows a simplified two-source MAR system. We will focus here on a system in which the transmitting nodes use orthogonal subchannels to transmit their signals, and the relay operates in half-duplex mode. This system model is similar to that used in [13]. The orthogonal subchannels can be synthesized in time or in frequency, but given their equivalence it is sufficient for us to focus on the case in which they are synthesized in time, that is, we will divide the total frame length into N nonoverlapping subframes of fractional length r_i, and we will allocate the ith subframe to the transmission (and relaying) of the message from source node i to the destination node. Figure 2 shows the block diagram of the cooperation scheme and the transmitted signals during one frame of such an MAR system with two source nodes. As shown in Figure 2, the first subframe is allocated to node 1 and has a fractional length r_1, while the second subframe is allocated to node 2 and has a fractional length $r_2 = 1 - r_1$. Each subframe is further partitioned into two equal-length blocks [13]. In the first block of subframe i of frame ℓ, node i sends a new block of symbols $B_i(w_{i\ell})$ to both the relay and the destination nodes, where $w_{i\ell}$ is the component of the ith user's message that is to be transmitted in the ℓth frame. In the second block of that subframe, the relay node transmits a function $f(\cdot)$ of the message it received from node i in the first block. (The actual function depends on the relaying strategy.) We will let P_i represent the power used by node i to transmit its message, and we will constrain it so that it satisfies the average power constraint $(r_i/2)P_i \leq \overline{P}_i$, where $\overline{P}_i$ is the maximum average power of node i. We will let P_{Ri} represent the relay power allocated to the transmission of the message of node i, and we will impose the average power constraint $\sum_{i=1}^{N}(r_i/2)P_{Ri} \leq \overline{P}_R$. (The function $f(\cdot)$ is normalized so that it has a unit power.) In this paper, we consider the following four relaying strategies.

(i) Regenerative decode-and-forward (RDF). The relay decodes the message $w_{i\ell}$, re-encodes it using the same code book as the source node, and transmits the codeword to the destination [6, 8].

(ii) Nonregenerative decode-and-forward (NDF). The relay decodes the message $w_{i\ell}$, re-encodes it using a different code book from that used by the source node, and transmits the codeword to the destination [15, 16].

(iii) Amplify-and-forward (AF). The relay amplifies the received signal and forwards it to the destination [8, 10]. In this case, $f(w_{i\ell})$ is the signal received by the relay, normalized by its power.

(iv) Compress-and-forward (CF). The relay transmits a compressed version of the signal it receives [11, 12].

Without loss of generality, we will focus here on a two-user system in order to simplify the exposition. However, as we will explain in Section 3.5, all the results of this paper can be applied to systems with more than two source nodes. For the two-source system, the received signals at the relay and the destination at block m can be expressed as

$$\mathbf{y}_R(m) = \begin{cases} K_{1R}\mathbf{x}_1(m) + \mathbf{z}_R(m) & m \bmod 4 = 1, \\ K_{2R}\mathbf{x}_2(m) + \mathbf{z}_R(m) & m \bmod 4 = 3, \\ \mathbf{0} & m \bmod 4 \in \{0, 2\}, \end{cases}$$
$$\mathbf{y}_0(m) = \begin{cases} K_{10}\mathbf{x}_1(m) + \mathbf{z}_0(m) & m \bmod 4 = 1, \\ K_{R0}\mathbf{x}_R(m) + \mathbf{z}_0(m) & m \bmod 4 = 2, \\ K_{20}\mathbf{x}_2(m) + \mathbf{z}_0(m) & m \bmod 4 = 3, \\ K_{R0}\mathbf{x}_R(m) + \mathbf{z}_0(m) & m \bmod 4 = 0, \end{cases} \tag{1}$$

where the vectors $\mathbf{y}_i$ and $\mathbf{x}_i$ contain the blocks of received and transmitted signals of node i, respectively; K_{ij}, $i \in \{1, 2, R\}$

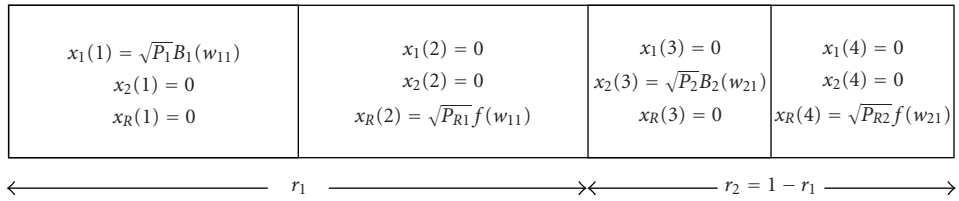


Figure 2: One frame of the considered orthogonal cooperation scheme for the case of 2 source nodes, and its constituent subframes.

and $j \in \{R, 0\}$, represents the channel gain between nodes i and j; $\mathbf{z}_j$ represents the additive zero mean white circular complex Gaussian noise with variance σ_j^2 at node j; and $\mathbf{0}$ is used to represent blocks in which the receiver of the relay node is turned off. For simplicity, we define the effective power gain $\gamma_{ij} = |K_{ij}|^2/\sigma_j^2$.

The focus of this paper will be on a system in which full channel state information (CSI) is available at the source nodes, and the channel coherence time is long. The CSI is exploited to jointly allocate the powers P_{Ri} and the resource allocation parameters r_i, with the goal of enlarging the achievable rate region. Under the assumption of equal channel resource allocation (i.e., $r_i = r_s$, for all i, s), expressions for the maximum achievable rate for a source node under each of the four relaying considered relaying strategies were provided in [13]. The extension of those expressions to the case of not necessarily equal resource allocation results in the following expressions for the maximum achievable rate of node i as a function of P_i, the transmission power of node i, P_{Ri}, the relay power allocated to node i, and r_i, the fraction of the channel resource allocated to node i.

(i) Regenerative decode-and-forward (RDF):

$$\overline{R}_{i,\mathrm{RDF}} = \frac{r_i}{2} \min\left\{\log\left(1+\gamma_{iR}P_i\right),\ \log\left(1+\gamma_{i0}P_i+\gamma_{R0}P_{Ri}\right)\right\}. \tag{2a}$$

(ii) Nonregenerative decode-and-forward (NDF):

$$\overline{R}_{i,\mathrm{NDF}} = \frac{r_i}{2} \min\left\{\log\left(1+\gamma_{iR}P_i\right), \log\left(1+\gamma_{i0}P_i\right) + \log\left(1+\gamma_{R0}P_{Ri}\right)\right\}. \tag{2b}$$

(iii) Amplify-and-forward (AF):

$$\overline{R}_{i,\mathrm{AF}} = \frac{r_i}{2} \log\left(1+\gamma_{i0}P_i + \frac{\gamma_{iR}\gamma_{R0}P_iP_{Ri}}{\left(1+\gamma_{iR}P_i+\gamma_{R0}P_{Ri}\right)}\right). \tag{2c}$$

(iv) Compress-and-forward (CF): assuming that the relay uses Wyner-Ziv lossy compression [17], the maximum achievable rate is

$$\overline{R}_{i,\mathrm{CF}} = \frac{r_i}{2} \log\left(1+\gamma_{i0}P_i + \frac{\gamma_{iR}\gamma_{R0}\left(\gamma_{i0}P_i+1\right)P_iP_{Ri}}{\gamma_{R0}\left(\gamma_{i0}P_i+1\right)P_{Ri}+P_i\left(\gamma_{i0}+\gamma_{iR}\right)+1}\right). \tag{2d}$$

The focus of the work in this paper will be on systems in which the relay node relays the messages of all source nodes in the system using the same preassigned relaying strategy. However, as we will demonstrate in Section 4, our results naturally extend to the case of heterogeneous relaying strategies, and hence facilitate the development of algorithms for the jointly optimal allocation of the relaying strategy.

3. JOINT POWER AND CHANNEL RESOURCE ALLOCATION

It was shown in [13] that for fixed channel resource allocation, the problem of finding the power allocation that maximizes the sum rate is convex, and closed-form solutions for the optimal power allocation were obtained. However, the direct formulation of the problem of joint allocation of both the power and the channel resource so as to enable operation at an arbitrary point on the boundary of the achievable rate region is not convex, and hence is significantly harder to solve. Despite this complexity, the problem is of interest because it is expected to yield significantly larger achievable rate regions than those obtained with equal channel resource allocation. In the next four subsections, we will study the problem of finding the jointly optimal power and resource allocation for each relaying strategy. We will show that in each case the problem can be transformed into a quasiconvex problem, and hence an optimal solution can be obtained using simple and efficient algorithms, that is, standard quasiconvex search algorithms [14]. Furthermore, for a fixed resource allocation, a closed-form solution for the optimal power allocation is obtained. By exposing the quasiconvexity of the problem and by obtaining a closed-form solution to the power allocation problem, we are able to achieve significantly larger achievable rate regions without incurring substantial additional computational cost.

The jointly optimal power and channel resource allocation at each point on the boundary of the achievable rate region can be found by maximizing a weighted sum of the maximal rates $\overline{R}_1$ and $\overline{R}_2$ subject to the bound on the transmitted powers, that is,

$$\begin{aligned} \max_{P_i,P_{Ri},r} \quad & \mu\overline{R}_1 + (1-\mu)\overline{R}_2, \\ \text{subject to} \quad & \frac{r}{2}P_{R1} + \frac{\hat{r}}{2}P_{R2} \le \overline{P}_R, \\ & \frac{r}{2}P_1 \le \overline{P}_1, \qquad \frac{\hat{r}}{2}P_2 \le \overline{P}_2, \\ & P_{Ri} \ge 0, \qquad P_i \ge 0, \end{aligned} \tag{3}$$

where $\overline{R}_i$ is the expression in (2a), (2b), (2c), or (2d) that corresponds to the given relaying strategy, $r = r_1$, $\hat{r} = r_2 = 1 - r$, and $\mu \in [0,1]$ weights the relative importance of $\overline{R}_1$ over $\overline{R}_2$. Alternatively, the jointly optimal power and channel resource allocation at each point on the boundary of the achievable rate region can also be found by maximizing $\overline{R}_i$ for a given target value of $\overline{R}_j$, subject to the bound on the transmitted powers, for example,

$$\begin{aligned} \max_{P_i, P_{Ri}, r} \quad & \overline{R}_1, \\ \text{subject to} \quad & \overline{R}_2 \geq R_{2,\text{tar}}, \\ & \frac{r}{2} P_{R1} + \frac{\hat{r}}{2} P_{R2} \leq \overline{P}_R, \\ & \frac{r}{2} P_1 \leq \overline{P}_1, \qquad \frac{\hat{r}}{2} P_2 \leq \overline{P}_2, \\ & P_{Ri} \geq 0, \qquad P_i \geq 0. \end{aligned} \tag{4}$$

Neither the formulation in (3) nor that in (4) is jointly convex in the transmitted powers and the channel resource allocation parameter r, and hence it appears that it may be difficult to develop a reliable efficient algorithm for their solution. However, in the following subsections, we will show that by adopting the framework in (4), the direct formulation can be transformed into a composition of a convex problem (with a closed-form solution) and a quasiconvex optimization problem, and hence it can be efficiently and reliably solved. The first step in that analysis is to observe that since the source nodes transmit on channels that are orthogonal to each other and to that of the relay, then at optimality they should transmit at full power, that is, the optimal values of P_1 and P_2 are $P_1^*(r) = 2\overline{P}_1/r$ and $P_2^*(r) = 2\overline{P}_2/\hat{r}$, respectively. In order to simplify our development, we will define $R_{2,\max}(r)$ to be the maximum achievable value for $\overline{R}_2$ for a given value of r and the given relaying strategy, that is, the value of the appropriate expression in (2a), (2b), (2c), or (2d) with $P_{R2} = 2\overline{P}_R/\hat{r}$ and $P_2 = 2\overline{P}_2/\hat{r}$.

3.1. Regenerative decode-and-forward

For the regenerative decode-and-forward strategy, the problem in (4) can be written as

$$\begin{aligned} \max_{P_{Ri}, r} \quad & \frac{r}{2} \min\left\{ \log\left(1 + \gamma_{1R} P_1^*\right), \right. \\ & \qquad \left. \log\left(1 + \gamma_{10} P_1^* + \gamma_{R0} P_{R1}\right) \right\}, \\ \text{subject to} \quad & \frac{\hat{r}}{2} \min\left\{ \log\left(1 + \gamma_{2R} P_2^*\right), \right. \\ & \qquad \left. \log\left(1 + \gamma_{20} P_2^* + \gamma_{R0} P_{R2}\right) \right\} \geq R_{2,\text{tar}}, \\ & \frac{r}{2} P_{R1} + \frac{\hat{r}}{2} P_{R2} \leq \overline{P}_R, \\ & P_{Ri} \geq 0. \end{aligned} \tag{5}$$

Unfortunately, the set of values for r, P_{R1}, and P_{R2} that satisfy the second constraint of (5) is bilinear, and hence the problem in (5) is not convex. However, if we define $\widetilde{P}_{R1} = rP_{R1}$ and $\widetilde{P}_{R2} = \hat{r}P_{R2}$, then the problem in (5) can be rewritten as

$$\begin{aligned} \max_{\widetilde{P}_{Ri}, r} \quad & \frac{r}{2} \min\left\{ \log\left(1 + \frac{2\gamma_{1R}\overline{P}_1}{r}\right), \right. \\ & \qquad \left. \log\left(1 + \frac{2\gamma_{10}\overline{P}_1 + \gamma_{R0}\widetilde{P}_{R1}}{r}\right) \right\}, \\ \text{subject to} \quad & \frac{\hat{r}}{2} \min\left\{ \log\left(1 + \frac{2\gamma_{2r}\overline{P}_2}{\hat{r}}\right), \right. \\ & \qquad \left. \log\left(1 + \frac{2\gamma_{20}\overline{P}_2 + \gamma_{R0}\widetilde{P}_{R2}}{\hat{r}}\right) \right\} \geq R_{2,\text{tar}}, \\ & \widetilde{P}_{R1} + \widetilde{P}_{R2} = 2\overline{P}_R, \\ & \widetilde{P}_{Ri} \geq 0. \end{aligned} \tag{6}$$

Formulating the problem as in (6) enables us to obtain the following result, the proof of which is provided in Appendix A.

Proposition 1. *For a given feasible target rate $R_{2,\text{tar}} \in (0, R_{2,\max}(0))$, the maximum achievable rate $\overline{R}_{1,\max}$ in (6) is a quasiconcave function of the channel resource sharing parameter r.*

In addition to the desirable property in Proposition 1, for any given channel resource allocation and for any feasible $R_{2,\text{tar}}$, a closed-form solution for the optimal power allocation can be found. In particular, for any given r, $\widetilde{P}_{R1}$ must be maximized in order to maximize R_1. Therefore, the optimal value of $\widetilde{P}_{R2}$ is the minimum value that satisfies the constraints in (6), and hence it can be written as

$$\widetilde{P}_{R2}^*(r) = \begin{cases} 0 & \text{if } \gamma_{2R} \leq \gamma_{20}, \\ \left(\dfrac{A - 2\gamma_{20}\overline{P}_2}{B}\right)^+ & \text{if } \gamma_{2R} > \gamma_{20}, \end{cases} \tag{7}$$

where $A = \hat{r}(2^{2R_{2,\text{tar}}/\hat{r}} - 1)$, $B = \gamma_{R0}$,and $x^+ = \max(0, x)$. The optimal value of $\widetilde{P}_{R1}$ is $\widetilde{P}_{R1}^*(r) = \min\{2\overline{P}_R - \widetilde{P}_{R2}^*(r), (2\overline{P}_1(\gamma_{1R} - \gamma_{10})/\gamma_{R0})^+\}$, where the second argument of the min function is the value of $\widetilde{P}_{R1}$ that makes the two arguments of the min function in the objective function of (6) equal. In Section 3.5, we will exploit the quasiconvexity result in Proposition 1 and the closed-form expression for $\widetilde{P}_{R2}^*(r)$ in (7) to develop an efficient algorithm for the jointly optimal allocation of power and the channel resource.

3.2. Nonregenerative decode-and-forward

Using the definition of $\widetilde{P}_{R1}$ and $\widetilde{P}_{R2}$ from the RDF case, the problem of maximizing the achievable rate region for the NDF relaying strategy can be written as

$$\begin{aligned} \max_{\widetilde{P}_{Ri}, r} \quad & \frac{r}{2} \min\left\{ \log\left(1 + \frac{2\gamma_{1R}\overline{P}_1}{r}\right), \log\left(1 + \frac{2\gamma_{10}\overline{P}_1}{r}\right) \right. \\ & \qquad \left. + \log\left(1 + \frac{\gamma_{R0}\widetilde{P}_{R1}}{r}\right) \right\}, \end{aligned}$$

$$\begin{aligned}\text{subject to}\quad & \frac{\hat{r}}{2}\min\left\{\log\left(1+\frac{2\gamma_{2R}\overline{P}_2}{\hat{r}}\right),\log(1+\frac{2\gamma_{20}\overline{P}_2}{\hat{r}}\right)\\ &\qquad +\log\left(1+\frac{\gamma_{R0}\widetilde{P}_{R2}}{\hat{r}}\right)\Big\}\geq R_{2,\text{tar}},\\ &\widetilde{P}_{R1}+\widetilde{P}_{R2}=2\overline{P}_R,\\ &\widetilde{P}_{Ri}\geq 0.\end{aligned}\tag{8}$$

Using the formulation in (8), we obtain the following result in Appendix B.

Proposition 2. *For a given feasible target rate $R_{2,\text{tar}}\in(0,R_{2,\max}(0))$, the maximum achievable rate $\overline{R}_{1,\max}$ in (8) is a quasiconcave function of r.*

Similar to the RDF case, for a given r and a feasible $R_{2,\text{tar}}$, a closed-form expression for the optimal $\widetilde{P}_{R2}$ can be obtained. This expression has the same form as that in (7), with the same definition for A, but with B defined as $B=\gamma_{R0}+2\gamma_{20}\gamma_{R0}\overline{P}_2/\hat{r}$. The optimal value for $\widetilde{P}_{R1}$ is $\widetilde{P}^*_{R1}(r)=\min\{2\overline{P}_R-\widetilde{P}^*_{R2}(r),(2\overline{P}_1(\gamma_{1R}-\gamma_{10})r/(\gamma_{R0}(r+2\overline{P}_1\gamma_{10})))^+\}$, where the second argument of the min function is the value of $\widetilde{P}_{R1}$ that makes the two arguments of the min function in the objective function of (8) equal.

3.3. Amplify-and-forward

In the case of amplify-and-forward relaying, problem (4) can be written as

$$\begin{aligned}\max_{\widetilde{P}_{Ri},r}\quad & \frac{r}{2}\log\left(1+\frac{2\gamma_{10}\overline{P}_1}{r}+\frac{2\gamma_{1R}\gamma_{R0}\overline{P}_1\widetilde{P}_{R1}}{r(r+2\gamma_{1R}\overline{P}_1+\gamma_{R0}\widetilde{P}_{R1})}\right),\\ \text{subject to}\quad & \frac{\hat{r}}{2}\log\left(1+\frac{2\gamma_{20}\overline{P}_2}{\hat{r}}+\frac{2\gamma_{2R}\gamma_{R0}\overline{P}_2\widetilde{P}_{R2}}{\hat{r}(\hat{r}+2\gamma_{2r}\overline{P}_2+\gamma_{R0}\widetilde{P}_{R2})}\right)\\ &\qquad\geq R_{2,\text{tar}},\\ &\widetilde{P}_{R1}+\widetilde{P}_{R2}=2\overline{P}_R,\\ &\widetilde{P}_{Ri}\geq 0.\end{aligned}\tag{9}$$

Using this formulation, we obtain the following result in Appendix C. (We point out that $\gamma_{i0}\overline{P}_i$ is the maximum achievable destination SNR on the direct channel of source node i.)

Proposition 3. *If the direct channels of both source nodes satisfy $\gamma_{i0}\overline{P}_i>1/2$, then for a given feasible target rate $R_{2,\text{tar}}\in(0,R_{2,\max}(0))$, the maximum achievable rate $\overline{R}_{1,\max}$ in (9) is a quasiconcave function of r.*

Similar to the cases of RDF and NDF relaying, for a given r and a feasible $R_{2,\text{tar}}$, in order to obtain an optimal power allocation we must find the smallest $\widetilde{P}_{R2}$ that satisfies the constraints in (9). If we define $C=A-2\gamma_{20}\overline{P}_2$, a closed-form solution for $\widetilde{P}_{R2}$ can be written as

$$\widetilde{P}^*_{R2}(r)=\left(\frac{C(\hat{r}+2\gamma_{2R}\overline{P}_2)}{2\gamma_{2R}\gamma_{R0}\overline{P}_2-\gamma_{R0}C}\right)^+.\tag{10}$$

Hence, the optimal value of $\widetilde{P}_{R1}$ is $\widetilde{P}^*_{R1}(r)=2\overline{P}_R-\widetilde{P}^*_{R2}(r)$.

Given $R_{2,\text{tar}}\in(0,R_{2,\max}(0))$, for $r\in(0,1)$, define $\psi(r)$ to be the optimal value of (4) for a given r if $R_{2,\text{tar}}\in(0,R_{2,\max}(r))$ and zero otherwise. Set $\psi(0)=0$ and $\psi(1)=0$. Set $t_0=0$, $t_4=1$, and $t_2=1/2$. Using the closed-form expression for the optimal power allocations, compute $\psi(t_2)$. Given a tolerance ε,

(1) set $t_1=(t_0+t_2)/2$ and $t_3=(t_2+t_4)/2$,
(2) using the closed-form expressions for the power allocations, compute $\psi(t_1)$ and $\psi(t_3)$,
(3) find $k^*=\arg\max_{k\in\{0,1,\ldots,4\}}\psi(t_k)$,
(4) replace t_0 by $t_{\max\{k^*-1,0\}}$, replace t_4 by $t_{\min\{k^*+1,4\}}$, and save $\psi(t_0)$ and $\psi(t_4)$. If $k^*\notin\{0,4\}$ set $t_2=t_{k^*}$ and save $\psi(t_2)$, else set $t_2=(t_0+t_4)/2$ and use the closed form expressions for the power allocations to calculate $\psi(t_2)$.
(5) if $t_4-t_0\geq\varepsilon$, return to (1), else set $r^*=t_{k^*}$.

Algorithm 1: A simple method for finding r^*

3.4. Compress-and-forward

Finally, for the compress-and-forward relaying strategy, the problem in (4) can be written as

$$\begin{aligned}\max_{\widetilde{P}_{Ri},r}\quad & \frac{r}{2}\min\left\{\log\left(1+\frac{2\gamma_{1R}\overline{P}_1}{r}\right),\ \log\left(1+\frac{2\gamma_{10}\overline{P}_1}{r}\right)\right.\\ &\qquad\left.+\log\left(1+\frac{\gamma_{R0}\widetilde{P}_{R1}}{r}\right)\right\},\\ \text{subject to}\quad & \frac{\hat{r}}{2}\min\left\{\log\left(1+\frac{2\gamma_{2R}\overline{P}_2}{\hat{r}}\right),\ \log\left(1+\frac{2\gamma_{20}\overline{P}_2}{\hat{r}}\right)\right.\\ &\qquad\left.+\log\left(1+\frac{\gamma_{R0}\widetilde{P}_{R2}}{\hat{r}}\right)\right\}\geq R_{2,\text{tar}},\\ &\widetilde{P}_{R1}+\widetilde{P}_{R2}=2\overline{P}_R,\\ &\widetilde{P}_{Ri}\geq 0.\end{aligned}\tag{11}$$

As we state in the following proposition (proved in Appendix D), the quasiconvex properties of the problem in (11) are similar to those of the amplify-and-forward case.

Proposition 4. *If the direct channels of both source nodes satisfy $\gamma_{i0}\overline{P}_i>1/2$, then for a given feasible target rate $R_{2,\text{tar}}\in(0,R_{2,\max}(0))$, the maximum achievable rate $\overline{R}_{1,\max}$ in (11) is a quasiconcave function of r.*

If we define $D=\gamma_{R0}(2\gamma_{20}\overline{P}_2+\hat{r})$, then the optimal solution for $\widetilde{P}_{R2}$ for a given r and a feasible $R_{2,\text{tar}}$ can be written as

$$\widetilde{P}^*_{R2}(r)=\left(\frac{C\hat{r}(\hat{r}+2(\gamma_{20}+\gamma_{2R})\overline{P}_2)}{D(2\gamma_{2R}\overline{P}_2-C)}\right)^+,\tag{12}$$

and the optimal $\widetilde{P}_{R1}$ is $\widetilde{P}^*_{R1}(r)=2\overline{P}_R-\widetilde{P}^*_{R2}(r)$.

Table 1: Parameters of the two-user channel models used in the numerical results.

	$\lvert K_{10}\rvert$	$\lvert K_{1R}\rvert$	$\lvert K_{20}\rvert$	$\lvert K_{2R}\rvert$	$\lvert K_{R0}\rvert$	$\sigma_R^2 = \sigma_0^2$	$\overline{P}_1$	$\overline{P}_2$	$\overline{P}_R$
Scenario 1	0.3	1.2	0.8	0.6	0.4	1	2	2	4
Scenario 2	0.3	1.2	0.6	0.8	0.4	1	2	2	4

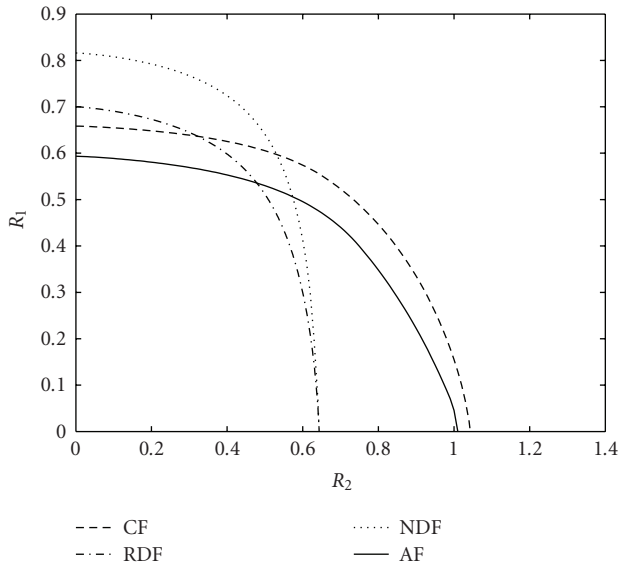


Figure 3: Achievable rate regions obtained via jointly optimal power and resource allocation in Scenario 1.

3.5. Summary and extensions

In the previous four subsections, we have shown that the problem of jointly allocating the power and the channel resource so as to enable operation at any point on the boundary of the achievable rate region is quasiconvex. In addition, we have shown that for a given resource allocation, a closed-form solution for the optimal power allocation can be obtained. These results mean that we can determine the optimal value for r using a standard efficient search method for quasiconvex problems (see, e.g., [14]). (In the AF and CF cases, these results are contingent on the maximum achievable SNR of both direct channels, being greater than -3 dB, which would typically be the case in practice. Furthermore, since the condition $\gamma_{i0}\overline{P}_i > 1/2$ depends only on the direct channel gains, the noise variance at the destination node, and the power constraints, this condition is testable before the design process commences.)

For the particular problem at hand, a simple approach that is closely related to bisection search is provided in Algorithm 1. At each step in that approach, we use the closed-form expressions for the optimal power allocation for each of the current values of r. Since the quasiconvex search can be efficiently implemented and since it converges rapidly, the jointly optimal values for r and the (scaled) powers $\widetilde{P}_{Ri}$ can be efficiently obtained.

In the above development, we have focused on the case of two source nodes. However, the core results extend directly

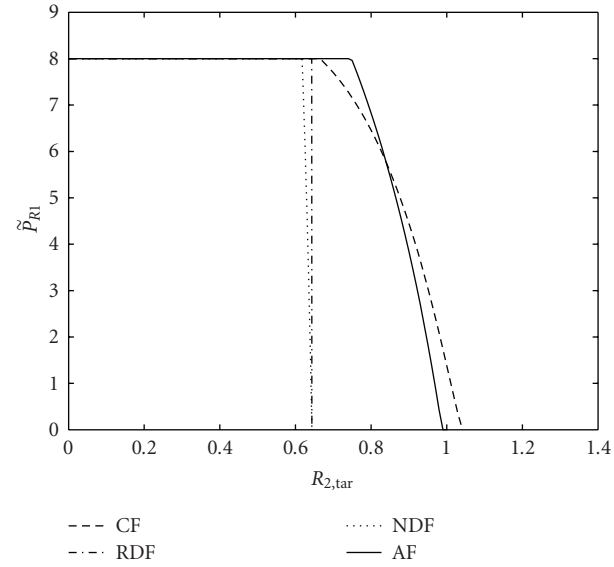


Figure 4: Powers allocated by the jointly optimal algorithm in Scenario 1.

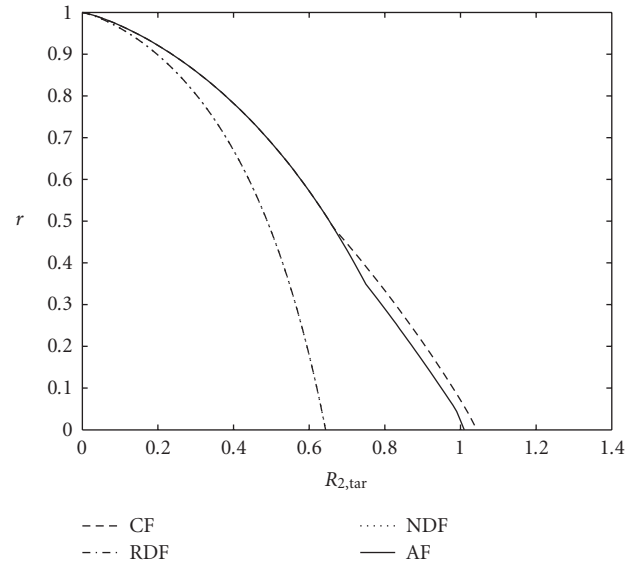
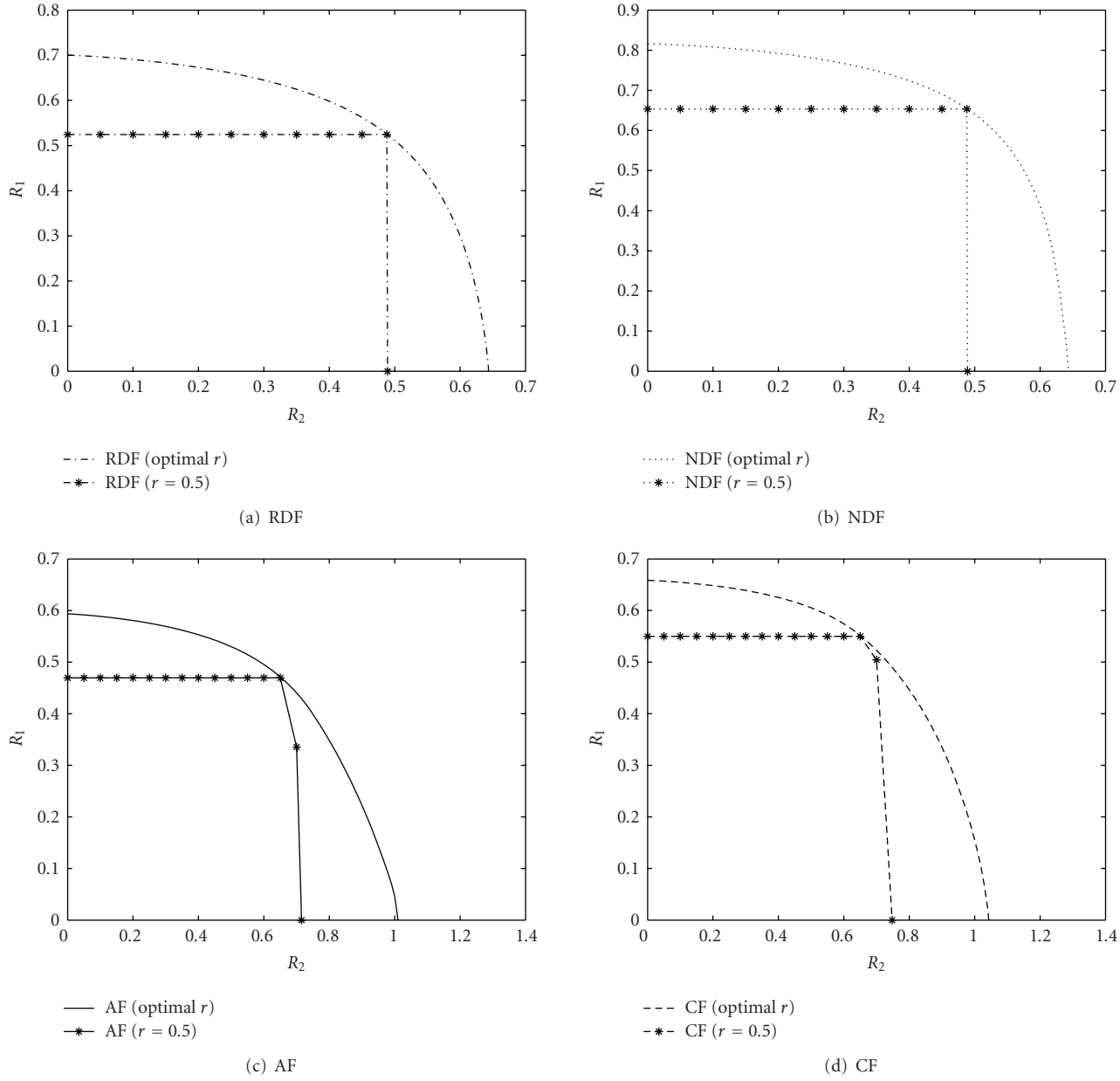


Figure 5: Resource allocation from the jointly optimal algorithm in Scenario 1.

to the case of $N > 2$ source nodes. Indeed, the joint power and resource allocation problem can be written in a form analogous to those in (6), (8), (9), and (11). To do so, we let $\overline{R}_i$ denote the appropriate maximal rate for node i from (2a), (2b), (2c), or (2d), and we define $\widetilde{P}_{Ri} = r_i P_{Ri}$, where P_{Ri} is the relay power allocated to the message of node i. If we choose to maximize the achievable rate of node j subject to target rate requirements for the other nodes, then the problem can be written as

$$\max_{\widetilde{P}_{Ri}, r_i} \quad \overline{R}_j,$$
$$\text{subject to} \quad \overline{R}_i \geq R_{i,\mathrm{tar}} \quad i = 1, 2, \ldots, N;\ i \neq j,$$



(a) RDF (b) NDF

(c) AF (d) CF

Figure 6: Comparisons between the achievable rate regions obtained by jointly optimal power and resource allocation and those obtained by power allocation only with equal resource allocation, for Scenario 1.

$$\begin{aligned} &\sum_{i=1}^{N} \widetilde{P}_{Ri} \leq 2\overline{P}_R, \\ &\widetilde{P}_{Ri} \geq 0, \\ &\sum_{i=1}^{N} r_i = 1. \end{aligned} \tag{13}$$

Using similar techniques to those in the previous subsections, it can be shown that this problem is quasiconvex in $(N-1)$ resource allocation parameters. The other parameter is not free as the resource allocation parameters must sum to one. (In the AF and CF cases, this result is, again, contingent on the condition $\gamma_{i0}\overline{P}_i > 1/2$ holding for all i). Furthermore, since for a given value of i, the expression $\overline{R}_i \geq R_{i,\text{tar}}$ depends only on $\widetilde{P}_{Ri}$ and r_i, for a given set of target rates for nodes $i \neq j$ and a given set of resource allocation parameters, a closed-form expression for the optimal $\widetilde{P}_{Ri}$ can be obtained (for the chosen relaying strategy). These expressions have a structure that is analogous to the corresponding expression for the case of two source nodes that was derived in the subsections above. As we will demonstrate in Section 4, problems of the form in (13) can be efficiently solved using $(N-1)$-dimensional quasiconvex search methods, in which the closed-form solution for the optimal powers given a fixed resource allocation is used at each step.

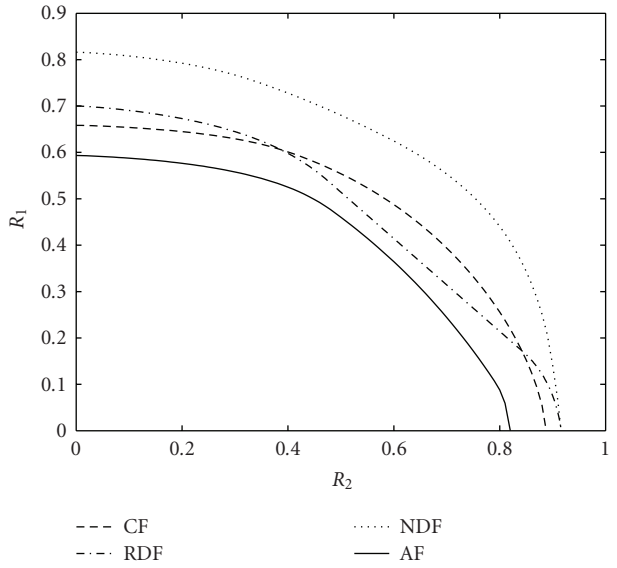


Figure 7: Achievable rate regions obtained via jointly optimal power and resource allocation in Scenario 2.

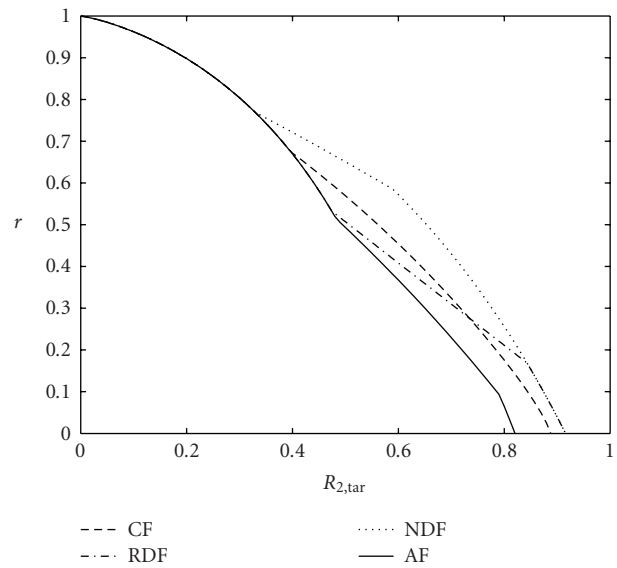


Figure 9: Resource allocation from the jointly optimal algorithm in Scenario 2.

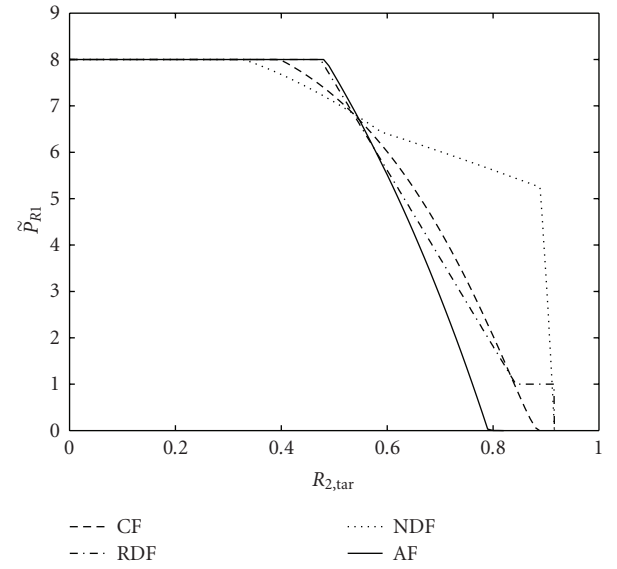


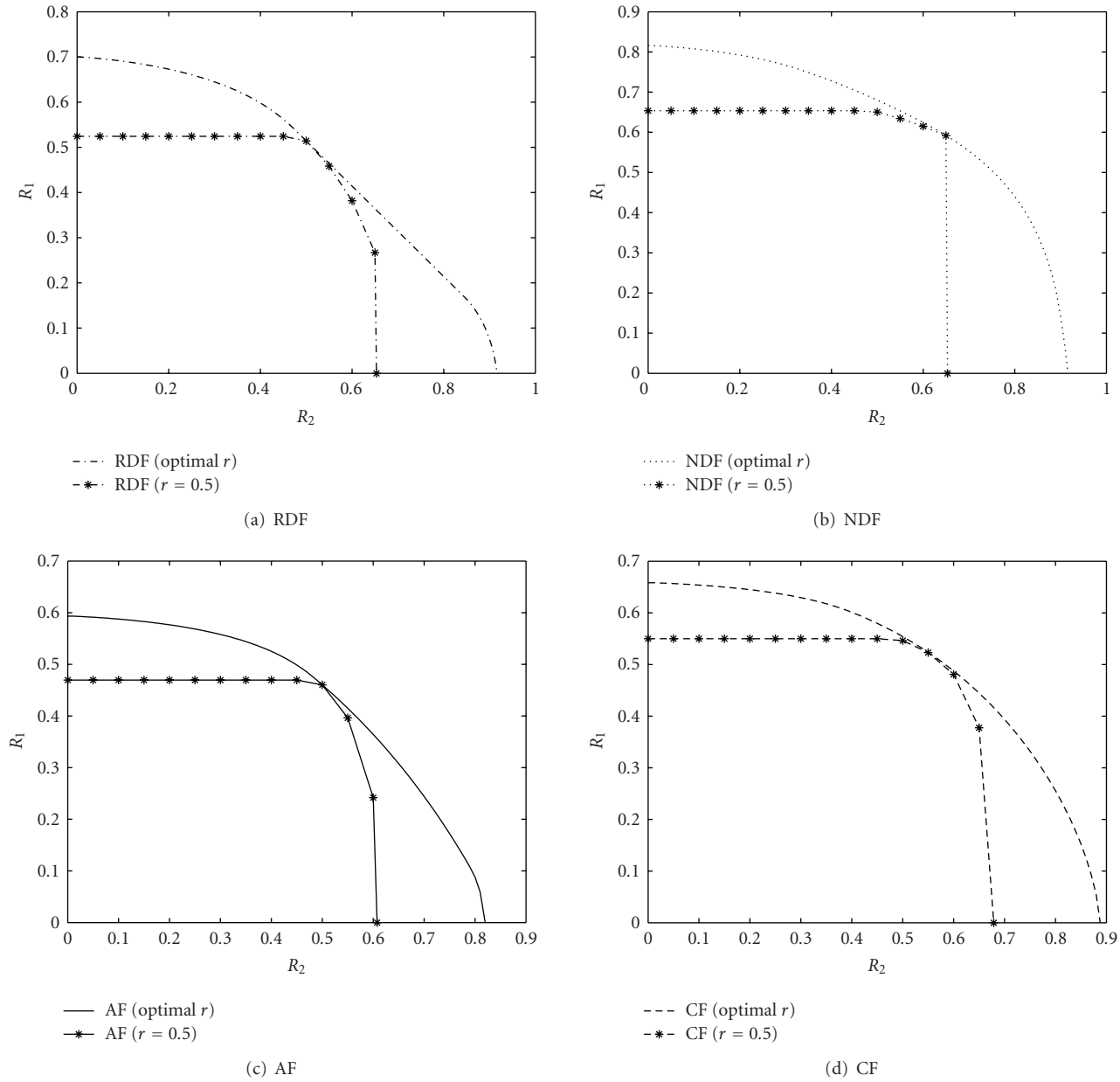
Figure 8: Powers allocated by jointly optimal algorithm in Scenario 2.

In the development above, we have considered systems in which the relay node uses the same (preassigned) relaying strategy for each node. However, since the source nodes use orthogonal channels, our results extend directly to the case of different relaying strategies, and we will provide an example of such a heterogeneous multiple access relay system in the numerical results below.

4. NUMERICAL RESULTS

In this section, we provide comparisons between the achievable rate regions obtained by different relaying strategies with the jointly optimal power and channel resource allocation derived in Section 3. We also provide comparisons between the achievable rate regions obtained with jointly optimal power and channel resource allocation and those obtained using optimal power allocation alone, with equal channel resource allocation, $r = 0.5$. We will provide comparisons for two different channel models, whose parameters are given in Table 1. Finally, we show that in some cases assigning different relaying strategies to different source nodes can result in a larger achievable rate region than assigning the same relaying strategy to all source nodes.

In Figure 3, we compare the achievable rate regions for the four relaying strategies, RDF, NDF, CF, and AF, in Scenario 1 in Table 1. In this scenario, the source-relay channel of node 1 has higher effective gain than its direct channel, whereas for node 2 the direct channel is better than the source-relay channel. Therefore, for small values of R_1 one would expect the values of R_2 that can be achieved by the CF and AF relaying strategies to be greater than those obtained by RDF and NDF, since the values of R_2 that can be achieved by RDF and NDF will be limited by the source-relay link, which is weak for node 2. Furthermore, for small values of R_2, one would expect RDF and NDF to result in higher achievable values of R_1 than CF and AF, since the source-relay link for node 1 is strong and does not represent the bottleneck in this case. Both these expected characteristics are evident in Figure 3. In Figure 4, we provide the power allocation $\widetilde{P}_{R1}$ for the four relaying strategies, and Figure 5 shows the channel resource allocation. (Note that, as expected, the optimal resource allocation is dependent on the choice of the relaying strategy.) It is interesting to observe that for the RDF strategy the relay power allocated to node 2 is zero, that is, $\widetilde{P}_{R1} = 2\overline{P}_R$ for all feasible values of $R_{2,\text{tar}}$. This solution is optimal because in Scenario 1 the achievable rate of node 2 for the RDF strategy is limited by the



(a) RDF (b) NDF

(c) AF (d) CF

Figure 10: Comparisons between the achievable rate regions obtained by jointly optimal power and resource allocation and those obtained by power allocation only with equal resource allocation, for Scenario 2.

source-relay link and there is no benefit to allocate any relay power to node 2. For the same reason, the relay power allocated to node 2 in the case of NDF relaying is also zero. However, in the case of NDF relaying, for small values of r, there is no need to use all the relay power to relay the messages of node 1, that is, $\widetilde{P}_{R1} < 2\overline{P}_R$, and it is sufficient to use only the amount of power $\widetilde{P}_{R1}$ that makes the arguments of the min function in (8) equal, that is, $\widetilde{P}_{R1} = 2\overline{P}_1(\gamma_{1R} - \gamma_{10})r/(\gamma_{R0}(r + 2\overline{P}_1\gamma_{10}))$. This can be seen in Figure 4 as the (steeply) decreasing dotted curve that represents the optimal $\widetilde{P}_{R1}$ for the case of NDF relaying. For values of R_2 in this region, the average power that the relay needs to use is strictly less than its maximum average power. We also observe from Figure 5 that the channel resource allocations for both RDF and NDF are the same. This situation arises because in both strategies the achievable rate of node 2 is limited by the achievable rate of the source-relay link. This rate has the same expression for both strategies, and hence, the same value of $\hat{r}$ will be allocated to node 2. A further observation from Figure 3 is that the achievable rate region for the CF relaying strategy is larger than that for AF and the achievable rate region for NDF is larger than that for RDF. This is consistent with the observations in [13], where the comparisons were made in terms of the expressions in (2a), (2b), (2c), and (2d) with $r = 1/2$.

To provide a quantitative comparison to the case of power allocation alone with equal resource allocation, we

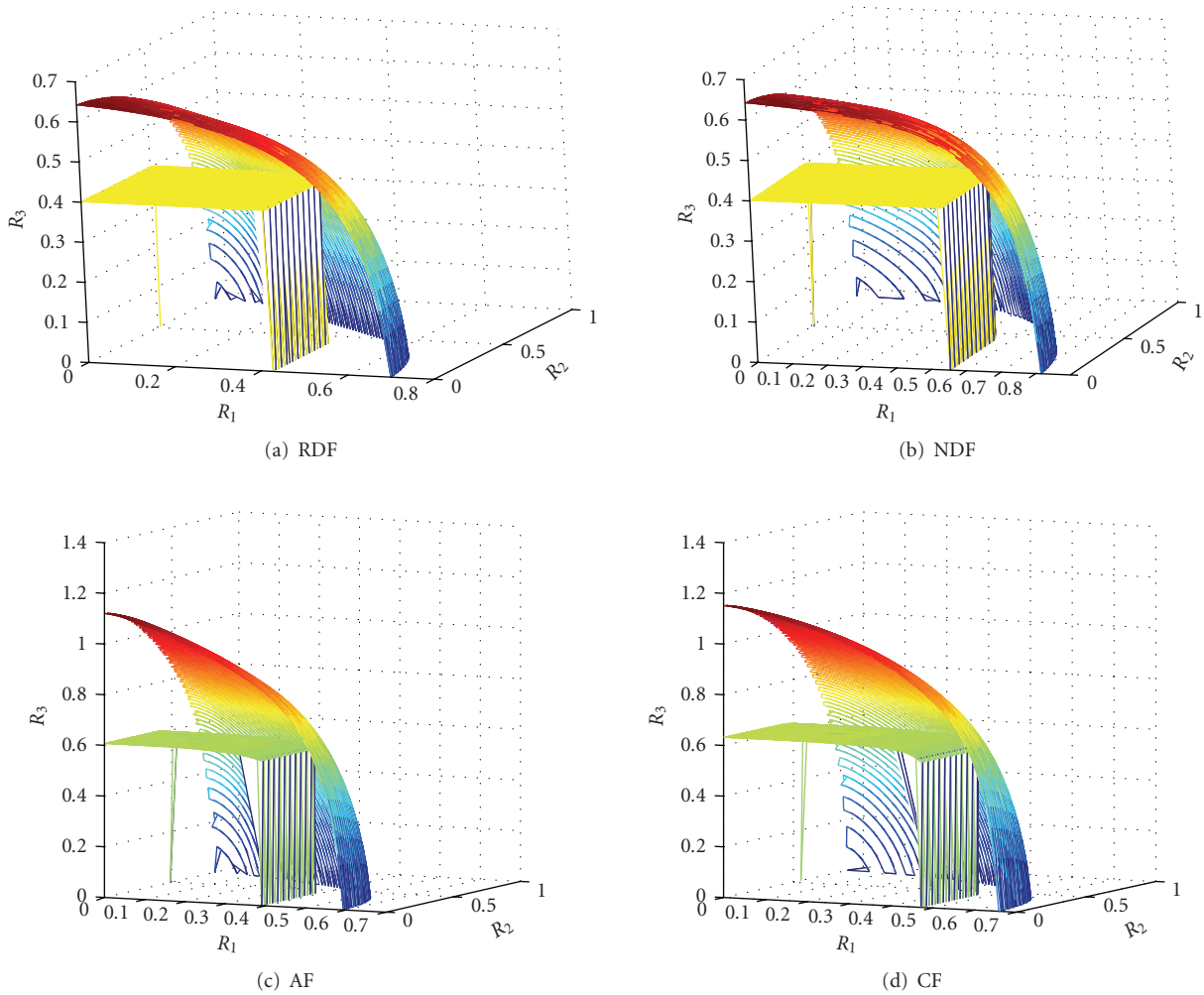


Figure 11: The achievable rate regions obtained by jointly optimal power and resource allocation and those obtained by power allocation alone with equal resource allocation for three-user system with $|K_{3R}| = 0.6$, $|K_{30}| = 0.9$, $\overline{P}_3 = 2$, and the remaining parameters from Scenario 2 in Table 1.

plot in Figure 6 the rate regions achieved by joint allocation and by power allocation alone for each relaying strategy. It is clear from the figure that the joint allocation results in significantly larger achievable rate regions. (The horizontal segments of the regions with $r = 0.5$ in Figure 6 arise from the allocation of all the relay power to node 1. In these cases, $R_{2,\text{tar}}$ can be achieved without the assistance of the relay, and hence all the relay power can be allocated to the message of node 1.) As expected, each of the curves for $r = 0.5$ in Figure 6 touches the corresponding curve for the jointly optimal power and channel resource allocation at one point. This point corresponds to the point at which the value $r = 0.5$ is (jointly) optimal.

In Figures 7–10, we examine the performance of the considered scheme in Scenario 2 of Table 1, in which the effective gain of the source-relay channel for node 2 is larger than that in Scenario 1, and that of the source-destination channel is smaller. As can be seen from Figure 7, increasing the gain of the source-relay channel of node 2 expands the achievable rate of the RDF and NDF strategies, even though the gain of the direct channel is reduced, whereas that change in the channel gains has resulted in the shrinkage of the achievable rate region for the CF and AF strategies. Therefore, we can see that the RDF and NDF strategies are more dependent on the quality of the source-relay channel than that of the source-destination channel (so long as the first term in the argument of the min function in (2a) and (2b) is no more than the second term), while the reverse applies to the CF and AF strategies. Figures 8 and 9 show the allocations of the relay power and the channel resource parameter, respectively. It is interesting to note that for the RDF strategy, when $R_{2,\text{tar}}$ is greater than a certain value, the relay power allocated to node 2 will be constant. The value of this constant is that which makes the two terms inside the min function on the left-hand side of the first constraint of (6) equal. This value can be calculated from the expression $\widetilde{P}_{R2} = 2(\gamma_{2R} - \gamma_{20})\overline{P}_2/\gamma_{R0}$. Figure 10 provides comparisons between the achievable rate regions obtained by the jointly optimal allocation and those obtained by optimal power allocation alone with equal resource allocation. As in

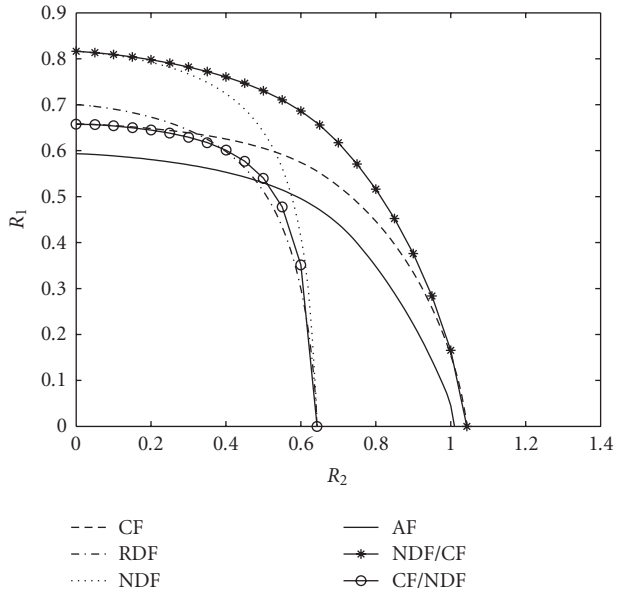


FIGURE 12: Comparison between the achievable rate regions when using the same relaying strategy for both users and when using different relaying strategies, for Scenario 1.

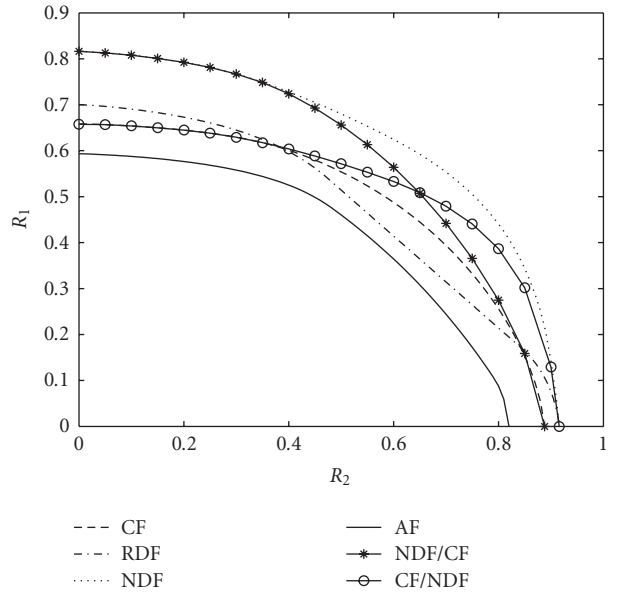


FIGURE 13: Comparison between the achievable rate regions when using the same relaying strategy for both users and when using different relaying strategies, for Scenario 2.

Figure 6, it is again clear that the joint allocation results in significantly larger achievable rate regions. (The horizontal segments in Figure 10 arise from all the relay power being allocated to node 1, because the corresponding values of $R_{2,\text{tar}}$ can be achieved without the assistance of the relay.)

In Figure 11, we extend this comparison to a three-user case. In order to obtain the jointly optimal power and channel resource allocations used to plot this figure, we used a two-dimensional quasiconvex search algorithm analogous to that in Algorithm 1 to solve instances of the optimization problem in (13). As in the two-user case, a substantially larger rate region can be achieved by joint allocation of the power and the channel resource. It is worth mentioning that the $R_3 = 0$ slice through the jointly optimized region is the same as that obtained in the corresponding two-user case (cf. Figure 7). This is because when $R_{3,\text{tar}} = 0$, no power and none of the channel resource will be allocated to the message of node 3. On the other hand, the $R_3 = 0$ slice through the region with fixed (and equal) resource allocation will be smaller than the corresponding region in Figure 7, because equal resource allocation in the three-user case corresponds to $r_i = 1/3$. This indicates that as the number of source nodes increases, so do the benefits of joint power and resource allocation.

In the above examples, we have considered the case in which the relay applies the same strategy to the messages of all source nodes. However, in Figures 12 and 13, we show that assigning different relaying strategies to the messages from different nodes may result in larger achievable rate regions. Figure 12 shows that in Scenario 1, if the messages of node 1 are relayed with the NDF strategy and the messages of node 2 are relayed with the CF startegy, the resulting achievable rate region will be larger than that of the homogeneous NDF and CF strategies. If the relaying strategies are reversed, it can be seen that the achievable rate region will be smaller than that of both the homogeneous NDF and CF strategies. Since the NDF achievable rate region dominates the CF achievable rate region in Scenario 2 (see Figure 7), it can be seen in Figure 13 that both combinations NDF/CF and CF/NDF provide smaller achievable rate regions than the pure NDF region. Therefore, in Scenario 2 NDF relaying for both source nodes provides the largest achievable rate region. The examples in Figures 12 and 13 suggest that one ought to jointly optimize the power allocation, the resource allocation, and the relaying strategy assigned for each node. Indeed, Figures 12 and 13 suggest that significant gains can be made by doing so. However, the direct formulation of that problem requires the joint allocation of power and the channel resource for each combination of relaying strategies, and hence the computational cost is exponential in the number of source nodes. Furthermore, as the achievable rate region of the overall system is the convex hull of the regions obtained by each combination of relaying strategies, time sharing between different combinations of relaying strategies may be required in order to maximize the achievable rate region. The approach in [18] to the design of relay networks based on orthogonal frequency division multiplexing (OFDM) offers some insight that may lead to more efficient algorithms for joint power, channel resource, and strategy allocation, but the development of such algorithms lies beyond our current scope.

5. CONCLUSION

In this paper, we have shown that the problem of jointly optimal allocation of the power and channel resource in

an orthogonal multiple access relay channel is quasiconvex, and hence simple efficient algorithms can be used to obtain the optimal solution. In addition, we obtained a closed-form expression for the optimal power allocation for a given resource allocation, and we used this expression to significantly reduce the complexity of the algorithm. The numerical results obtained using the proposed algorithm show that significant rate gains can be obtained over those schemes that apply only power allocation and equal channel resource allocation. Finally, we provided an example of the joint allocation of the power, the channel resource, and the relaying strategy, and showed that this has the potential to further enlarge the achievable rate region.

APPENDICES

A. PROOF OF PROPOSITION 1

Assume that the solutions to (6) with $r = r_\alpha$ and $r = r_\beta$ are both greater than certain target rate $C_{1,\text{tar}}$. Let x_α and x_β denote the corresponding optimal values of $\widetilde{P}_{R1}$. Then, we have that

$$\begin{aligned}\frac{r}{2}\min\left\{\log\left(1+\frac{2\gamma_{1R}\overline{P}_1}{r}\right),\log\left(1+\frac{2\gamma_{10}\overline{P}_1+\gamma_{R0}P_{R1}}{r}\right)\right\}&\geq C_{1,\text{tar}},\\ \frac{\hat{r}}{2}\min\left\{\log\left(1+\frac{\gamma_{2r}\overline{P}_2}{\hat{r}}\right),\log\left(1+\frac{2\gamma_{20}\overline{P}_2+\gamma_{R0}P_{R2}}{\hat{r}}\right)\right\}&\geq C_{2,\text{tar}},\end{aligned} \tag{A.1}$$

for $(r = r_\alpha, \widetilde{P}_{R1} = x_\alpha)$ and $(r = r_\beta, \widetilde{P}_{R1} = x_\beta)$. The inequalities in (A.1) can be written as

$$\begin{aligned} f_1(x_\alpha) &\geq g_1(r_\alpha), \qquad f_1(x_\beta) \geq g_1(r_\beta),\\ f_2(x_\alpha) &\geq g_2(r_\alpha), \qquad f_2(x_\beta) \geq g_2(r_\beta),\end{aligned} \tag{A.2}$$

where

$$\begin{aligned} f_1(x) &= \min\{2\gamma_{1R}\overline{P}_1,\ 2\gamma_{10}\overline{P}_1+\gamma_{R0}x\},\\ g_1(r) &= r\left(2^{2C_{1,\text{tar}}/r}-1\right),\\ f_2(x) &= \min\{2\gamma_{2r}\overline{P}_2,\ 2\gamma_{20}\overline{P}_2+\gamma_{R0}(2\overline{P}_R-x)\},\\ g_2(r) &= \hat{r}\left(2^{2C_{2,\text{tar}}/\hat{r}}-1\right).\end{aligned} \tag{A.3}$$

Examining these functions, we observe that $f_1(x)$ and $f_2(x)$ are both concave functions. By differentiating $g_1(r)$ twice with respect to r we obtain

$$\frac{d^2g_1(r)}{dr^2} = \frac{4C_{1,\text{tar}}^2\ln(2)^2 2^{2C_{1,\text{tar}}/r}}{r^3} \geq 0, \tag{A.4}$$

and hence, $g_1(r)$ is a convex function in r. Similarly, we can show that $g_2(r)$ is convex in r.

Now, if we consider $r_\gamma = \mu r_\alpha + \hat{\mu} r_\beta$ and $x_\gamma = \mu x_\alpha + \hat{\mu} x_\beta$, where $\mu \in [0,1]$ and $\hat{\mu} = 1-\mu$, then

$$\begin{aligned} f_1(x_\gamma) &\geq_a \mu f_1(x_\alpha) + \hat{\mu} f_1(x_\beta)\\ &\geq \mu g_1(r_\alpha) + \hat{\mu} g_1(r_\beta)\\ &\geq_b g_1(r_\gamma),\end{aligned} \tag{A.5}$$

where a follows from the concavity of $f_1(x)$ and b follows from the convexity of $g_1(r)$. Similarly, it can be shown that

$$f_2(x_\gamma) \geq g_2(r_\gamma). \tag{A.6}$$

Hence, for any two values of r (namely, r_α and r_β), if there exist values of x (namely, x_α and x_β), such that the conditions in (A.1) are satisfied, then for any value of r that lies between r_α and r_β (namely, $r_\gamma = \mu x_\alpha + \hat{\mu} x_\beta$), there exists a value for x that lies between x_α and x_β (namely, $x_\gamma = \mu x_\alpha + \hat{\mu} x_\beta$) such that the conditions in (A.1) are satisfied. Therefore, the set of values of r for which the solution of the problem in (6) is greater than a certain target rate $C_{1,\text{tar}}$ is a convex set. Hence, the problem in (6) is quasiconcave in r.

B. PROOF OF PROPOSITION 2

We begin by showing that the function $(r/2)\log((1+a/r)(1+bx/r))$ is quasiconcave in the variables r and x, where a and b are nonnegative constants. To do so, we assume that the pairs (r_α, x_α) and (r_β, x_β) satisfy

$$\frac{r}{2}\log\left(\left(1+\frac{a}{r}\right)\left(1+\frac{bx}{r}\right)\right) \geq M, \tag{B.1}$$

where M is a nonnegative constant. We can write (B.1) as

$$f_0(r,x) \geq g_0(r), \tag{B.2}$$

where

$$f_0(r,x) = r+bx, \qquad g_0(r) = \frac{r^2 2^{2M/r}}{r+a}. \tag{B.3}$$

The function $f_0(r,x)$ is a linear function, while the function $g_0(r)$ can be shown to be convex function using the fact that

$$\begin{aligned}&\frac{d^2g_0(r)}{dr^2}\\ &= \left(2a^2 2^{2a/r}\left(r^2(1-\ln(2))^2+2ar\ln(2)(2\ln(2)-1)\right.\right.\\ &\qquad\left.\left.+\ln(2)^2(r^2+2a^2)\right)\right)\times\left((r+a)^3r^2\right)^{-1}\\ &\geq 0.\end{aligned} \tag{B.4}$$

Now,if we consider $r_\gamma = \mu r_\alpha + \hat{\mu} r_\beta$ and $x_\gamma = \mu x_\alpha + \hat{\mu} x_\beta$, where $\mu \in [0,1]$ and $\hat{\mu} = 1-\mu$, then

$$\begin{aligned} f_0(r_\gamma, x_\gamma) &= \mu f_0(r_\alpha, x_\alpha) + \hat{\mu} f_0(r_\beta, x_\beta)\\ &\geq \mu g_0(r_\alpha) + \hat{\mu} g_0(r_\beta)\\ &\geq_a g_0(r_\gamma),\end{aligned} \tag{B.5}$$

where a follows from the convexity of $g_1(r)$. Therefore, the set of pairs (r,x) that satisfy (B.1) is a convex set, and hence the function $(r/2)\log((1+a/r)(1+bx/r))$ is quasiconcave in the variables r and x.

By obtaining its second derivative, it is straight forward to show that $(r/2)\log(1+(2\gamma_{1R}\overline{P}_1/r))$ is concave in r. Since the minimum of a concave function and a quasiconcave function is a quasiconcave function, then we can say that the function

$$\begin{aligned}\frac{r}{2}\min\left\{\log\left(1+\frac{2\gamma_{1R}\overline{P}_1}{r}\right),\ \log\left(1+\frac{2\gamma_{10}\overline{P}_1}{r}\right)\right.\\ \left.+\log\left(1+\frac{\gamma_{R0}\widetilde{P}_{R1}}{r}\right)\right\}\end{aligned} \tag{B.6}$$

is quasiconcave in r and P_{R1}. Similarly, the function

$$\frac{\hat{r}}{2}\min\left\{\log\left(1+\frac{2\gamma_{2r}\overline{P}_2}{\hat{r}}\right),\log\left(1+\frac{2\gamma_{20}\overline{P}_2}{\hat{r}}\right)+\log\left(1+\frac{\gamma_{R0}(2\overline{P}_R-\widetilde{P}_{R1})}{\hat{r}}\right)\right\} \tag{B.7}$$

can be shown to be quasiconcave in r and P_{R1}. Therefore, the problem in (8) is quasiconcave in r and P_{R1}. That is, the set of all pairs (r, P_{R1}) for which the solution of the problem in (8) is greater than a target rate $C_{1,\text{tar}}$, that is, the set of all pairs that satisfy

$$\begin{aligned}\frac{r}{2}\min\left\{\log\left(1+\frac{2\gamma_{1R}\overline{P}_1}{r}\right),\ \log\left(1+\frac{2\gamma_{10}\overline{P}_1}{r}\right)+\log\left(1+\frac{\gamma_{R0}\widetilde{P}_{R1}}{r}\right)\right\}&\geq C_{1,\text{tar}},\\ \frac{\hat{r}}{2}\min\left\{\log\left(1+\frac{2\gamma_{2r}\overline{P}_2}{\hat{r}}\right),\ \log\left(1+\frac{2\gamma_{20}\overline{P}_2}{\hat{r}}\right)+\log\left(1+\frac{\gamma_{R0}\widetilde{P}_{R2}}{\hat{r}}\right)\right\}&\geq C_{2,\text{tar}},\end{aligned} \tag{B.8}$$

is a convex set.

C. PROOF OF PROPOSITION 3

Consider the function

$$f(r,x)=r\log\left(1+\frac{a}{r}+\frac{bcx}{r(r+b+cx)}\right), \tag{C.1}$$

where a, b, and c are positive constants and $(r,x)\in(0,1)\times\mathbb{R}_{++}$. We will avoid the cases where $r=0$ or $r=1$ because these cases correspond to scenarios in which one of the source nodes does not transmit. In those scenarios, the problem is easy to solve because all the relay power and all the channel resource will be allocated to the transmission of the message of the other source node. We will show that $f(r,x)$ is quasiconvex using the the second-order condition for the quasiconvexity which states that [14] for any vector $\mathbf{z}$ such that $\mathbf{z}^T\nabla f=0$, if the function f satisfies $\mathbf{z}^T\nabla^2 f\mathbf{z}<0$, then f is quasiconcave.

For the function f, we denote the gradient by $\nabla f=[f_r,\ f_x]^T$, where $f_w=\partial f/\partial w$. Since $\nabla f\in\mathbb{R}^2$, the subspace orthogonal to ∇f will be a one-dimensional subspace. Since the vector $\mathbf{z}=[-f_x,\ f_r]^T$ is orthogonal to ∇f, then all the vectors in the subspace orthogonal to ∇f are parallel to the vector $\mathbf{z}$. Examining the quantity $\mathbf{z}^T\nabla^2 f\mathbf{z}$, we have that

$$\mathbf{z}^T\nabla^2 f\mathbf{z}=-A\frac{f(r,x)^2}{r^2}+B\frac{f(r,x)}{r}-C, \tag{C.2}$$

where A, B, and C are positive quantities that depend on the constants a, b, and c and the variables r and x. Equation (C.2) can be written as

$$\mathbf{z}^T\nabla^2 f\mathbf{z}=-\left[\left(\sqrt{A}\frac{f(r,x)}{r}-\sqrt{C}\right)^2+(2\sqrt{AC}-B)\frac{f(r,x)}{r}\right]. \tag{C.3}$$

From (C.3), it can be seen that it is sufficient that $2\sqrt{AC}>B$ for the quantity $\mathbf{z}^T\nabla^2 f\mathbf{z}$ to be negative and consequently for the function $f(r,x)$ to be quasiconcave in (r,x). Since both the quantities $2\sqrt{AC}$ and B are positive, we can examine the quantities $4AC$ and B^2. In particular, it can be shown that

$$\begin{aligned}4AC-B^2=4(&\underbrace{5abcxr^2-bcxr^3}+\underbrace{3ab^2r^2-b^2r^3}+\underbrace{abr^3-br^4}\\&+4ab^3cx+10ab^2cxr+2b^3cxr+2a^2c^2x^2r\\&+6a^2b^2r+6a^2br^2+2a^2r^3+2a^2b^3\\&+4bac^2x^2r+4ab^2c^2x^2+b^2r^2cx+2a^2bc^2x^2\\&+4a^2r^2cx+8a^2bcxr+4a^2b^2cx+2arb^3\\&+2b^2c^2x^2r+2b^3c^2x^2)\times r^2b^3c^4(b+r)^2\\&\times(rb+rcx+r^2+ab+acx+ar+bcx)^{-5}\\&\times(b+cx+r)^{-4}.\end{aligned} \tag{C.4}$$

The underbraced terms in (C.4) contain the negative terms in (C.4), each paired with a corresponding positive term. It can be seen that if $a\geq r$, then each of these underbraced terms is nonnegative. Therefore, $a\geq 1$ is a sufficient condition for $4AC>B^2$, and hence is a sufficient condition for the function $f(r,x)$ to be quasiconcave.

By making the substitutions $a=2\gamma_{10}\overline{P}_1$, $b=2\gamma_{1R}\overline{P}_1$, $c=\gamma_{R0}$, and $x=P_{R1}$, the sufficient condition becomes $2\gamma_{10}\overline{P}_1\geq 1$, that is, if the maximum achievable SNR of the direct channel of node 1 is at least -3 dB, then the objective function in (9) is quasiconcave in (r,P_{R1}). Similarly, we can obtain that $2\gamma_{20}\overline{P}_2\geq 1$ is a sufficient condition for the function on the left hand side of the first constraint in (9) to be quasiconcave in $(\hat{r},P_{R2})$. Therefore, the problem in (9) is quasiconcave in (r,P_{R1}) if the maximum achievable SNR of the direct channel of both nodes is at least -3 dB.

D. PROOF OF PROPOSITION 4

Following a similar proof to that in Appendix C, consider the function

$$f(r,x)=r\log\left(1+\frac{a}{r}+\frac{bc(a+r)x}{r(r^2+(a+b)r+c(a+r)x)}\right), \tag{D.1}$$

where a, b, and c are positive constants and $(r,x)\in(0,1)\times\mathbb{R}_{++}$. Define $\mathbf{z}$ to be the vector orthogonal to the gradient subspace of the function $f(r,x)$, that is, $\mathbf{z}=[-f_x,\ f_r]^T$. Examining the quantity $\mathbf{z}^T\nabla^2 f\mathbf{z}$, we have that

$$\begin{aligned}\mathbf{z}^T\nabla^2 f\mathbf{z}&=-\frac{A}{r^2}f(r,x)^2+\frac{B}{r}f(r,x)-C,\\&=-\left[\left(\sqrt{A}\frac{f(r,x)}{r}-\sqrt{C}\right)^2+(2\sqrt{AC}-B)\frac{f(r,x)}{r}\right],\end{aligned} \tag{D.2}$$

where of course A, B, and C are different positive functions of a, b, c, r, and x than those in Appendix C.

From (D.2), it can be seen that it is sufficient that $2\sqrt{AC}>B$ for $\mathbf{z}^T\nabla^2 f\mathbf{z}$ to be negative, and consequently for

the function $f(r,x)$ to be quasiconcave in (r,x). Since both $2\sqrt{AC}$ and B are positive, we can examine

$$\begin{aligned}
&4AC - B^2 \\
&\quad = 4b^3c^4r^5\big(\underbrace{20ba^3r^4 - 3bar^6}+\underbrace{18b^2a^3r^3 - 2b^2ar^5} \\
&\qquad + \underbrace{20bcxa^2r^4 - bcxr^6}+\underbrace{8b^2a^2r^4 - b^2r^6} \\
&\qquad + \underbrace{7ba^5r^2 - br^7}+12a^2r^2b^2c^2x^2 + 2r^5cxb^2 \\
&\qquad + 2r^4c^2x^2b^2 + 4ar^4bc^2x^2 + 2a^2r^4c^2x^2 \\
&\qquad + 2r^4cxb^3 + 10ar^4b^2cx + 2a^2r^6 \\
&\qquad + 24a^2r^3b^2cx + 8ar^3b^2c^2x^2 + 11a^5bcxr \\
&\qquad + 10a^4xcrb^2 + 8a^5r^3 + 2a^6r^2 \\
&\qquad + 16a^2r^3bc^2x^2 + 3a^3xcrb^3 + 37a^4r^2cxb \\
&\qquad + 16a^4x^2c^2rb + 9b^2r^2a^4 + 5a^2xcr^2b^3 \\
&\qquad + 2a^6x^2c^2 + a^2r^2b^4 + 8a^3x^2c^2rb^2 \\
&\qquad + 26a^3b^2cxr^2 + 4a^2br^5 + 4a^2r^5cx \\
&\qquad + 24a^3br^2c^2x^2 + 44a^3br^3cx + 8x^2c^2a^3r^3 \\
&\qquad + 12a^4r^2c^2x^2 + 24a^4r^3cx + 5b^3r^3a^2 \\
&\qquad + ar^5bcx + 16a^3r^4cx + 12a^4r^4 \\
&\qquad + 4ar^3b^3cx + 8a^3r^5 + 2a^4x^2c^2b^2 \\
&\qquad + 4a^6rcx + 8a^5c^2x^2r + 4a^5x^2c^2b \\
&\qquad + 16a^5r^2cx + 5b^3r^2a^3 + 21a^4r^3b\big) \\
&\quad \times (2ar + r^2 + a^2 + ab + rb)^{-2}(cx + r)^{-4} \\
&\quad \times (acx + rcx + ar + rb + r^2)^{-5}.
\end{aligned} \tag{D.3}$$

The underbraced terms of (D.3) contain the negative terms in (D.3), each paired with a corresponding positive term. It can be seen that if $a \geq r$, each of these underbraced terms is nonnegative. Therefore, $a \geq 1$ is a sufficient condition for $4AC > B^2$, and hence for the function $f(r,x)$ to be quasiconcave.

Making the substitutions $a = 2\gamma_{10}\overline{P}_1$, $b = 2\gamma_{1R}\overline{P}_1$, $c = \gamma_{R0}$, and $x = P_{R1}$, the sufficient condition becomes $2\gamma_{10}\overline{P}_1 \geq 1$. That is, if the maximum achievable SNR of the direct channel of node 1 is at least -3 dB, then the objective function in (11) is quasiconcave in (r, P_{R1}). Similarly, we can obtain that $2\gamma_{20}\overline{P}_2 \geq 1$ is a sufficient condition for the function on the left hand side of the first constraint in (11) to be quasiconcave in $(\hat{r}, P_{R2})$. Therefore, the problem in (11) is quasiconcave in (r, P_{R1}) if the maximum achievable SNR of the direct channel of both nodes is at least -3 dB.

ACKNOWLEDGMENTS

This work was supported in part by a Premier's Research Excellence Award from the Government of Ontario. The work of the second author is also supported in part by the Canada Research Chairs program. A preliminary version of part of this work appears in the *Proceedings of the 2008 International Symposium on Information Theory*.

REFERENCES

[1] G. Kramer, M. Gastpar, and P. Gupta, "Cooperative strategies and capacity theorems for relay networks," *IEEE Transactions on Information Theory*, vol. 51, no. 9, pp. 3037–3063, 2005.

[2] G. Kramer and A. J. van Wijngaarden, "On the white Gaussian multiple-access relay channel," in *Proceedings of the IEEE International Symposium on Information Theory (ISIT '00)*, p. 40, Sorrento, Italy, June 2000.

[3] L. Sankaranarayanan, G. Kramer, and N. B. Mandayam, "Capacity theorems for the multiple-access relay channels," in *Proceedings of the 42nd Annual Allerton Conference on Communication, Control and Computing*, Monticello, Ill, USA, September-October 2004.

[4] L. Sankaranarayanan, G. Kramer, and N. B. Mandayam, "Cooperation vs. hierarchy: an information-theoretic comparison," in *Proceedings of the IEEE International Symposium on Information Theory (ISIT '05)*, pp. 411–415, Adelaide, Australia, September 2005.

[5] T. M. Cover and J. A. Thomas, *Elements of Information Theory*, John Wiley & Sons, New York, NY, USA, 1991.

[6] A. Sendonaris, E. Erkip, and B. Aazhang, "User cooperation diversity—part I: system description," *IEEE Transactions on Communications*, vol. 51, no. 11, pp. 1927–1938, 2003.

[7] A. Sendonaris, E. Erkip, and B. Aazhang, "User cooperation diversity—part II: implementation aspects and performance analysis," *IEEE Transactions on Communications*, vol. 51, no. 11, pp. 1939–1948, 2003.

[8] J. N. Laneman, D. N. C. Tse, and G. W. Wornell, "Cooperative diversity in wireless networks: efficient protocols and outage behavior," *IEEE Transactions on Information Theory*, vol. 50, no. 12, pp. 3062–3080, 2004.

[9] L. Sankaranarayanan, G. Kramer, and N. B. Mandayam, "Hierarchical sensor networks: capacity bounds and cooperative strategies using the multiple-access relay channel model," in *Proceedings of the 1st Annual IEEE Communications Society Conference on Sensor and Ad Hoc Communications and Networks (SECON '04)*, pp. 191–199, Santa Clara, Calif, USA, October 2004.

[10] K. Seddik, A. K. Sadek, W. Su, and K. J. R. Liu, "Outage analysis and optimal power allocation for multinode relay networks," *IEEE Signal Processing Letters*, vol. 14, no. 6, pp. 377–380, 2007.

[11] T. M. Cover and A. El Gamal, "Capacity theorems for the relay channel," *IEEE Transactions on Information Theory*, vol. 25, no. 5, pp. 572–584, 1979.

[12] L. Lai, K. Liu, and H. El Gamal, "The three-node wireless network: achievable rates and cooperation strategies," *IEEE Transactions on Information Theory*, vol. 52, no. 3, pp. 805–828, 2006.

[13] S. Serbetli and A. Yener, "Relay assisted F/TDMA ad hoc networks: Node classification, power allocation and relaying strategies," *IEEE Transactions on Communications*, vol. 56, no. 6, pp. 937–947, 2008.

[14] S. Boyd and L. Vandenberghe, *Convex Optimization*, Cambridge University Press, Cambridge, UK, 2004.

[15] P. A. Anghel, M. Kaveh, and Z.-Q. Luo, "An efficient algorithm for optimum power allocation in a decode-and-forward cooperative system with orthogonal transmissions," in *Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP '06)*, vol. 4, pp. 685–688, Toulouse, France, May 2006.

[16] Y. Liang and V. V. Veeravalli, "Gaussian orthogonal relay channels: optimal resource allocation and capacity," *IEEE Transactions on Information Theory*, vol. 51, no. 9, pp. 3284–3289, 2005.

[17] A. Wyner and J. Ziv, "The rate-distortion function for source coding with side information at the decoder," *IEEE Transactions on Information Theory*, vol. 22, no. 1, pp. 1–10, 1976.

[18] T. C.-Y. Ng and W. Yu, "Joint optimization of relay strategies and resource allocations in cooperative cellular networks," *IEEE Journal on Selected Areas in Communications*, vol. 25, no. 2, pp. 328–339, 2007.

Hindawi Publishing Corporation
EURASIP Journal on Advances in Signal Processing
Volume 2008, Article ID 793126, 14 pages
doi:10.1155/2008/793126

Research Article

Resource Sharing via Planed Relay for HWN*

Chong Shen, Susan Rea, and Dirk Pesch

Centre for Adaptive Wireless Systems, Department of Electronic Engineering, Cork Institute of Technology, Ireland

Correspondence should be addressed to Chong Shen, chong.shen@cit.ie

Received 23 October 2007; Revised 22 February 2008; Accepted 1 April 2008

Recommended by J. Wang

We present an improved version of adaptive distributed cross-layer routing algorithm (ADCR) for hybrid wireless network with dedicated relay stations (HWN*) in this paper. A mobile terminal (MT) may borrow radio resources that are available thousands mile away via secure multihop RNs, where RNs are placed at pre-engineered locations in the network. In rural places such as mountain areas, an MT may also communicate with the core network, when intermediate MTs act as relay node with mobility. To address cross-layer network layers routing issues, the cascaded ADCR establishes routing paths across MTs, RNs, and cellular base stations (BSs) and provides appropriate quality of service (QoS). We verify the routing performance benefits of HWN* over other networks by intensive simulation.

1. INTRODUCTION

Time Division Multiple Access (TDMA)-based digital cellular standard global system for mobile (GSM) was first deployed in 1990 with a new 900-MHz band. However, due to uneven nature of the time-varying spatial distribution [1], network performance metrics are not sufficient for today's wireless network where more ad hoc features are being introduced.

To effectively manage problems stated above, we propose to combine the advantages of different networks so that the Mobile Terminal (MT) can utilise an optimised MANET, the base-station-oriented network (BSON) and the relay services. Figure 1 presents hybrid wireless network with relay nodes (HWN*), the relay nodes (RNs) of core network compose a mesh-like structure connected to the internet protocol (IP) networks through RN gateways, while base stations (BSs) are connected to the IP networks via switches. In rural places without infrastructure support as indicated in Figure 1, two MTs may communicate directly, or through intermediate MTs. When an MT transmits packets to a BS through RNs, the RNs extend the signalling coverage of BSON thus we can expect an enhanced resource-sharing performance.

An adaptive distributed cross-layer routing (ADCR) algorithm is proposed for HWN* based on [2] using the minimal number of hops and considering routing model dynamic switching to reduce latency, preserve communications, deliver good overall throughput/per node throughput, and extend the HWN* coverage. A cross-layer network design [3] that seeks to enhance the system performance by jointly designing MAC and NETWORK layers is adopted. We analyse in design stage the theoretical cellular network media access capacity, multihop traffic relaying issues, and inter network traffic handovers [4]. The cascaded ADCR then includes three subpacket transmission modes labeled as one-hop ad-hoc transmission (OHAHT) for point-to-point ad hoc direct communication, multihop combined transmission (MHCT) for radio resource relaying using fixed RNs or MTs, and cellular transmission (CT) for traditional cellular service. In rural places without infrastructure RN support, the MHCT transmission mode can be implemented on selforganised ad hoc nodes for supporting multihop communication as long as: (i) The resource of relaying MTs is contention-free, (ii) the migration range of relaying MTs is limited, and (iii) the speed changes of relaying MTs in sampling times have limited influence on routing.

The paper begins with a heterogenous wireless networks RN incorporation discussion, including the comparison work between proposed HWN* framework stage I and HWN* framework stage II. We present two pre-engineered RNs positioning algorithms in Section 3. In Section 4, we

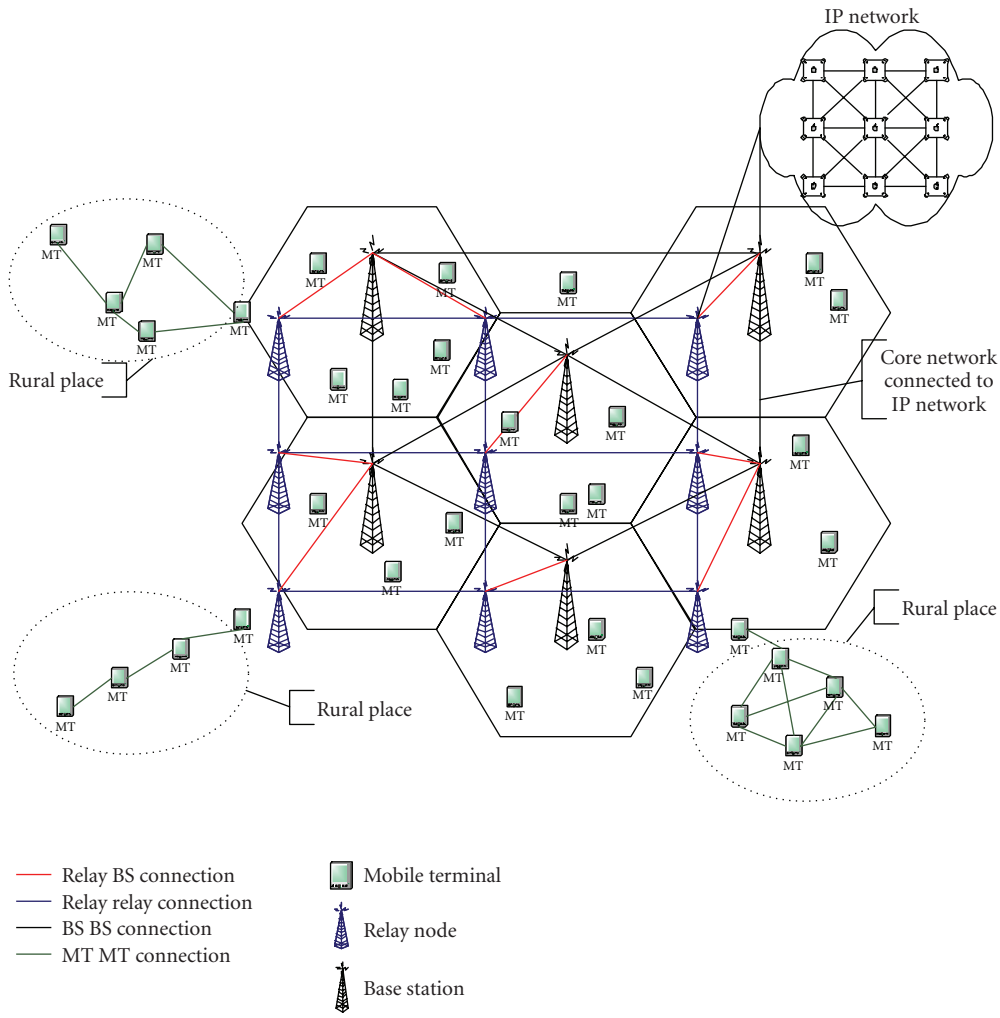


Figure 1: The hybrid wireless network with fixed relay stations.

discuss three traffic transmission modes with emphasis on MHCT mode in copy with newly included packet relaying environment. The ADCR performance of the HWN* under various scenarios is evaluated in Section 5 to address network capacity, per MT throughput, access speed, and end-to-end delay. Finally in Section 6, conclusions are made with future research outlook.

2. HETEROGENOUS WIRELESS NETWORKS

The further balance of radio resource in heterogeneous networks or hybrid wireless networks requires assistant equipment functioned-like internetwork switcher, thus we introduced a new node structure (RN), and further divided it into *heterogeneous RN* that uses different radio access technology (RAT) with common or different sets of transmission resources for its links and *homogeneous relay node* that uses the same radio access technology and mode in a common set of transmission resources for its entire links. For example, the IST-WINNER [5] project proposes to share the same RAT with BSs, RNs, and MTs to realise a dynamic spectrum usage. Multiple noninterfering relay frequencies operate in parallel through the use of intelligent radios. The spectrum where an RN operates can be leased for a limited time depending on network status. The spectrum on which it is operating is reclaimed when network performance improves. Two RNs operating on noninterfering spectrums form a network relay link with multiple orthogonal bands. Multiple nodes within range of each other may also transmit simultaneously on different channels without relying on a media access protocol or distributed scheduling algorithm to resolve contention.

Focus on different design objectives, the *iCAR* [6] is derived from existing cellular networks and enables the network to achieve theoretical capacity through adaptive traffic load balancing. The *SOPRANO* [7] is a scalable architecture that assumes the use of asynchronous code-division multiple access (CDMA) with spreading codes to support high-data-rate internet and multimedia traffic. It is similar to iCAR other than IP network support and cross network connection methods. We summarise in Table 1 the main research improvements from the HWN* stage I [4] to the HWN* stage II. The comparison between the

Table 1: Research improvement of HWN* framework stage II.

project	HWN* framework stage I	HWN* framework stage II
Main objectives	Incorporate a MANET to increase system capacity while realising differentiated QoS services	Stage I + Investigation on places without infrastructure support
Basic infrastructure	BSON, BSON with RN, MANET with RN	BSON, BSON with RN, MANET with RN and MANET
Routing issues	BS switch and RN assisted traffic diversion	Cascaded and distributed routing with three transmission modes
Mode movement issues	Attractor points model	Costudy of user movement model and RN placement algorithm
Congestion control	QoS-based session congestion control algorithm	QoS-based session congestion control algorithm
RN positioning scheme	Fix point RN positioning	RN positioning considering node movement pattern
Load balance	QoS-based multihop load balancing	QoS-based multihop load balancing
Call admission	The BS coordinated admission	Distributed session admission algorithm

iCar, multipower architecture for cellular network (MuPAC), hybrid wireless network (HWN) without RN support, WINNER, SOPRANO, and MCN can be found in [4, 8] with the identification of technologies used.

Consider a cellular handover scenario in Figure 2 where MT A is currently connected to MT B and is moving out of Cell 1 into Cell 6. A request for a BS handover will be sent as soon as the power level by MT A goes below a certain threshold (trajectory indicated by red dotted line). A successful handover will take place within a few hundred milliseconds depending on speed before the received power from BSs reaches an unacceptable level. When MT A arrives in Cell 6, if the congestion persists in cell 6 for a period of time during which the MT moves farther away from the other neighbouring cell border, thus causing the received power level from BS A to fall below the acceptable level, handover will fail and the call will be permanently terminated.

However, in MHCT mode of HWN*, the data session does not have to be dropped even though the congestion in Cell 6 persists. For example, when MT A moves into the congested Cell 6, apart from trying cellular connections, it also associates itself with an RN using either ad hoc frequency or cellular frequency, then the RN may continue transmission with any BS via the multihop relaying structure and the relaying path can be also extended to the area with no cellular coverage. For example, the routing path for an MT in rural place can be even from MT → MTs → core network; and the corresponding frequencies used can be ad hoc frequency → ad hoc frequencies → either ad hoc frequency or cellular frequency. In addition, OHAHT of point-to-point ad hoc communications can be another routing mechanism option to further balance traffic load. The simulation results presented in [4] have already proven that inter network traffic management can significantly improve the grade of service, reduce the traffic blocking probability, while maintaining the QoS.

The relay concept extends service range, optimises cell capacity, minimises transmit power, covers shadowed areas, supports inter network load balancing, and supports MANET routing. Theoretically, both the HWN* system capacity and the transport capacity per MT, when compared to a cellular network, should be improved because the RNs provide relay capability as the substitution of a poor-quality single-hop wireless link with a better-quality link being encouraged whenever possible. Also a higher end-to-end data rate could be obtained if an MT had two simultaneously communicating interfaces.

Using three scaling approaches, we can implement network/simulation dimensioning and estimate how many RNs should be deployed when the number of MTs changes. The three parameters are the number of RNs m, the number of MTs n, and the system capacity C. The asymptotic scaling for the per user throughput as n becomes large is

$$m \leq \sqrt{\frac{n}{\log n}}. \tag{1}$$

The per user throughput is of the order $C/\sqrt{n/\log n}$ and can be realised by allowing only ad hoc communications which do not necessarily need RN support, when

$$\sqrt{\frac{n}{\log n}} \leq m \leq \frac{n}{\log n}. \tag{2}$$

The order for the per user throughput is Cm/n, therefore the total additional bandwidth provided by m RNs is effectively shared among n MTs. Finally, when

$$\frac{n}{\log n} \leq m, \tag{3}$$

the order of the per user throughput is only $C/\log n$ which implies that further investments in relay nodes will not lead to an improvement in throughput and bandwidth optimisation.

3. RN-PLACEMENT ALGORITHMS

We explore the relay node placement and HWN* initialisation problem in this section. The network spectral efficiency was taken by [9] as the objective to optimise RN positioning. The paper made the assumption that the quality

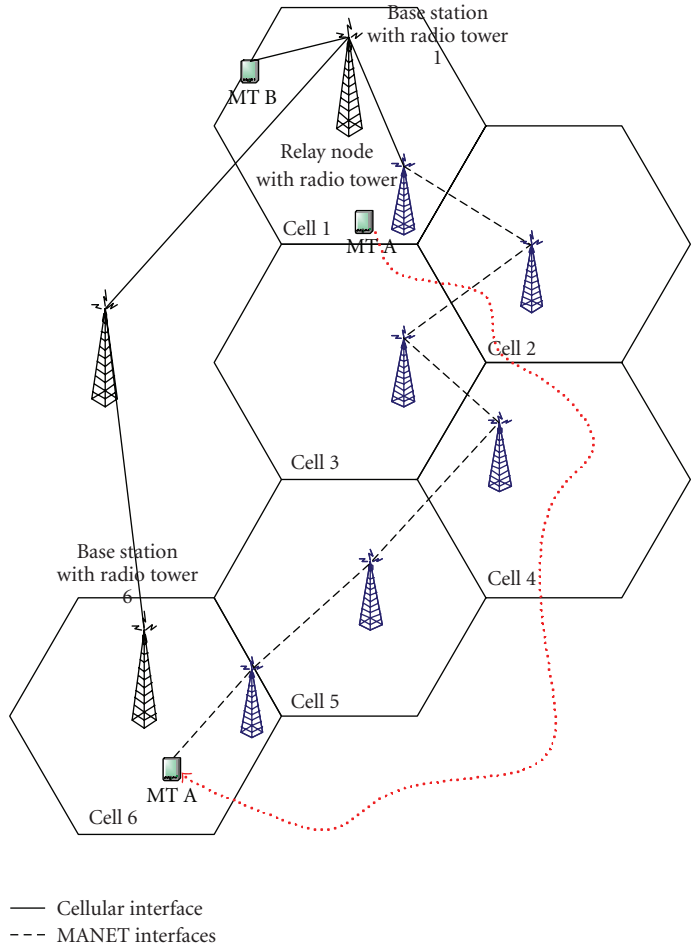


Figure 2: Multihop combined transmission example of cellular resource relaying using fixed RNs.

on the links connecting BS ↔ RN is always better than the link between RN ↔ RN. This assumption can be satisfied by establishing line-of-sight (LOS) links between BS and RN or by designing links that enhance the antenna gains. However, the solution imposes extra difficulty on network planning by complicating transceiver design. In this section, two RN positioning algorithms are proposed, which are packing-based RN placement and heuristic RN placement considering user movement behaviours. The algorithms implementation is to use a minimum number of RNs that enable the relaying of maximum traffic under the media contention from both cellular and ad hoc perspectives.

It is well known from planar geometry that to cover a two-dimensional district with equal-sized circles, the best possible packing solution can be obtained by surrounding each circle by six circles as shown in Figure 3 left. But to have connections between the RNs, an overlap between relay cells is required. We therefore consider a situation where the location of the RNs is centered with maximum coverage. The deployments shown in Figure 3 (left side) are two examples of such pre-engineered approach with a number of RNs in the HWN*. The first deployment tries to cover the entire area while the second one tries to cover densely populated regions.

Heuristic RN placement that we devised has a straightforward design philosophy based on two most important factors, which are user movement behaviour and bandwidth utilisation. By imposing such a plan, we can improve the availability of MTs at disadvantaged locations and enlarge network dimensioning possibility. It is first assumed that RNs can acquire *SIR* information via local estimation according to the distance. The RN positioning is formulated as a constrained optimisation problem, of which the goal is to maximise the overall network throughput and per node throughput so that majority MTs are better served with guaranteed QoS. The attractor points mobility model deployed on MTs uses macro- and microcontrols to improve user movement experiences, it may be not practical to calculate each MT's trajectory, but probabilities of user reaching a set of frequently visited points can be useful. Coincidentally, the hottest areas are places where most media contention happens, and RN can be located in these points to mitigate the contention. The next step of the heuristic algorithm is to decide the number of RNs needed in solving bandwidth contenting with guaranteed QoS. As shown in Figure 4, after getting traffic load information, the RN number used for further simulation studies is actually estimated through network dimensioning analysis discussed

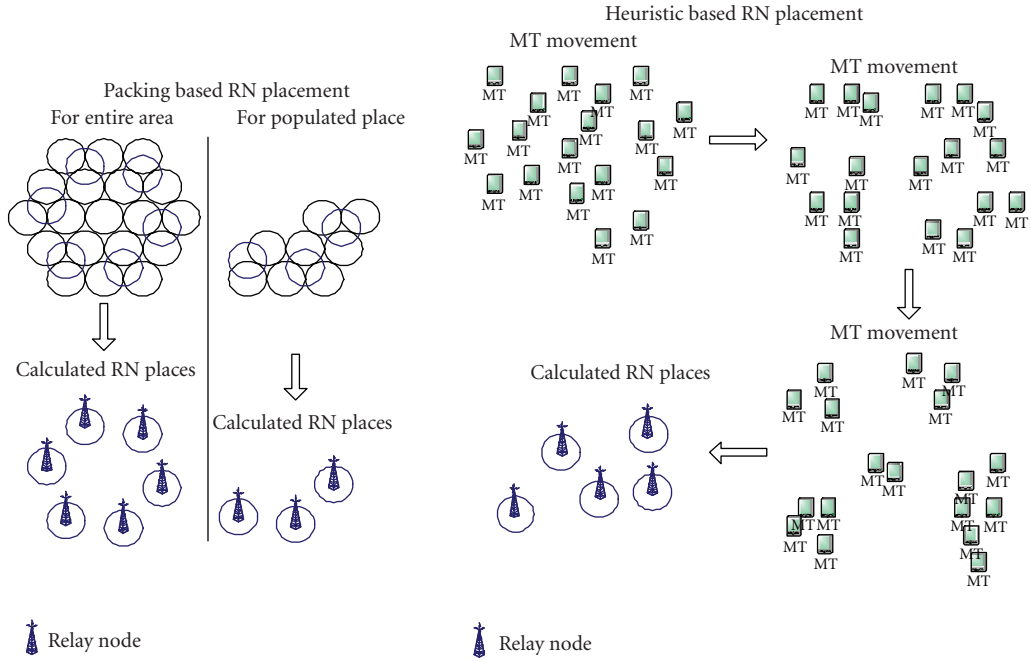


Figure 3: Packing-based RN placement and heuristic RN placement.

in Section 2. The migration experiments are carried out to produce a set of candidate points. A hard distant limit δ is introduced and if distance between one candidate point and any BS is smaller or equal to δ, this point will be eliminated from final list.

The HWN*, after RN placement, is then formed in two stages, which are serving RN, BS association stage, and route identification stage. More details on network formation can be found in [4].

4. ADAPTIVE DISTRIBUTED CROSS-LAYER ROUTING

The QoS flows can consume all the bandwidth on certain links, thus creating congestion for, or even starvation of, best effort sessions. Statically, partitioning the link resources can result in low network throughput if the traffic mix changes over time. Thus, a mechanism that dynamically distributes link resources across traffic classes based on the current load conditions in each traffic class is critical for performance. By proposing a cascaded adaptive distributed cross-layer routing (ADCR) for HWN*, we discourage applications from using any route that is heavily loaded with low-priority traffic. Traditional routing strategies that use global state information are not considered. Problems associated with maintaining global state information and the staleness of such information are avoided by having individual MTs infer the network states based on route discovery statistics collected locally, and perform traffic routing using this localised view of the network QoS state. Each application, categorised by service class with the choice of three possible transmission modes, maintains a set of candidate paths to each possible destination and routes flows along these paths. The selection of the candidate paths is a key issue in localised routing and has a considerable impact on how the ADCR performs. The high-priority traffic is given high priority in accessing comparatively expensive cellular resource, while low-priority traffic tries to access low-cost ad hoc resource. Per MT bandwidth is used as the only metric for route local statistics collection since it is one of the most important metrics in QoS routing, furthermore, important metrics such as end-to-end delay, jitter can be expressed as a function of the bandwidth.

We divide traffic sessions into simple service classes which are high-profile users (HPUs), normal-profile users (NPUs), and low-profile users (LPUs). Principally, HPUs get the best QoS, next comes NPUs with smaller medium access opportunities. LPUs are a best-effort class with unused medium resources by other classes. HPUs have the highest access priority in any communication modes of HWN*, and traffic admission of NPUs and LPUs has to consider ongoing HPUs sessions. The NPUs are configured to have a higher probability than LPUs in terms of resource acquisition and this probability is decided by an association level (AL) set. In case of network congestion, CT mode may temporarily become unavailable to NPUs when HPUs are not fully accommodated, while LPUs sessions may be only granted MHCT and OHAHT mode access to mitigate network congestion, reduce transmission delay, and improve per MT throughput. More details of resource acquisition, QoS-based media access control, traffic class coordination, and traffic class association were explained in [4].

The RN has the right to reserve QoS-guaranteed free channels for packet transmission and it maintains a status table that refers to other RNs and it provides information on changing busy conditions or relay failure. The purpose of bandwidth reservation is to let RNs that receive the relaying

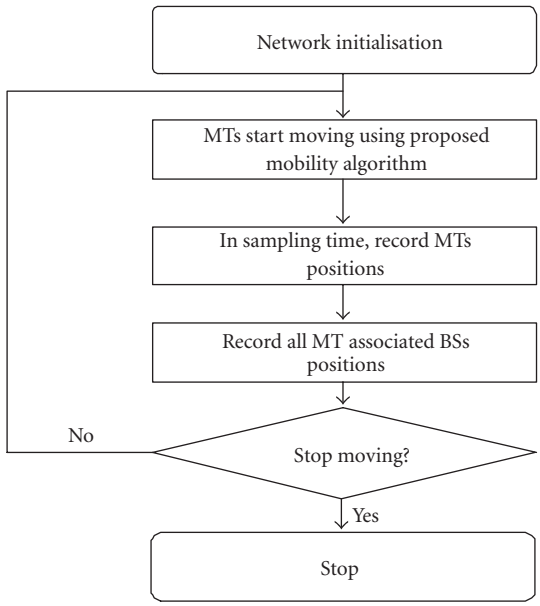


Figure 4: Flowchart of heuristic RN placement.

discovery command check if they can provide the bandwidth required for the connection.

To avoid having higher traffic classes being influenced by lower traffic classes in terms of queueing delay, we place a waiting time limitation on each traffic class and force starving packet switch transmission model [4]. A traffic flow maintains two queues: a slot queue and a packet queue, and we decouple slot queue for traffic class identification from packet queue for transmission. Start and finish tags are associated with slots but not packets. When a packet arrives for a flow, it gets added to the packet queue, and a new slot is added to the slot queue. Corresponding start and finish tags are assigned to the new slot. The way to raise priority in slot queue is that the packets related to a high profile have shorter backoff time to increase the probability of early medium access. As for the status table maintenance, information flooding is restricted to a limited scope. Once a positive acknowledgment message is confirmed by requesting RN, the relay paths will not be changed unless resource contention happens. Given the fact that maintaining global RNs channel status in each RN slows down RN response time, we only require each RN update neighbouring RNs' information, periodically.

The cascaded ADCR scheme includes three subpacket transmission models, which are the OHAHT, the MHCT, and the CT as illustrated in Figure 5. The communication commands are defined as

(i) ACK/ACCEPT/REJECT/REJHO for the message-delivery acknowledgment, packet acceptance, packet rejection, and after-rejection handover request.

(ii) SEARCH/SETUP/DATA/BREAK for destination node finding, new connection establishment, packet delivery, and connection teardown.

(iii) MOS for MT to choose adaptive transmission mode.

(iv) FAIL used to acknowledge any failure on RN or MT.

(v) LREQ to request a label during the routing, The label is a short, fixed-length identifier. Multiple labels can identify a path or connection from the source MT to the destination MT. The structure of a label message contains flag, flow, cost, traffic class, mobility information, and time tO lIVE (TTL).

(vi) LREP to request a label replay during the label routing in MHCT model.

Time-sensitive multimedia applications have restrictions on end-to-end transmission delay, while FTP data transfers need a minimum guarantee on packet losses. The ADCR should therefore consider differentiated QoS issues while guaranteeing HPUs that agree to pay more than NPUs and LPUs. However, due to the high priority of premium traffic, the global network behaviour as a consequence of this service class, including routing and scheduling of premium packets, may impose significant influences on traffic of other classes. These negative influences, which could degrade the performance of low-priority classes with respect to some important metrics such as the packet loss probability and the packet delay, are often called the interclass effects. To reduce the interclass effects, we proposed in [4] a mechanism based on association level (AL) calculation for load balancing of different service classes. The AL is a set of parameters monitoring channel availabilities, an AL that scores higher than the threshold means that the channels are already occupied by ongoing sessions. The simulation results demonstrated that the proposed mechanism distributes the premium bandwidth requirements more efficiently, and the traffic is better organised and balanced before routing. Figure 5 also presents corresponding process of an MT's association with its serving BS and RN, and simplified ADCR algorithm. As presented in script, if the source MT continues transmitting directly until the SIR falls to a certain level, the traffic re-routing or handovers will be initiated. In rerouting, the model selection priority for HPUs is CT > MHCT > OHAHT, while priorities for NPUs and LPUs are MHCT > CT > OHAHT and OHAHT > MHCT > CT, separately. Also, inter- and intranetwork handover triggers are discussed in paper [4].

4.1. One-hop ad-hoc transmission

In OHAHT, the requesting MT first broadcasts SEARCH messages to every node in its transmission range including its associated RN and BS. For example, MT A in Figure 5 broadcasts SEARCH messages, if the destination MT B is within its transmission range and there is no ad hoc-based media contention between MT A and MT B, MT B can respond to MT A with an ACK message. Once MT A confirms the acknowledgment, it starts a connection SETUP session immediately.

```
BS & RN Association()
{
For N MTs in a cellular cell, 0 < i < N;
SIR_i = Received signal quality evaluation of MT i from the serving BS;
For N MTs in a cellular cell, 0 < i < N;
TTL = 1;
/* Receive neighbouring information from surrounding RNs*/
For N MTs, 0 < i < N, M RNs, 0 < j < M;
SIR_ij = Received signal quality evaluation of MT i from surrounding RNs;
Sort SIR_ij in descending order from high SIR to Low SIR;
Associated RN = the RN with Max(SIR_ij);
}
ADCR Routing()
{
DB(i) = Node i's distance from serving BS;
DR(i) = Node i's distance from serving RN;
/* Identify which traffic class the packet belongs to, HPU, NPU or LPU*/
Traffic_Class_Discovery ();
/* Individual packet routing with three sub models, OHAHT, MHCT and CT*/
One_hop_adhoc_transmission();
Multi_hop_combined_transmission();
Cellular_transmission();
Node i is scheduled to initiated a packet transmission at time T(k);
switch(service_class)
{
case HPU():
/* Evaluate QoS requirement and urgency based on weighted calculations */
Evaluation();
/*Check media access constraints for three transmission models*/
Media_check();
/*Try to use the transmission models in order of CT, MHCT then OHAHT*/
HPU_routing();
break;
case NPU():
Evaluation();
Media_check();
/*Try to use the transmission models in order of MHCT, CT then OHAHT*/
NPU_routing();
break;
case LPU():
Evaluation();
Media_check();
/*Try to use the transmission models in order of OHAHT, MHCT then CT*/
LPU_routing();
break;
}
```

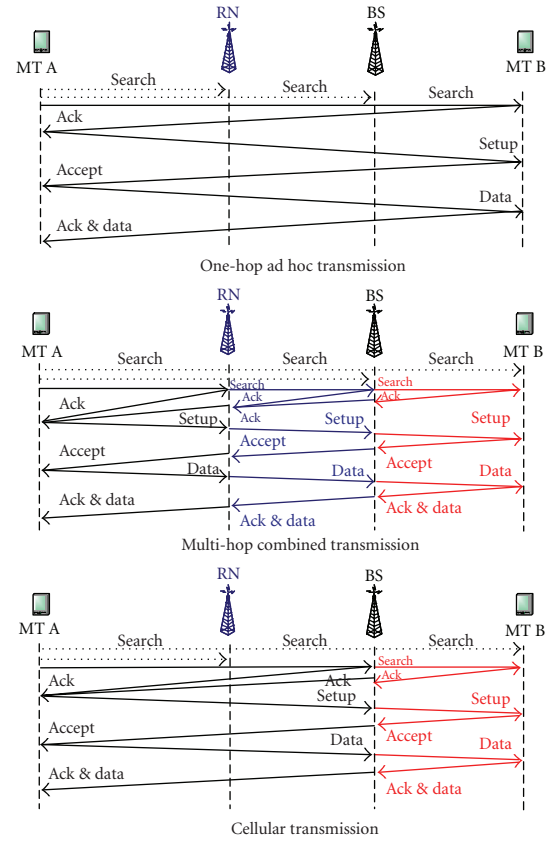


Figure 5: Computerised ADCR algorithm and simplified transmission modes illustration.

4.2. Multihop combined transmission

The MHCT can involve RNs acting as intermediate nodes for message relaying. Figure 5 shows the connection setup process for communication between MT A and MT B via the RN infrastructure. MT A first broadcasts SEARCH messages to every node to find MT B. After the SEARCH session, MT A may find that the cellular resources can be borrowed through RNs by receiving three ACK messages from the serving BS of MT B, RNs, and the MT B. The positive acknowledgment requires MT B to send an ACK to its serving BS, then the serving BS sends an ACK to the RN infrastructure and finally the RNs feedback the ACK to MT A. Once the positive ACK is confirmed, the MT A starts a connection SETUP from $MTA \rightarrow RN$, then $RN \rightarrow BS$, and finally $BS \rightarrow MTB$. The DATA-transmission process follows the same packet delivery route, and further route discovery is prohibited to reduce the signalling overhead.

The label routing concept [10] originated from ATM network is introduced to MHCT mode since RN switching provides faster packet forwarding than routing because its operation is relatively simple. The label switching protocol uses signalling protocol distribute labels and set up new route after the path is computed by the routing module. This requires that the path is pre-established with signalling before it can be used. In reactive MHCT mode with frequent topology changes on both sender and receiver, a high rate of path setup and tear down signaling may occur. It simply can not use separate signalling to set up a new route. Instead, the path finding process dynamically initialised by the LREQ packet carrying a unique label and flow information, where low-path setup delay is guaranteed. The flow information

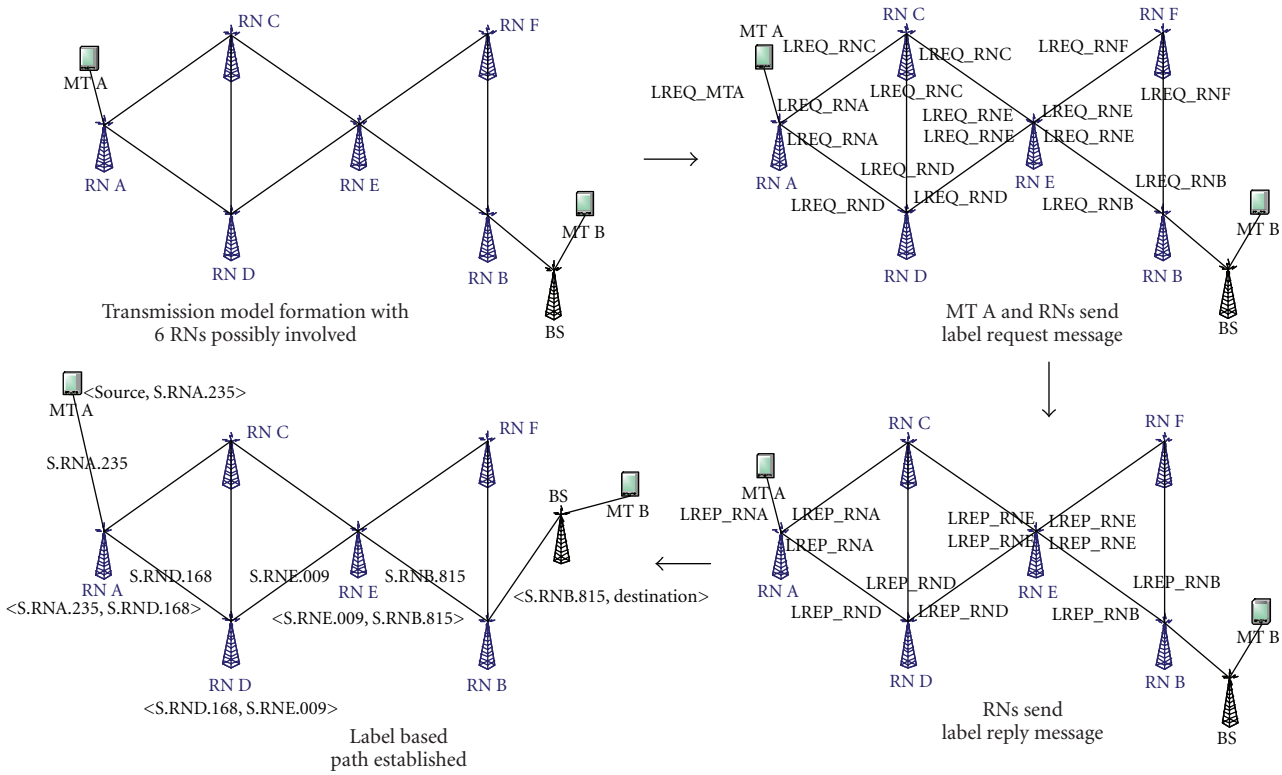


Figure 6: Label routing illustration.

contains source address and a flow number is chosen by the source node since the default source address is assumed to be unique.

In finding the destination MT, the source MT creates an LREQ message in which the packet contains IDs, sequence number, and service class of the source MT. This packet also contains traffic flow information, a broadcast ID, and a hop count that is initialised to zero. All RNs that receive this message will increment the hop count. If an RN does not have any information about the destination node, it will record the neighbour's ID where the first copy of LREQ is from and send this LREQ to its neighbours. LREQs from the same node with the same broadcast ID will not be processed more than once. Figure 6 gives an example of label routing in MHCT. In this example, there are eight nodes with duplex connection link.

The MT A first creates an LREQ message and sends it out to its associated RN. Figure 6 illustrates the propagation of LREQ across the RNs and the reverse path at every RN. The reverse path entry is created for the transmission of the reserved label for this path. This label is embedded in the label-reply message LREP. The reserve path entry will be maintained long enough for the LREQ to traverse the path and for RNs to send an LREP to the source MT. Once a path is found in the relay structure, the source MT will check the sequence number (SEQ) of the destination MT in the current path in order to avoid old path information. It should be at least as great as the value entry in the LREQ. Otherwise, the existing path in the table will be discarded. If $SEQ \geq SEQ_{LREQ}$, it will also check in the current path whether the QoS requested by the source MT has been satisfied. If not, this request will be discarded. If the source MT still can not find the destination MT B, MT A will increment the hop count in the LREQ by one and then broadcast it to its neighbors. Any duplicated LREQ with same source node ID and same broadcast ID will be discarded. Normally relay-based label routing should have a maximum hop count. However, there is no energy constraints and node mobility issues in our relay infrastructure, thus theoretically any hop count threshold can be possible. We specify the hop count in LREQ as not being larger than 10 as a simulation limitation to avoid computation complexity and if the sender of an LREQ does not receive the reply message, each node only resends the LREQ once for each connection request.

The RN only creates an LREP with the total hop count of this path if hop count, sequence number, and path QoS are all acceptable, the new sequence number of the destination MT is the largest one between SEQ and SEQ_{LREQ}, the best QoS, and a label from its label pool. Then this LREP will be sent back to the source MT along the reverse path entry. The third plot in Figure 6 shows the propagation of the LREP along the reserve paths. Note that both RN C and RN F fail to send the LREP due to hop count, sequence number or QoS issues. The path between the source MT and the destination MT is composed of multiple segments and all data packets are relayed by these segments. Each segment is a real connection between two nodes and labeled by the sending-side node of the LREP in this segment. For example in the the path

Table 2: Characteristics of QoS differentiated users.

	Low-profile user	Normal-profile user	High-profile user
Portion	Voice 20% Web 10% Video 5%	Voice 15% Web 8% Video 10%	Voice 10% Web 7% Video 15%
	Voice dwell/session time: 60 s/120 s	Web dwell/session time: 120 s/trace	Video dwell/session time: 120 s/240 s

$MTA \leftrightarrow RNA \leftrightarrow RND \leftrightarrow RNE \leftrightarrow RNB \leftrightarrow MTB$ showed in the last plot of Figure 6, RNs A, D, E, and B set up the labels of the segments between A and D, D and E, and E and E, respectively. MT A and RN A, MT B, RN B, and its associated BS are the other two segments. Since the topology of the relay structure is meshed, the source MT can receive more than one LREP. There is a hop count field in the LREP. This field records the total number of hops of the path. The source MT will choose the smallest hop count from the LREPs in the specific limited time. All LREPs that are received after this time threshold will be ignored. And if some available LREPs have the same hop count, the path that has the largest destination sequence number, which means it is the latest path, will be chosen.

The MHCT mode can be also implemented in multihop ad hoc transmissions in copy with rural environment without infrastructure node support. The basic mechanism is almost the same except MT replaces fixed RN and acts as traffic switching nodes. The source MT first tries to establish a connection destination node. If there is no path which can reach the destination node in its local label routing table, or the mobility constrains of MT relaying are violated, the source MT will initiate another path discovery until TTL reaches.

4.3. Cellular transmission

The last plot in Figure 5 shows the connection setup of CT model between MT A and MT B via cellular BSs. MT A first broadcasts SEARCH messages to every node to find MT B. After the SEARCH session, the MT A finds that it is able to communicate with MT B directly via BSs, while the connection can be setup through a virtual wireless backbone. The positive acknowledgment of a connection requires MT B to send an ACK to its serving BS, then the serving BS informs the serving BS of MT A or the BS feedbacks the ACK to MT B when both MT A and MT B share the same serving BS. Once the positive ACK is confirmed, MT A starts connection SETUP from $MTA \rightarrow BS$, then $BS \rightarrow BS$, and finally $BS \rightarrow MTB$. The DATA transmission process follows the same packet-switched delivery route. Dynamic channel allocation can be realised in a distributed manner given that the channel usage does not break the two-channel interference constrains [11] which are cosite constraint where there are minimum channel separations within a cell and non-cosite constraint where minimum channel separation between two adjacent BSs is kept.

5. SIMULATION

We present various schemes and results of the simulation that have been implemented for the ADCR in this section. The OMNET++ simulator [12] is used and we generalise all video streaming as real-time services, while web services are referred to as nonreal-time services. Table 2 presents the default QoS profile used consisting of 30%, 64 Kbps streaming video, 45% general voice calls, and 25% nonreal-time web services. The service request portion is distributed and shared among HPUs, NPUs, and LPUs.

The MTs are randomly distributed in 13 regular hexagonal cells (1 km length, 2.6 km^2) in an 8 km $\times$ 8 km grid. The HWN* attractor point mobility model (HPMM) [4] is implemented. At the simulation start, an MT schedules an *ACK* message to itself before it determines a new position. After saving the messages, the MT sends a *MOVE* message to the physical layer and reschedules the *ACK* to be delivered in a move interval. This metropolitan environment consists of n points which MTs will move towards. The mobility model implementation provides an approach which influences user mobility in a distributed manner with micro mobility, instead of grouping MTs with macro mobility.

BS is placed in the centre of each cell, and from 0 to 1300 MTs are scattered in HWN*. To ensure frequency reuse, 7 frequencies are allocated to each cell with 128 available channels. MT travels from 0 to 80 km/h since a relative speed higher than 160 km/h is not suitable for the 802.11 radio propagation model, which has limited compensation for channel fading. A node can not continue relaying packet if its speed changes to more than 10 km/h. The log-normal standard deviation σ is set as 10 dB, shadowing correlation distance χ_s is set to 50 m, and the mean SIR value r_d is set to 17 dB. Default energy model provided by OMNET++ is implemented, specifically, for a 250 m transmission range the transmit power used is 0.282 W. Transmit power used for a transmission range of d is proportional to d^4 [13].

5.1. HWN* capacity analysis

The first experiment is to present two pre-engineered RN positioning strategies' influence on the HWN* capacity under various traffic input. The HWN* network operations are considered, including the process of RN & BS registration, traffic balancing, routing path discovery, transmission mode selection, and data delivery.

When packing-based RN positioning scheme is implemented in the HWN*, per cell capacity is expected greater than random RN placement HWN* and normal cellular network under any traffic input. This is because these MTs, which are not serviced in a cell, can use the packed relay path to access other media resources strategically. With the traffic input being increased higher, packing RN-based HWN* achieves complete connectivity regardless of cellular service penetration percentage. Figure 7 records per cell capacity performance of three scenarios with traffic load

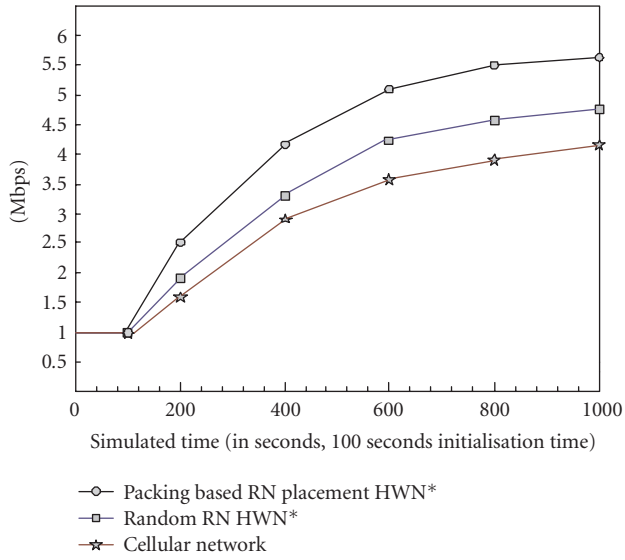


Figure 7: Average capacity comparison of packing RN HWN*, random RN HWN*, and cellular network.

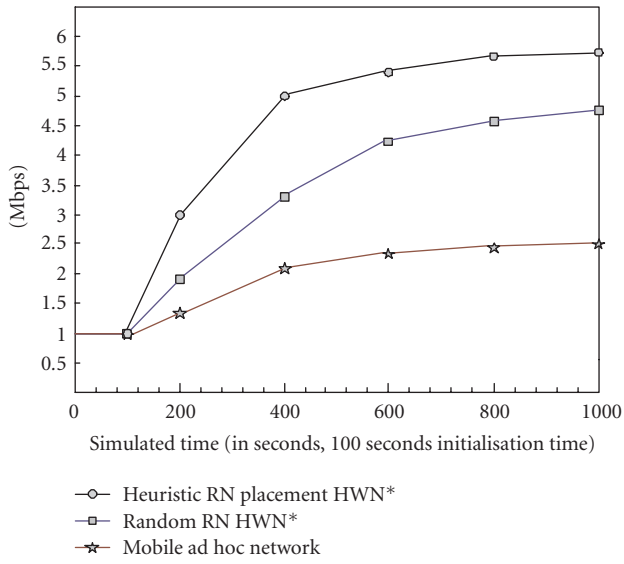


Figure 8: Average capacity comparison of heuristic RN HWN*, random RN HWN*, and mobile ad hoc network.

being increased. The capacities of both packing RN-based HWN* and random RN-based HWN* go up till maximum throughput reaches around 5.6 Mbps and 4.7 Mbps, respectively. As we can see from the trend of capacity lines, when the traffic input grows higher, packing RN-based HWN* outperforms the random RN HWN* in terms of network fairness, and its maximum capacity gets close to the theoretical gain with a more uniform communication experience.

Using the same simulation parameters, we also compare per-cell per-second capacity of heuristic placement RN HWN*, random placement RN HWN*, and mobile ad hoc network. The AODV module provided by OMNET++ has been simulated to realise MANET routing. Figure 8 Presents the result. Overall, heuristic RN placement has the highest capacity followed by packing algorithm, random HWN*, cellular network, and MANET (also refer to Figure 7). The extremely low capacity of the MANET is the results of high-contention level, erratic connections, and AODV protocol overhead. Heuristic RN-based HWN* outperforms packing RN HWN* under any traffic input, which indicates more traffic is adaptively routed. The maximum capacity of this structure achieves 5.7 Mbps.

For packet delivery ratio in the HWN*, the system throughput (ST) is defined as the delivery ratio:

$$\mathrm{ST} = \frac{\mathrm{Total\_number\_of\_data\_received}}{\mathrm{Total\_number\_of\_data\_sent}} 100\%. \quad (4)$$

In this experiment, we only implement UDP traffic (with no handshaking mechanism) on each MT instead of the default QoS0-based traffic profile, and network operations of the proposed HWN* are simulated. The packets are sent at constant bit rate (CBR) with a packet size of 1500 bytes and the MTs are added from 0 to 500 gradually as an input parameter to increase the offered load. Figure 9 shows the impact of increased traffic on the packet delivery ratio. It

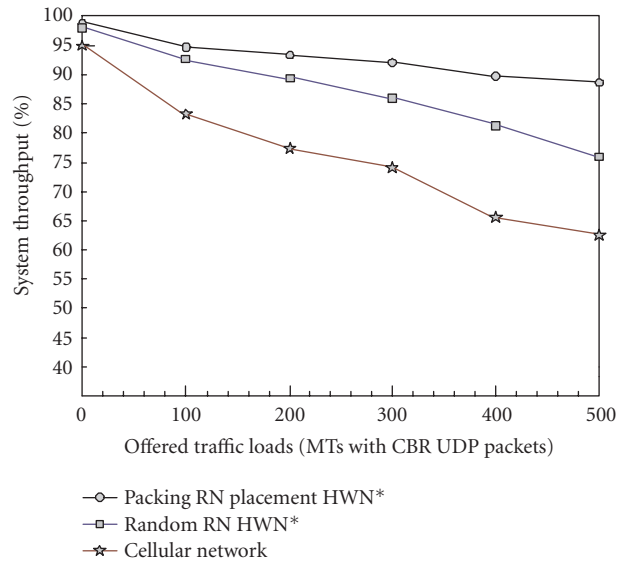


Figure 9: Packing RN HWN*, random RN HWN*, and cellular network throughput versus offered load.

indicates under any traffic input, the ADCR with packing based-RNs placement gives a higher throughput than the HWN* with random RN placement and pure cellular system. The packet delivery ratio decreases when the UDP traffic load increases, this is mainly due to the congestion. However, packing RN-based HWN* outperforms random RN HWN* or TDMA network by 12% and 26%, respectively, when the maximum traffic load is achieved.

In Figure 10, we present the throughput performance for heuristic-based RN placement HWN* with the ADCR, random RN positioning HWN* and MANET with the AODV algorithm, respectively. The curve of heuristic RN

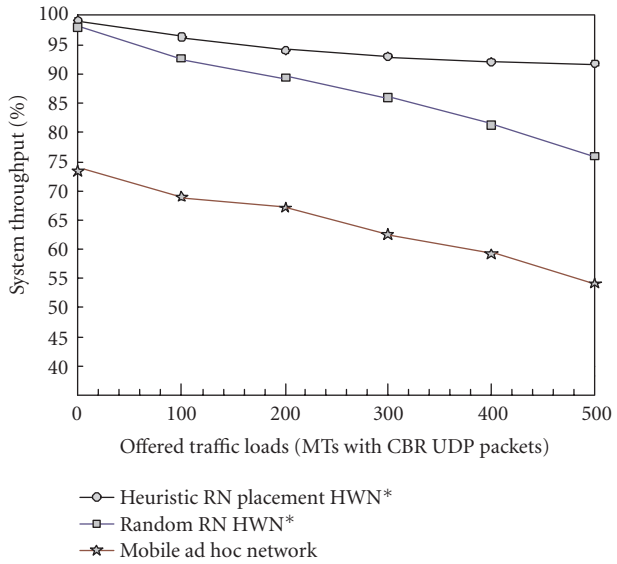


Figure 10: Heuristic RN HWN*, random RN HWN*, and MANET throughput versus offered load.

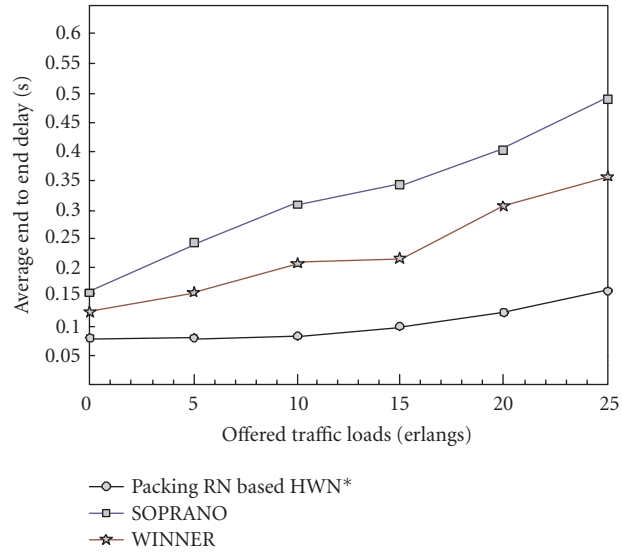


Figure 11: Average end-to-end transmission delay of HWN*, WINNER, and SOPRANO.

HWN* corresponds to a case where all the transmitted packets are maximally received, which can be considered to be an upper throughput bound on this proposed scheme. One can see that the increase of traffic load does not affect too much of the scheme' performance. Overall, heuristic algorithm has the highest throughput followed by packing RN HWN*, random RN HWN*, cellular network, and MANET (also refer to Figure 9). Furthermore, we notice that MANET exhibits a jittering performance with very low throughput under any traffic conditions. When the maximum traffic load achieves, the heuristic-based RN structure outperforms packing-based structure by 3%.

5.2. Packet transmission delay

The average packet transmission end-to-end delay of a traffic flow should be directly proportional to the number of hops traversed by the flow, and inversely proportional to the flow's end-to-end throughput, this is an interesting metric to study as the HWN* network itself has a complicated transmission arrangements, which can be seen as hybrid traffic migration of MANET, cellular network, and enhanced packet relay services. The average End-to-end Delay (AED) is defined as

$$\text{AED} = \frac{\text{Total\_number\_of\_data\_received}}{\text{Total\_delivery\_time}}. \quad (5)$$

Simplified WINNER and SOPRANO hybrid network infrastructures are therefore simulated with traffic routing functionality. The WINNER concept system realises packet switch through cooperative relaying, and RN operates same resource management functions as cellular BS. In decentralised SOPRANO, route path calculation is exclusively carried out in local MT. A minimum energy routing protocol, as recommended in [14], which maximally saves the transmission power, is simulated for SOPRANO. Figure 11 presents the average end-to-end delay versus quantised load offered for three hybrid networks. There is a significant improvement in the delay performance of HWN* ADCR, compared to the cooperative relaying of the WINNER and the minimum energy routing of the SOPRANO. At 15 Erlangs average offered load, the corresponding average end-to-end delays are 0.10, 0.21, and 0.033 seconds and the delay in other systems is almost two or three times larger than the HWN* ADCR under any traffic load. The ADCR scheme adaptively selects paths with better quality and prevents wasting transmission time.

Figure 12 presents the end-to-end delay comparison result between packing RN HWN* with ADCR, heuristic RN HWN* ADCR and heuristic HWN* ADCR + MHCT ad hoc. Interestingly, the MHCT ad hoc mode does not bring too much negative impact on system performance and the position of RNs either does not marginally influent delay performance.

Apart from RN placement plan, the number of relay nodes is another practical parameter to affect the HWN* system performance. The packing-based RN placement HWN* is chosen as the test scenario causes a random RN number to cut off heuristic algorithm that may cause large transmission delay in hot spot. Figure 13 presents delay performance of fully loaded RN, two third RN loaded and one third RN loaded scenarios under increasing traffic load. It clearly indicates that the delay is much less in fully loaded RN plan, compared to the other two scenarios with less infrastructure nodes. One can estimate that an increase of one RN reduces end-to-end delay while improves HWN* throughput at least 3% average in a small system domain including seven cellular BSs [15]. However, excessive installation of RN may not be a preferable approach because a tradeoff exists between management cost and expected system performance.

Table 3: Success route acquire ratio comparison between different user classes.

	HPUs	NPUs	LPUs	Simple HWN*
5 Erlangs/cell	100.0%	100.0%	98.0%	98.5%
10 Erlangs/cell	98.1%	93.2%	91.9%	86.2%
15 Erlangs/cell	97.3%	87.0%	86.7%	77.7%
20 Erlangs/cell	96.1%	84.4%	82.5%	57.5%
25 Erlangs/cell	95.0%	77.1%	72.1%	45.1%

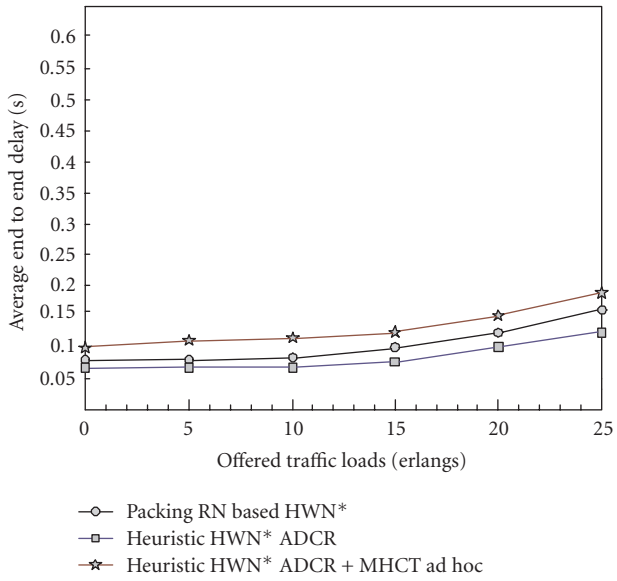


Figure 12: Average end-to-end transmission of various HWN* scenarios.

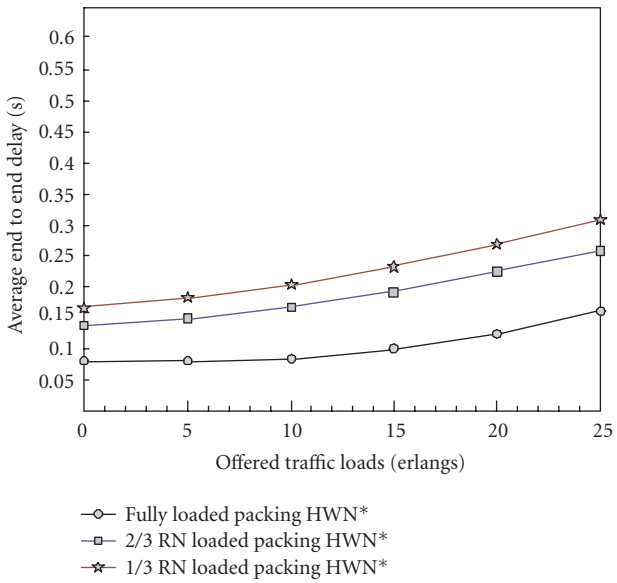


Figure 13: Average end-to-end transmission under various RN loads.

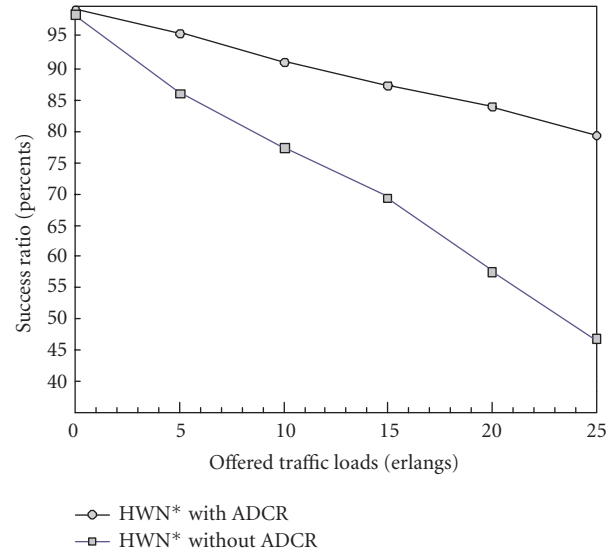


Figure 14: Comparisons of success route acquire ratio between HWN* ADCR and HWN* without ADCR.

5.3. QoS-based routing analysis

Experiments are conducted to verify that if the ADCR meets the goal of providing QoS differentiation among different users based on their class profile. To setup a comparison benchmark, we simulated a simple HWN* without any dedicated resource management and routing algorithms. Each packet session in this network has the same privileges when accessing the media resource. The arriving packet sessions are accommodated on a first-come-first-serve basis until all available channels have been occupied. An MT terminates the routing process when it can not find an alternative route. Figure 14 shows the route acquire success ratio with traffic load being constantly increased. The ADCR scheme has the most successful acquire ratio since it always returns a path on-demand and the performance improvement is marginal when the system is heavily loaded. The simple routing algorithm in the HWN* performs worse than the ADCR due to its limitation on route selection and lack of alternate paths.

Table 3 presents the individual class successful path acquiring probabilities against traffic loads offered for

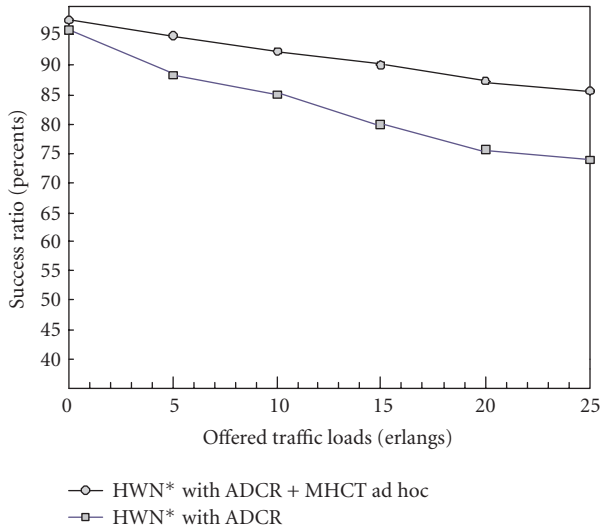


Figure 15: Comparisons of success route acquire ratio between HWN* ADCR and HWN* ADCR + MHCT ad hoc.

different user classes. It can be seen that different results are experienced by user applications in different service classes and for unclassified users in the simple HWN*. Under low- and medium-traffic intensities, the success rates are similar among HPUs, NPUs, and LPUs, since sufficient routes are available and LPUs are not largely affected by HPUs and NPUs communications. However, in the high-traffic intensity case, HPUs and NPUs applications encounter large resource competition in the MAC layer, which consumes a considerable fraction of the radio resource. This may adversely affect the route finding performance of LPUs, in particular when an HPU and NPU traffic hot spot occurs, LPUs are pushed to use the ad hoc communications modes, where the routing process are comparatively unstable compared against the infrastructure-based modes.

An interesting scenario is also investigated where we enable multihop ad hoc communication with MHCT soft relay implementation. When considering route acquire ratio, 15% of the MTs is configured residing in places outside infrastructure support, while the rest of MTs still migrates using attractor point mobility model. Configuration files for simulation such as user profiles and traffic input are kept the same as the files used in previous QoS-based routing analysis. Figure 15 presents the result comparison between HWN* with ADCR, and HWN* with ADCR + MHCT ad hoc. Different levels of route acquire ratio degradation happen in both scenarios. However, under maximum traffic input, the HWN* with ADCR + MHCT ad hoc system outperforms the HWN* ADCR without multihop ad hoc communication by 11%.

6. CONCLUSION

In this article, we have presented an overview of heterogenous wireless networks with the comparison between the HWN* stage I and HWN* stage II works. We have devised a cascaded selforganising routing scheme, the ADCR, and enabled the communication in rural places using multihop ad hoc communication. The routing algorithm employs a service class-based approach to discourage applications from using any route that is heavily loaded with low-priority traffic, with three subtransmission modes. Simulation results demonstrate that the ADCR further balances radio resource, reduces transmission delay, and potentially increases the network capacity. The future work will address cross-layer resource selfoptimisation issues in cooperative networks with dedicated relay nodes.

REFERENCES

[1] R. Beck and H. Panzer, "Strategies for handover and dynamic channel allocation in micro-cellular mobile radio systems," in *Proceedings of the IEEE 39th Vehicular Technology Conference (VTC '89)*, vol. 1, pp. 178–185, San Francisco, Calif, USA, May 1989.

[2] C. Shen, S. Rea, and D. Pesch, "Adaptive cross-layer routing for HWN with dedicated relay station," in *Proceedings of the International Conference on Wireless Communications, Networking and Mobile Computing (WiCOM '06)*, pp. 1–5, Wuhan, China, September 2006.

[3] E. Setton, T. Yoo, X. Zhu, A. Goldsmith, and B. Girod, "Cross-layer design of ad hoc networks for real-time video streaming," *IEEE Wireless Communications*, vol. 12, no. 4, pp. 59–65, 2005.

[4] C. Shen, S. Rea, and D. Pesch, "HWN* mobility management considering QoS, optimization and cross layer issues," *Journal of Communications Software and Systems*, vol. 4, no. 3, 2007.

[5] WINNER, "D4.3: identification, definition and assessment of cooperation schemes between RANs," Final deliverable, IST-2003-507581 WINNER, June 2005, https://www.ist-winner.org/.

[6] H. Y. Hsieh and R. Sivakumar, "Performance comparison of cellular and multi-hop wireless networks: a quantitative study," in *Proceedings of the ACM SIGMETRICS International Conference on Measurement and Modeling of Computer Systems (SIGMETRICS '01)*, pp. 113–122, Cambridge, Mass, USA, June 2001.

[7] C. Murthy and B. Manoj, *Ad Hoc Wireless Networks: Architectures and Protocols*, Prentice-Hall, Englewood-Cliffs, NJ, USA, 2004.

[8] B. S. Manoj, K. J. Kumar, C. D. Frank, and C. S. R. Murthy, "On the use of multiple hops in next generation wireless systems," *Wireless Networks*, vol. 12, no. 2, pp. 199–221, 2006.

[9] H. Hu and K. Yanikomeroglu, "Performance analysis of cellular networks with digital fixed relays," M.S. thesis, Carleton University, Ottawa, Canada, May 2006.

[10] A. Acharya, A. Misra, and S. Bansal, "A label-switching packet forwarding architecture for multi-hop wireless LANs," in *Proceedings of the 5th ACM International Workshop on Wireless Mobile Multimedia (WOWMOM '02)*, pp. 33–40, Atlanta, Ga, USA, September 2002.

[11] C. Shen, D. Pesch, and J. Irvine, "Distributed dynamic channel allocation with fuzzy model selection," in *Proceedings of the Information Technology and Telecommunications Conference (ITT '04)*, Limerick, Ireland, October 2004.

[12] "Omnet++ discrete event simulation system," http://www.omnetpp.org/.

[13] G. L. Stuber, *Principles of Mobile Communications*, Kluwer Academic Publishers, Norwell, Mass, USA, 1996.

[14] A. N. Zadeh, B. Jabbari, R. Pickholtz, and B. Vojcic, "Self-organizing packet radio ad hoc networks with overlay (SOPRANO)," *IEEE Communications Magazine*, vol. 40, no. 6, pp. 149–157, 2002.

[15] C. Shen, D. Pesch, and J. Irvine, "Autonomic TDD link optimising using hybrid wireless network and genetic algorithms," in *Proceedings of the 62nd IEEE Vehicular Technology Conference (VTC '05)*, vol. 1, pp. 262–266, Dallas, Tex, USA, September 2005.

Hindawi Publishing Corporation
EURASIP Journal on Advances in Signal Processing
Volume 2008, Article ID 763264, 9 pages
doi:10.1155/2008/763264

Research Article

Multiradio Resource Management: Parallel Transmission for Higher Throughput?

Alessandro Bazzi, Gianni Pasolini, and Oreste Andrisano

WiLab, IEIIT-BO/CNR, DEIS, University of Bologna, V.le Risorgimento 2, 40136 Bologna, Italy

Correspondence should be addressed to Gianni Pasolini, gianni.pasolini@unibo.it

Received 30 November 2007; Accepted 23 April 2008

Recommended by Moe Win

Mobile communication systems beyond the third generation will see the interconnection of heterogeneous radio access networks (UMTS, WiMax, wireless local area networks, etc.) in order to always provide the best quality of service (QoS) to users with multimode terminals. This scenario poses a number of critical issues, which have to be faced in order to get the best from the integrated access network. In this paper, we will investigate the issue of parallel transmission over multiple radio access technologies (RATs), focusing the attention on the QoS perceived by final users. We will show that the achievement of a real benefit from parallel transmission over multiple RATs is conditioned to the fulfilment of some requirements related to the kind of RATs, the multiradio resource management (MRRM) strategy, and the transport-level protocol behaviour. All these aspects will be carefully considered in our investigation, which will be carried out partly adopting an analytical approach and partly by means of simulations. In this paper, in particular, we will propose a simple but effective MRRM algorithm, whose performance will be investigated in IEEE802.11a-UMTS and IEEE802.11a-IEEE802.16e heterogeneous networks (adopted as case studies).

1. INTRODUCTION

It is a shared opinion among researchers that mobile communication systems beyond the third generation (3G) will see the interconnection of heterogeneous radio access networks in order to always provide the best quality of service in the most efficient way. The realization of such a scenario will allow, in fact, to pursue not only the "always best connected" paradigm, but also to increase the efficiency in the networks usage by fully exploiting the peculiarities, in terms of capacity, cost, coverage, and support of users' mobility, of the different radio access technologies (RATs) that could be deployed in the same coverage area.

Several steps have already been taken in the direction of RATs integration: protocols to make wireless local area networks (WLANs) and 3G cellular networks interact are currently under standardisation (see, e.g., [1, 2]), and user terminals able to operate with more than one communication technology are already a reality.

Nonetheless, this scenario poses a number of critical issues, which are mainly related to the architecture of future heterogeneous networks and to the radio resource management strategies to be adopted in order to take advantage of the multiaccess capability.

From the viewpoint of the heterogeneous network architecture, the simplest solution is the so-called "loose coupling": different networks are connected through gateways, still maintaining their independence. This scenario, that is based on the mobile IP paradigm, is only a little step ahead the current situation of completely independent RATs and does not allow seamless handovers between two RATs.

A more interesting and promising solution is the so-called "tight-coupling": in this case different RATs are connected to the same controller and each of them supports a different access modality to the same "core network." This solution is significantly more complex but will allow fast handovers and a really effective multiple-resources management, which in the following will be referred to as multiradio resource management (MRRM).

As far as MRRM is concerned, it is straightforward to understand that the availability of a heterogeneous access network adopting the tight-coupling architecture will make possible to take advantage of the multiradio transmission diversity (MRTD) [3, 4], which consists in the splitting of the

data flow exchanged by two end-to-end entities over more than one RAT.

MRTD can be accomplished, in particular, in a twofold manner: (1) dynamic switching between the available RATs, which are used alternatively, and (2) parallel transmission over multiple RATs [3]. In the former case, the entity performing MRRM dynamically selects the RAT via which data units are going to be transmitted, whereas in the latter case there is a parallel usage of more than one RAT for the same data flow (with or without data duplication for the transmissions over the different RATs).

The aim of this paper is, in particular, to investigate the benefits and the critical aspects of the "parallel transmission MRTD without data duplication" in a tight-coupled heterogeneous network in the case of best effort traffic.

An example investigation of "parallel transmission MRTD" is reported, for instance, in [5], where the provision of video streaming and web browsing services is considered, and the most relevant data (video base-layer and www main-objects, which are only a small fraction of the total but of great importance) are carried by an UMTS RAT, whereas a WLAN, which is faster but less reliable, is used to transmit video enhancement-layers and www inline-objects.

In this paper, differently from [5], we do not assume that the data splitting is performed by the traffic source on the basis of the data importance. Here, on the contrary, we did the more realistic assumption that the traffic source (which could be far from the end user) does not know whether multiple RATs are available at the user side or not.

We assumed, therefore, that the possible data splitting is performed locally at the Network level, by the entity managing the RATs (if more than one) covering the user region. This is even more realistic considering that users could be moving, thus dynamically entering or exiting multiple RATs areas.

Investigations on MRTD are also carried out in [6, 7], where the emphasis is on the exploitation of the radio channel diversity on a per packet basis, not considering, however, the impact of protocol layers higher than the data link.

Other studies on parallel transmission focus on the physical layer only, for instance [8, 9].

Differently, in this paper we consider the whole protocol stack, from the physical layer to the application one, with particular reference to the Transport layer protocol which, as will be seen in the following, deserves a particular attention in multiple RATs scenarios.

More in general, the achievement of a real benefit from "parallel transmission MRTD" is conditioned to the fulfilment of some requirements related to the kind of RATs, the MRRM strategy, and the transport-level protocol behaviour. All these aspects have been carefully considered in our investigation, which has been carried out partly adopting an analytical approach and partly by means of simulations.

The paper is organized as follows. In Section 2, the scenario considered for our investigations is outlined along with the assumptions and the description of the investigation methodology. In Section 3, the issue of the transport protocol behaviour with multiple RATs is addressed. In Section 4, an analytical investigation on the achievable performance level is carried out. In Section 5, an original MRRM strategy is proposed and its effectiveness is assessed. Finally, in Section 6 the final conclusions are drawn.

2. INVESTIGATION ASSUMPTIONS AND METHODOLOGY

In this paper, the three most relevant actual or upcoming RATs have been considered as case studies: the well-known wideband code division multiple access (WCDMA), UMTS technology for 3G cellular communications [10], the IEEE802.11a technology for WLANs [11], and the IEEE802.16e technology (also known as Mobile-WiMax) for broadband mobile access [12].

The scenario considered in this paper consists of a tight-coupled heterogeneous access network constituted by two RATs, either WLAN-UMTS or WLAN-WiMax.

The assumptions we made with reference to this scenario are summarized hereafter:

Technologies

As far as the three above-mentioned communication technologies are concerned, the following choices and assumptions have been made in the rest of the paper.

(1) *UMTS*. The WCDMA version of UMTS was considered, with a channelisation bandwidth of 5 MHz in the 2 GHz band. The 384 kbps bearer has always been assumed for data transmissions.

(2) *WiMax*. We considered the IEEE802.16e Wireless MAN-OFDMA version operating with 2048 OFDM subcarriers and a channelisation bandwidth of 7 MHz in the 3.5 GHz band; the time division duplexing (TDD) scheme was adopted as well as a frame duration of 10 milliseconds and a 2 : 1 downlink:uplink asymmetry rate of the TDD frame.

(3) *WLAN*. The IEEE802.11a WLAN technology has been considered as foreseen by the specification, that is, with a channelisation bandwidth of 20 MHz in the 5 GHz band and a nominal transmission rate going from 6 Mbps to 54 Mbps.

Since our interest is focused on the access network side, in this paper we assumed that packet losses and delays introduced by the core network are negligible. Packet losses and delays introduced by the access network have been, on the contrary, accurately taken into account.

MRRM

We assumed that, according to the principle of "parallel transmission MRTD," each user can simultaneously operate with both available RATs by means of a multimode user terminal.

Here we considered, in particular, the parallel transmission "without data duplication" modality. This means that the data flow of a single communication is split into two disjoint subflows addressed to the two different RATs.

We made the (realistic) assumption that the entity performing MRRM is periodically informed on the number of IP packets transmitted by each technology as well as on the number of IP packets still waiting (in the data link level transmitting queues) to be transmitted; by the knowledge of these parameters a decision on the traffic distribution over the two RATs is taken, as detailed later on.

Service

In this paper, we did not consider other traffic categories than the best effort one; users were connected to both RATs at the same time, ideally expecting to perceive a total throughput as high as the sum of those possible with each RAT singularly.

In order to make easier the interpretation of numerical results, in the following we considered, without loss of generality, only one active user performing an infinite file download.

Investigation methodology

Results have been obtained partly analytically and partly through simulations, adopting the simulation platform SHINE that has been developed in the framework of several research projects at WiLab, Bologna, Italy [13]. The aim of SHINE is to reproduce the behaviour of RATs, carefully considering all aspects related to each single level of the protocol stack and all characteristics of a realistic environment. This simulation tool, described in [14], has already been adopted to investigate a UMTS-WLAN heterogeneous network in the case of "dynamic switching MRTD" (see [15], e.g.).

Performance metric

The performance metric we adopted to investigate the above-described multiple RATs scenario is the throughput provided by the integrated network. As we focused our attention, in particular, on best effort traffic, we assumed that the TCP protocol is adopted at the transport layer and we derived, as performance metric, the TCP level throughput perceived by the final user.

Let us observe, now, that a huge number of different TCP versions are available nowadays; as will be shown in the following section, the choice of the particular TCP version adopted in the considered scenario is not irrelevant and must be carefully considered.

3. TRANSPORT LEVEL ISSUES

The most widespread versions of the TCP transport protocol (e.g., New Reno (NR) TCP [16]) work at best when packets are delivered in order or, at least, with a sporadic disordering. A frequent out-of-order delivery of TCP packets originates, in fact, useless duplicates of transport level acknowledgments; after three duplicates a packet loss is supposed by the transport protocol and the fast recovery-fast retransmit phase is entered at the transmitter side.

This causes a significant reduction of the TCP congestion-window size and, as a consequence, a reduction of the throughput achievable at the transport level.

This aspect of the TCP behaviour has been deeply investigated in the literature (e.g., [17, 18]) and modern communication systems often include a reordering entity at the data link level of the receiver side (see, e.g., the WiMax standard [12]) to prevent possible performance degradation.

Let us observe, now, that when "parallel transmission MRTD" is adopted, each RAT works autonomously at data link and physical levels, with no knowledge of other active RATs. During the transmission phase, in fact, the packets flow coming from the upper layers is split into subflows that are passed to the different data link level queues of the active RATs and then transmitted independently one of the others.

It follows that the out-of-order delivery of packets and the consequent performance degradation are very likely, owing to possible differences of the queues occupation levels as well as of the medium access strategies and the transmission rates of the active RATs.

The independency of the different RATs makes very difficult, however, to perform a frame reordering at the data link level of the receiver and, at the same time, it would be preferable to avoid, for the sake of simplicity, the introduction of an entity that collects and reorders TCP level packets coming from different RATs. For this reason, the adoption of particular versions of TCP, especially designed to solve this problem, is advisable in multiple RATs scenarios.

Here, we considered the adoption of the delayed duplicates New Reno version of TCP (DD-TCP) [18], which simply delays the transmission of TCP acknowledgments when an out-of-order packet is received, hoping that the missing packet is already on the fly. The drawback of this solution is, of course, that the fast recovery-fast retransmit phases are delayed also when they are necessary.

The DD-TCP differs from the NR-TCP only at the receiving side of the transport level peer-to-peer communication; this implies that the NR-TCP can be maintained at the transmitter side. Thus, this solution could be adopted, at least, on multimode user terminals, where the issue of out-of-order packet delivery is more critical owing to the higher traffic load that usually characterises the downlink phase.

In order to investigate the impact of DD-TCP on the performance achievable with the "parallel transmission MRTD," here we considered a downlink best effort connection simultaneously exploiting two RATs.

As our aim was to highlight only the effect of the transport-level behaviour, the heterogeneous network considered for this specific investigation was somewhat anomalous: the two considered RATs were, in fact, both IEEE802.11a WLANs whose access points (APs) were located in the same place. Since the two simultaneous connections provide the same throughput, the MRRM strategy we adopted in this case randomly distributed TCP/IP packets between the two RATs with equal (i.e., 50%) probability.

The outcome of this investigation is reported in Figure 1, where the amount of acknowledged TCP packets is reported as a function of the time for both DD-TCP and NR-TCP. The case of a single AP (i.e., of a single RAT) is also shown

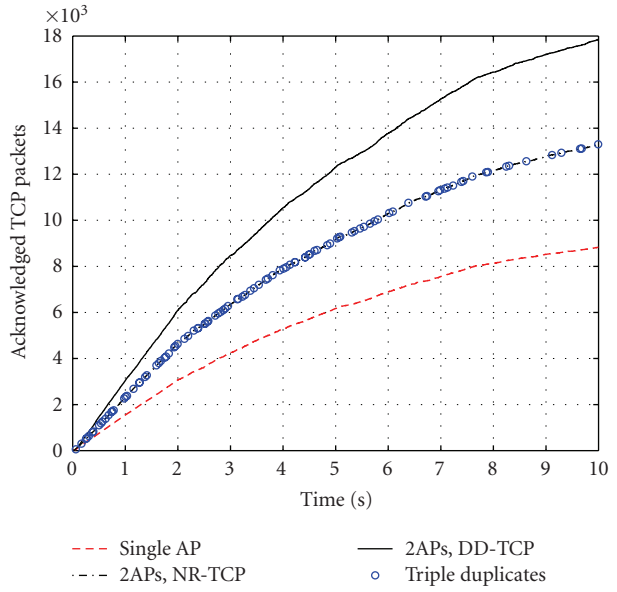


Figure 1: Acknowledged TCP packets of the download performed by one user that moves away from 2 colocated WLAN APs versus time; single RAT connection compared to parallel transmission over two RATs adopting either New Reno or delayed duplicates New Reno as TCP protocols. Triple duplicate events are marked with "o."

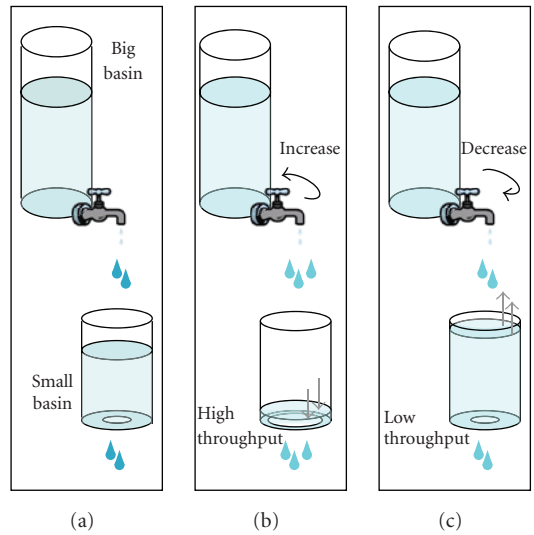


Figure 2: Representation of the TCP mechanism.

for comparison purpose (in this case DD-TCP and NR-TCP provide the same performance); the circles ("o") indicate the triple-duplicates events.

To derive the results reported in Figure 1, we considered a user that, starting from the APs position, moves away at a speed of 3 m/s. It follows that increasing time instants correspond to increasing distances from the APs and, as a consequence, to a decreasing slope of the curves, which is induced by the WLAN link adaptation strategy that, as the user moves away from the APs, selects more reliable but slower modulation/coding schemes.

Observing Figure 1, it is important to notice that triple duplicates are generated only when NR-TCP over two RATs is adopted, and that they occur during the whole simulated time interval, no matter the distance from the APs (i.e., independently on the signal quality); this means that all triple duplicates here observed are a consequence of out-of-order packet delivery events. We verified in fact that, thanks to the WLAN automatic repeat request (ARQ) mechanism, no data link level fragment, and consequently no TCP/IP packet, is lost in the investigated scenario during the whole simulation time, even when the maximum distance is reached (after 10 seconds).

As can be observed, triple duplicates heavily affect the achieved performance level; the comparison with the curve related to a single AP shows, in fact, that the number of acknowledged packets is not doubled when considering NR-TCP with two RATs.

When DD-TCP is adopted, on the contrary, no triple duplicate event occurs and the amount of TCP packets acknowledged in a given time interval, which is strictly related to the provided throughput, is doubled with respect to the single connection case. Let us underline that this is not a trivial result, since we are splitting a single TCP flow over two independent technologies and reassembling it directly at the TCP level of the receiver.

Please note that the DD-TCP protocol was chosen, among other possibilities, since it is a very simple solution. It is beyond the scope of this paper to investigate the most suitable TCP version to overcome the triple duplicate problem in multiple RATs networks.

4. THROUGHPUT ANALYSIS

Let us consider, now, a really heterogeneous network, which is in general constituted by RATs whose characteristics could be very different in terms, for instance, of medium access strategies and transmission rates.

It is straightforward to understand that, in this case, the random distribution of packets with uniform probability over the different RATs would hardly be the best solution. Indeed, to fully exploit the availability of multiple RATs and get the best from the integrated access network, an efficient MRRM strategy must be designed, able to properly balance the traffic distribution over the different access technologies.

In order to clarify this statement, a brief digression on the TCP protocol behaviour is reported hereafter, starting from a simple metaphor.

Let us represent the application-level queue as a big basin (in the following, big basin) filled with water that represents the data to be transmitted (see Figure 2(a)). Another, smaller basin (in the following, small basin) represents, instead, the data path from the source to the receiver: the size of the data link level queue can be represented by the small basin size and the transmission speed by the width of the hole at the small basin bottom.

In this representation the TCP protocol works like a tap controlling the amount of water to be passed to the small basin in order to prevent overflow events (a similar metaphor

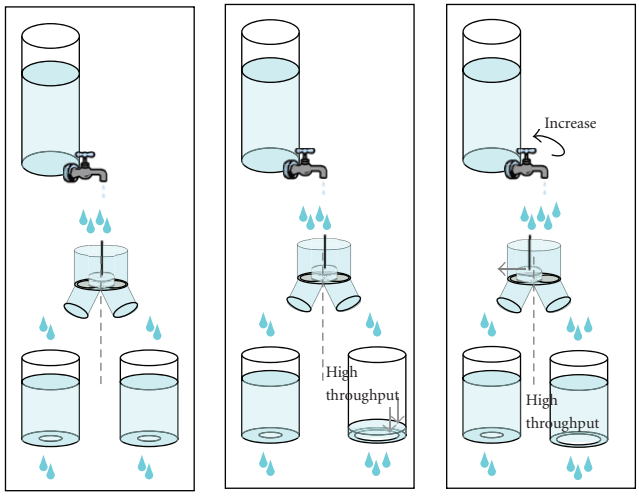


Figure 3: Representation of the TCP mechanism with parallel transmission over two RATs.

is used, e.g., in [19]). It follows that the water flow exiting from the tap represents the TCP level throughput, and the water flow exiting from the small basin represents the data link level throughput.

As long as the small basin is characterised by a wide hole, as depicted in Figure 2(b), the tap can increase the water flow, reflecting the fact that when a high data link level throughput is provided by the communication link, the TCP level throughput can be correspondingly increased.

When, on the contrary, a small hole (→ a low data link level throughput) is detected, the tap (→ the TCP protocol) reduces the water flow (→ the TCP level throughput), as described in Figure 2(c). This way, the congestion control is performed, and the data link level queues saturation is avoided.

Now the question is: what happens when two basins (i.e., two RATs) are available instead of one and the water flow is equally split between them?

Having in mind that the tap has to prevent the overflow of either of the two small basins, it is easy to understand that, in the presence of two small basins with the same hole widths, the tap could simply double the water flow, as depicted in Figure 3(a). Reasoning in terms of throughput and multiple RATs, this is the case investigated in Figure 1, where two equal and equally loaded RATs were considered.

In the presence of a small basin with a hole wider than the other (see Figure 3(b)), on the other hand, the tap behaviour is influenced by the small basin characterised by the lower emptying rate (the leftmost one in Figure 3(b)), which is the most subject to overflow. This means that the availability of a further "wider holed" basin is not fully exploited in terms of water flow increase. Reasoning in terms of TCP protocol, in fact, the congestion window moves following the TCP level acknowledgments related to packets received in the correct order. This means that, as long as a gap is present in the received packet sequence (one or more packets are missing because of a RAT slower than the other), the congestion window does not move at the transmitter side, thus reducing the provided throughput.

Coming back to the water flow metaphor, it is immediate to understand that, in order to fully exploit the availability of the further, "more performing," small basin, the water flow splitting modality must be modified in such a way that the water in the two small basins is kept at almost the same level (see Figure 3(c)). This consideration introduces in our metaphor the concept of resource management, which is represented in Figure 3(c) by the presence of a valve which dynamically changes the subflows discharge.

This concept, translated in the telecommunication-correspondent MRRM concept, will be thoroughly worked out in the remainder of the paper. To do this, however, an analytical formulation of TCP protocol behaviour in the presence of multiple RATs is needed, which is reported in the following subsection.

4.1. Throughput analytical derivation

Starting from the above-reported considerations, we can derive a simple analytical framework to model the average throughput T perceived by the final user in the case of two heterogeneous RATs, denoted in the following as RAT_A and RAT_B, managed by an MRRM entity which splits the packets flow between RAT_A and RAT_B with probabilities P_A and $P_B = 1 - P_A$, respectively.

Focusing the attention on a generic user, let us denote with T_i the maximum data link level throughput supported by RAT_i in the direction of interest (uplink or downlink), given the particular conditions (signal quality, network load due to other users,...) experienced by the user. Dealing with a dual mode user, we will denote with T_A and T_B the above-introduced metric referred to RAT_A and RAT_B, respectively.

Let us assume that a block of N transport-level packets of B bits has to be transmitted and let us denote, furthermore, with O the amount of overhead bits added by protocol layers from transport to data link. After the MRRM operation, the N packets flow is split into two subflows of, in average, $N \cdot P_A$ and $N \cdot P_B$ packets, which are addressed to RAT_A and RAT_B.

It follows that, in average, RAT_A and RAT_B empty their queues in $D_A = (N \cdot (B + O) \cdot P_A)/T_A$ and $D_B = (N \cdot (B + O) \cdot P_B)/T_B$ seconds, respectively.

Thus, the whole N packets block is delivered to the considered user in a time interval that corresponds to the longest between D_A and D_B.

This means that the average TCP level throughput provided by the integrated access network to the final user can be expressed as

$$T = \begin{cases} \dfrac{N \cdot B}{D_A} = \dfrac{T_A}{P_A}\xi, & \text{when } D_A > D_B, \text{ that is when } \dfrac{T_A}{P_A} < \dfrac{T_B}{P_B}, \\ \dfrac{N \cdot B}{D_B} = \dfrac{T_B}{P_B}\xi, & \text{in the opposite case, when } \dfrac{T_A}{P_A} \geq \dfrac{T_B}{P_B}, \end{cases} \tag{1}$$

or in a more compact way as

$$T = \min\left\{\frac{T_A \xi}{P_A}, \frac{T_B \xi}{P_B}\right\}, \tag{2}$$

where the factor $\xi = B/(B + O)$ takes into account the degradation due to the overhead introduced by protocol layers from transport to Data Link.

Let us observe, now, that the term $T_A\xi/P_A$ of (2) is a monotonic increasing function of $P_B = 1 - P_A$, while the term $T_B\xi/P_B$ is monotonically decreasing with P_B.

Since $T_A/P_A < T_B/P_B$ when P_B tends to 0 and $T_A/P_A > T_B/P_B$ when P_B tends to 1, it follows that the maximum TCP level throughput $T_{\max}$ is achieved when $T_A/P_A = T_B/P_B$, that is, when

$$P_A = P_A^{(\max)} = \frac{T_A}{T_A + T_B}, \tag{3}$$

and consequently

$$P_B = P_B^{(\max)} = 1 - P_A^{(\max)} = \frac{T_B}{T_A + T_B}, \tag{4}$$

having denoted with $P_A^{(\max)}$ and $P_B^{(\max)}$ the values of P_A and P_B that maximize T.

Recalling (2), the maximum achievable TCP level throughput is immediately derived as

$$T_{\max} = \min\left\{\frac{T_A\xi}{P_A}, \frac{T_B\xi}{P_B}\right\}\Big|_{P_A = P_A^{(\max)}} = (T_A + T_B)\xi, \tag{5}$$

thus showing that a TCP level throughput as high as the sum of the single TCP level throughputs can be achieved.

Equations (3) and (4) show that an optimal choice of P_A and P_B is possible, in principle, on condition that accurate and updated values of the data link level throughputs T_A and T_B are known (or, equivalently, accurate and updated values of the TCP level throughputs $T_A\xi$ and $T_B\xi$).

4.2. Model validation

In order to validate the above-described analytical framework, a simulative investigation has been carried out considering two different scenarios: the first one integrates a WLAN RAT and a WiMax RAT, while the second one integrates a WLAN RAT and an UMTS RAT.

All wireless access points, that is, the WLAN AP, the UMTS Node B, and the WiMax base station, are placed in the same position and the single user here considered is located near them (this means high perceived signal to noise ratio).

Packets are probabilistically passed by the MRRM entity to the WLAN data link/physical levels with probability P_{WLAN} (which corresponds to P_A of the general analytical framework) and to the other technology (i.e., WiMax in the first case or UMTS in the second one) with probability $1 - P_{\text{WLAN}}$ (which corresponds to P_B of the general analytical framework), both in the uplink and in the downlink.

The simulations outcomes are reported in Figure 4, where the average throughput perceived at the TCP level is shown as a function of P_{WLAN}.

In the same figure, we also reported the curves obtained through (2), in which we assumed that $T_A\xi$ is referred (in both scenarios) to the WLAN RAT, and $T_B\xi$ is referred, depending on the scenario, to the WiMax RAT (WLAN-WiMax scenario) or to the UMTS RAT (WLAN-UMTS scenario). The values of $T_A\xi$ and $T_B\xi$, to be feeded to (2), have been obtained by means of simulations for each one of the considered technologies, obtaining $T_{\text{WLAN}} = T_A\xi = 18.53$ Mbps, $T_{\text{WiMax}} = T_B\xi = 12.76$ Mbps (first scenario) and $T_{\text{UMTS}} = T_B\xi = 0.36$ Mbps (second scenario).

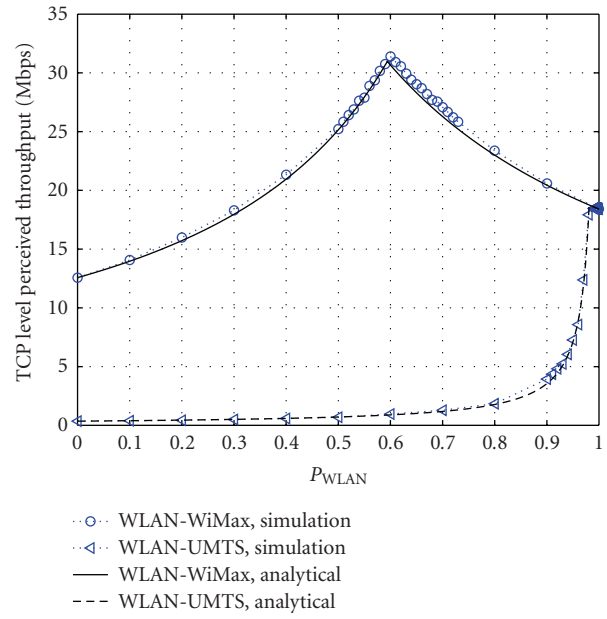


Figure 4: TCP level throughput adopting a WLAN connection and a WiMax or UMTS one, as a function of the probability that the packet is transferred through the WLAN.

With reference to Figure 4, let us observe, first of all, the very good matching between the simulation results and the analytical curves derived from (2), which confirms the accuracy of the whole framework. The accuracy of (3) and (5) can also be easily checked. Focusing the attention, for instance, on the WLAN-WiMax case, it is easy to derive (from (3)) $P_A^{(\max)} = P_{\text{WLAN}} = 0.59$ and (from (5)) $T_{\max} = 31.29$ Mbps, in perfect agreement with the coordinates of the maximum that can be observed in the curve related to the WLAN-WiMax scenario.

Let us observe, moreover, the rapid throughput degradation following an uncorrect choice of P_{WLAN}. This means that the correct assessment of P_{WLAN} heavily impacts the system performance.

Focusing the attention on the curve related to the WLAN-UMTS heterogeneous network, we can argue that in the investigated conditions the high difference of the data link throughputs provided by the two RATs makes the TCP behaviour so inefficient that the adoption of the WLAN technology alone is almost the best solution; a significant performance degradation can be noted, in fact, when P_{WLAN} is lower than $\sim$0.98. For this reason, in the next session we will focus on the WLAN-WiMax heterogeneous network only.

Please note that, although we limited our investigation to the case of two active technologies, all conclusions can also be generalised for a greater number of RATs.

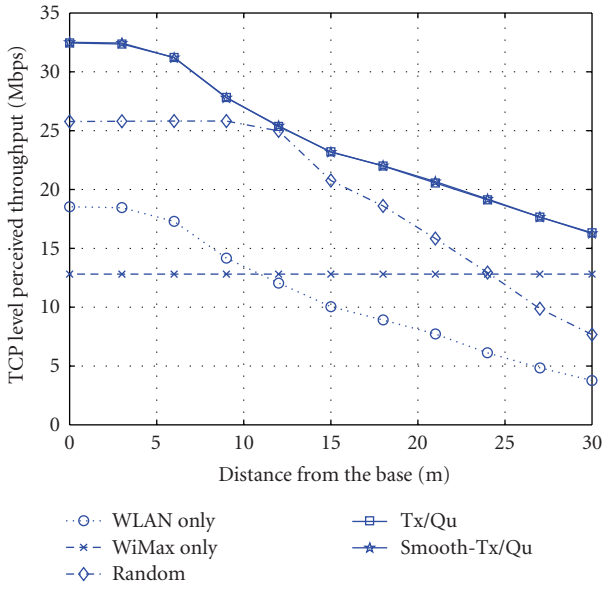


Figure 5: WLAN-WiMax heterogeneous networks. TCP level throughput varying the distance of the user from the AP/base station, for different MRRM schemes. No mobility.

5. MRRM STRATEGIES

In Section 4, we showed that, depending on the characteristics of the considered RATs, there exists an optimal traffic distribution policy for each (dual mode) user, which depends, in particular, on the average throughput that every single RAT can provide to it.

In principle, starting from knowledge of the maximum data link or TCP throughput that can be provided to the user by each RAT, the MRRM entity could perform the optimal traffic balance on the basis of (3) and (4).

Let us observe, however, that the maximum (data link or TCP) throughput that can be provided to a single user by a given RAT is time variable, since it depends on a number of dynamically changing parameters, such as the amount of served users (which affects the data link level queue occupation), its position (which could affect the physical level transmission rate if a link adaptation algorithm is adopted), the presence of interference, and so forth. It follows that its assessment could be difficult and scarcely accurate.

When a new connection is established, in fact, no knowledge of the throughput that the incoming user will perceive is available, hence no optimal traffic balance could be performed at the connection activation. When the communication is ongoing, on the other hand, the not optimal traffic balance performed at the connection setup could bring to an under-utilisation of one (or more) RAT, thus making the related throughput measurement not consistent with the throughput potentially available and consequently preventing the correct choice of the splitting probabilities.

Focusing again the attention to the case study of the two heterogeneous networks previously considered, the question is therefore how to dynamically and automatically select the correct value for P_{WLAN}.

In this paper, we propose an original MRRM strategy, that we called *Smooth-Tx/Qu*, and we compare its performance with those of benchmark cases. More specifically, the following MRRM strategies are considered and compared in the following.

(i) *Random*: packets are randomly distributed with equal probability among active connections (please note that a random distribution corresponds to $P_{\text{WLAN}} = 0.5$ and observing the curve of Figure 4 related to the WLAN-UMTS case, we can argue that in some cases this is absolutely a wrong choice). This policy is considered only for comparison purpose.

(ii) *Transmissions over Pending Packets (Tx/Qu)*: packets are always passed to the technology with the higher value of the ratio between the number of transmitted packets and the number of packets waiting in the data link queue; thus, system queues are kept filled proportionally to the transmission speed;

(iii) *Smoothed Transmissions over Pending Packets (Smooth-Tx/Qu)*: it is an evolution of the Tx/Qu strategy. The only difference is that in this case the number of transmitted packets is halved every T_{half} seconds (in our simulations we adopted $T_{\text{half}} = 0.125$ s); periodically halving the amount of transmitted packets allows to reduce the impact of old transmissions, thus improving the achieved performance in a scenario where transmission rates could change (due to users mobility, e.g.).

In Figure 5, the above-detailed MRRM strategies are compared in a scenario consisting of a heterogeneous network with one IEEE802.11a AP and one WiMax base station located in the same position. The user is performing an infinite file download and does not change its position; its distance from the colocated AP/base station is reported on the x-axis, while the average perceived TCP level throughput is reported on the y-axis.

Before discussing the results reported in Figure 5, a preliminary note on the considered distance range (0–30 m) is needed.

Let us observe, first of all, that WiMax is a long range communications technology, with a coverage range in the order of kilometers. Nonetheless, since our focus is on the heterogeneous WLAN-WiMax access network, we must consider coverage distances in the order of few dozens of meters (i.e., the coverage range of a WLAN), where both RATS are available; for this reason the x-axis of Figure 5 ranges from 0 to 30 meters.

The different curves of Figure 5 refer, in particular, to the three MRRM strategies above described and, for comparison, to the cases of a single WLAN RAT and of a single WiMax RAT.

Obviously, when considering the case of a single WiMax RAT, the throughput perceived by a user located in the region of interest is always at the maximum achievable level, as shown by the flat curve in Figure 5. As expected, on the contrary, the throughput provided by the WLAN in the same range of distances rapidly decreases for increasing distances.

Table 1: TCP level average throughput in Mbps for various distribution schemes in different conditions. Single user, 1 WLAN AP and 1 WiMax Base Station (BS) colocated. 10 seconds simulated.

User position	WLAN only	WiMax only	Random	Tx/Qu	Smooth-Tx/Qu
(1) Still, near the AP/BS	18.53	12.76	25.23	32.28	32.37
(2) Still, 30 m far from the AP/BS	3.81	12.76	7.95	16.35	16.40
(3) Moving away at 1 m/s, starting from the AP/BS	11.83	12.76	20.99	24.94	25.01
(4) Near the AP/BS for half simulation, then 30 m far (instantaneously)	10.04	12.61	14.44	18.43	21.03

The most important results reported in Figure 5, however, are related to the three upper curves (two of them are superimposed), which refer to the previously described MRRM strategies when applied in the considered heterogeneous WLAN-WiMax access network.

As can be immediately observed, the two dynamic strategies proved to be really effective, greatly outperforming the random distribution strategy. Please observe that the achievable throughput in these cases is even slightly higher than the sum of the throughputs provided by each technology alone.

At a first glance, it could seem strange that a throughput (slightly) higher than the sum of the two throughputs provided in the single RAT cases can be achieved; however, this phenomenon can be easily explained considering the fact that the adopted DD-TCP solution (slightly) reduces the number of TCP level acknowledgments transmitted in the uplink phase (in average a higher number of packets are acknowledged by a single DD-TCP acknowledgment with respect to NR-TCP). Since in a WLAN the uplink and downlink phases contend for the wireless medium, a reduction of the uplink traffic turns into a downlink throughput increase. This marginal aspect, which is strictly related to the particular medium access control strategy adopted by the WLAN, is neglected in the analytical framework developed in Section 4.

As a final consideration on Figure 5, we can observe that the *Smooth-Tx/Qu* and the *Tx/Qu* strategies are almost equivalent in this case; this is due to the fact that the related curves have been obtained considering a still user.

The impact of user mobility is immediately evident considering the results reported in Table 1, which are related to the same scenario (single user and a WLAN-WiMax heterogeneous network with colocated WLAN AP and WiMax base station) in different conditions. Four scenarios are, in particular, considered:

(1) the user stands still near the AP/base station (optimal signal reception);

(2) the user stands still at 30 m from the AP/base station (optimal WiMax signal, but medium quality WLAN signal);

(3) the user moves from the AP/base station far away at a speed of 1 m/s (low mobility);

(4) the user stands still near the AP/base station for half the simulation time, then it moves instantaneously 30 m far away (reproducing the effect of a high-speed mobility).

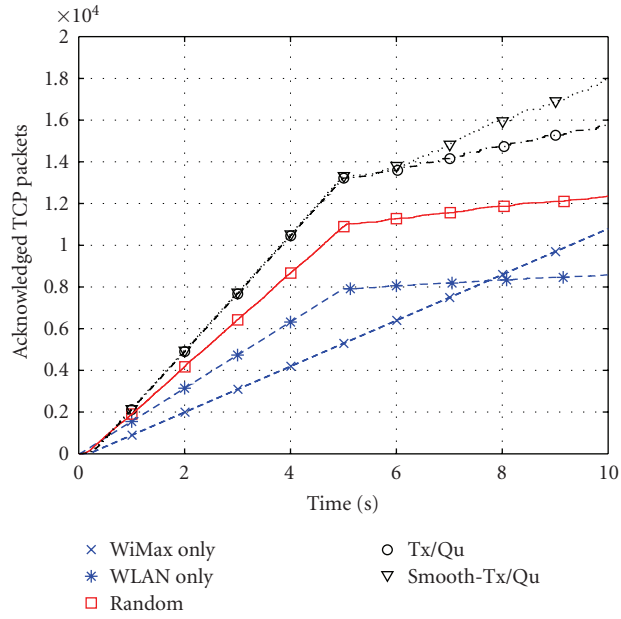


Figure 6: Acknowledged TCP packets of the download performed by one user that instantaneously (after 5 seconds) moves 30 m away from colocated WLAN AP and WiMax BS versus time; single WLAN RAT connection, single WiMax RAT connection, and different distribution strategies are compared.

Results are shown for all the above-described MRRM strategies as well as for the benchmark scenarios with a single WLAN RAT and a single WiMax RAT and refer to the average (over the 10 seconds simulated time interval) throughput perceived in each considered case.

As can be observed, while the random distribution confirms its poor performance (please note that when it is adopted with the user standing still at a distance of 30 m, the perceived throughput is lower than that obtained using WiMax only), the proposed dynamic MRRM methods provide satisfying performance. Focusing the attention on the last case (correspondent to high mobility), the gain achieved with the *Smooth-Tx/Qu* method is clearly evident, although the *Tx/Qu* method may be sufficient in most cases.

To get a more accurate picture of the system behaviour in a high mobility scenario, in Figure 6 the amount of acknowledged TCP level packets is shown as a function of the time, in the above-described scenario 4. Please note that the throughput values shown in the fourth row of Table 1 can be obtained from Figure 6 through the following equation: $T = (N_{\text{acked}} \cdot N_{\text{bit}})/D$, where T is the average throughput in

bits per second, N_{acked} is the total number of acknowledged packets, N_{bit} is the number of payload bits per TCP packet (i.e., 1460*8), and D is the total duration of the simulation (i.e., 10 seconds).

Observing the curves related to the *Tx/Qu* and the *Smooth-Tx/Qu* strategies, the effectiveness of the latter approach appears, once more, clearly evident. In the former case, in fact, the splitting probabilities update takes place very slowly in time, thus reducing the total achievable throughput.

6. CONCLUSIONS

In this paper, we faced the issue of RATs integration in tight-coupled heterogeneous networks. The "parallel transmission multiradio diversity" has been particularly investigated with the aim to highlight benefits and critical aspects. Results, obtained through simulations, refer to a TCP session whose traffic is split over different access technologies without the need of any modifications to communication protocols.

Here, we proposed original multiradio resource management strategies and derived their performance in extremely relevant scenarios, such as those constituted by WLAN-UMTS and WLAN-WiMax heterogeneous networks.

The main outcomes of our investigations can be summarised as follows:

(i) the parallel transmission allows a total throughput as high as the sum of throughput of the single RATs;

(ii) the parallel transmission generates a disordering of upper layers packets at the receiver side; this is an issue to be carefully considered when the parallel transmission refers to a TCP connection;

(iii) the performance of parallel transmission is very sensitive to the algorithm adopted to split upper layers packet over the considered RATs;

(iv) when different RATs with remarkable difference in achievable throughput are considered, the adoption of parallel transmission as defined in this paper should be preferably avoided;

(v) the proposed dynamic MRRM strategy, in spite of its simplicity, proved to be really effective, fully exploiting the pool of resources provided by the integrated heterogeneous network.

REFERENCES

[1] 3GPP TS 22.234 v8.1.0, "Requirements on 3GPP system to Wireless Local Area Network (WLAN) interworking," June 2007.

[2] Unlicensed Mobile Access specifications, http://www.umatoday.com/.

[3] K. Dimou, R. Agero, M. Bortnik, et al., "Generic link layer: a solution for multiradio transmission diversity in communication networks beyond 3G," in *Proceedings of the 62nd IEEE Vehicular Technology Conference (VTC '05)*, vol. 3, pp. 1672–1676, Dallas, Tex, USA, September 2005.

[4] J. Sachs, H. Wiemann, J. Lundsjö, and P. Magnusson, "Integration of multiradio access in a beyond 3G network," in *Proceedings of the 15th IEEE International Symposium on Personal, Indoor and Mobile Radio Communications (PIMRC '04)*, vol. 2, pp. 757–762, Barcelona, Spain, September 2004.

[5] J. Luo, R. Mukerjee, M. Dillinger, E. Mohyeldin, and E. Schulz, "Investigation of radio resource scheduling in WLANs coupled with 3G cellular network," *IEEE Communications Magazine*, vol. 41, no. 6, pp. 108–115, 2003.

[6] L. Badia, C. Taddia, G. Mazzini, and M. Zorzi, "Multiradio resource allocation strategies for heterogeneous wireless networks," in *Proceedings of the Wireless Personal Multimedia Communications Conference (WPMC '05)*, Aalborg, Denmark, September 2005.

[7] R. Veronesi, "Multiuser scheduling with multiradio access selection," in *Proceedingsof the 2nd International Symposium on Wireless Communications Systems (ISWCS '05)*, pp. 455–459, Siena, Italy, September 2005.

[8] M. Saquib, S. Das, G. Mandyam, and M. Z. Win, "Fade resistant transmission over time-varying wireless channels using parallel sequences," in *Proceedings of the IEEE International Conference on Communications (ICC '04)*, vol. 1, pp. 385–389, Paris, France, June 2004.

[9] M. Saquib and M. Z. Win, "Fade-resistant transmission over time-varying wireless channels," *IEEE Signal Processing Letters*, vol. 11, no. 6, pp. 561–564, 2004.

[10] http://www.3gpp.org/.

[11] IEEE Std 802.11a 1999, "Information technology—telecommunications and information exchange between systems—local and metropolitan area networks specific requirements part 11: wireless lan medium access control (MAC) and physical layer (PHY) specifications amendment 1: high-speed physical layer in the 5 GHz band".

[12] IEEE Std 802.16e-2005 and IEEE Std 802.16-2004/Cor1-2005, "IEEE Standard for Local and metropolitan area networks Part 16: Air Interface for Fixed and Mobile Broadband Wireless Access Systems Amendment 2: Physical and Medium Access Control Layers for Combined Fixed and Mobile Operation in Licensed Bands and Corrigendum 1," February 2006.

[13] Wireless Communications Laboratories, Bologna, Italy, http://www.wilab.org/.

[14] A. Bazzi, G. Pasolini, and C. Gambetti, "SHINE: simulation platform for heterogeneous interworking networks," in *Proceedings of the IEEE International Conference on Communications (ICC '06)*, vol. 12, pp. 5534–5539, Istanbul, Turkey, June 2006.

[15] O. Andrisano, A. Bazzi, M. Diolaiti, C. Gambetti, and G. Pasolini, "UMTS and WLAN integration: architectural solution and performance," in *Proceedings of the 16th IEEE International Symposium on Personal, Indoor and Mobile Radio Communications (PIMRC '05)*, vol. 3, pp. 1769–1775, Berlin, Germany, September 2005.

[16] S. Floyd and T. Henderson, "The NewReno Modification to TCP's Fast Recovery Algorithm," RFC 2582, April 1999.

[17] J. C. R. Bennett, C. Partridge, and N. Shectman, "Packet reordering is not pathological network behavior," *IEEE/ACM Transactions on Networking*, vol. 7, no. 6, pp. 789–798, 1999.

[18] M. N. Mehta and N. H. Vaidya, "Delayed duplicate-acknowledgments: a proposal to improve performance of TCP on wireless links," Tech. Rep., Department of Computer Science, Texas A&M University, College Station, Tex, USA, December 1997.

[19] A. S. Tanenbaum, *Computer Networks*, Prentice Hall, Upper Saddle River, NJ, USA, 2002.