Asymmetric Cooperation among Wireless Relays with Linear Precoding

Natasha Devroye, Neelesh B. Mehta, *Senior Member, IEEE*, and Andreas F. Molisch, *Fellow, IEEE*

***Abstract*—Wireless relays extend coverage, improve spectral efficiency, and enhance reliability and rates of wireless cellular communication systems. In this work, we introduce the fundamental notion of *asymmetric cooperation* among cooperating relays in cellular downlinks - different relays are party to different but overlapping knowledge about the messages transmitted from the base station. We argue that asymmetric cooperation arises naturally in most two-phase protocols in which the base station first transmits information to multiple relays that then cooperatively forward the information to the recipient mobile stations in the cell. For a system in which two relays are of the decode-and-forward type and cooperate using linear precoding to communicate with two mobile stations, we formulate the general, but complicated, throughput optimization problem and derive several results that considerably simplify the optimization. We show that under different channel configurations and fairness criteria, asymmetric cooperation is often the throughput-maximizing option. Under typical configurations, a 20-30% throughput enhancement is achieved compared to conventional full-cooperation systems.**

***Index Terms*—Relays, fading channels, cooperation, linear precoding, asymmetry, SDMA, MIMO.**

I. Introduction and Motivation

RELAYS, both fixed and mobile, promise great gains in wireless cellular and ad-hoc networks. Relays promise to extend cell coverage, boost transmission rates, improve spectral efficiency, and achieve all this at much lower costs than deploying full-fledged base stations (BSs) or nodes [1]–[5]. Consequently, recent literature [6]–[8] and standards such as IEEE 802.16j [9] have proposed augmenting cellular networks with fixed or mobile wireless relays. In ad-hoc networks, relays have traditionally been considered as wireless devices that receive and subsequently retransmit a signal or message in a multi-hop scenario. However, recent results have extended their use to more general multi-terminal cooperative networks [10]–[13].

In this work, we investigate the gains of cooperation between relays in a cellular setting. We consider a simple but fundamental scenario in which a single-antenna BS wishes to communicate distinct messages to multiple single-antenna mobile stations (MSs) via multiple single-antenna relays. As often considered in the literature and standards, downlink communication is performed in a canonical two-hop or two-phase manner [1], [2], [14]–[16].

Manuscript received November 7, 2007; revised May 2008; accepted August 11, 2008. The associate editor coordinating the review of this paper and approving it for publication was R. Nabar.

N. Devroye is with the Dept. of Electrical and Computer Engineering, University of Illinois at Chicago, Chicago, IL, USA (e-mail: devroye@uic.edu).

N. B. Mehta is with the Dept. of Electrical Communication Engineering, Indian Institute of Science (IISc), Bangalore, India.

A. F. Molisch is with Mitsubishi Electric Research Laboratories (MERL), Cambridge, MA, USA. A. F. Molisch is also at the Dept. of Electroscience, Lund University, Sweden.

Digital Object Identifier 10.1109/T-WC.2008.071305

In the first phase, the messages are transmitted by the BS to the relays in a TDMA fashion, which is throughput-optimal for a single antenna broadcast channel [17], [18]. In the second phase, the relays *cooperate* together in transmitting the received message(s) to the MSs. In this paper, we introduce an important and, hitherto ignored, fundamental notion of *asymmetric cooperation* that inevitably arises in such scenarios. This asymmetry arises because the relays might receive unequal (but overlapping) amounts of data from the BS in the first phase. The extent of this asymmetry depends on the differences in the fading states of the BS to relay channels and the duration and rate of BS transmission. This is illustrated in Fig. 1, which considers one BS, two relays, and two MSs. In Fig. 1(a), the relay cooperation is *symmetric* as both relays have the messages for both mobiles. In contrast, in Fig. 1(b), the cooperation between the relays is *asymmetric* as relay 1 has messages for both MSs, while relay 2 has a message for only one MS (MS 2). Intuitively, while the former scenario in which full transmitter cooperation is possible can achieve a higher relay-to-MS throughput in phase two than the latter asymmetric cooperation scenario, enabling full transmitter cooperation requires a larger phase one transmission time, as both messages must be conveyed to the relays rather than a subset of them.

Many recent papers have studied two-hop downlink cellular systems. Given the large body of literature in this area, we refer the interested reader to the informative overviews in [2], [10], [19], [20], and references therein, for more theoretical and practical perspectives on the problem. However, *asymmetric* relay cooperation has not been considered to the best of our knowledge. The most closely related papers are grouped and discussed below.

- **Cellular systems with non-cooperative relays:** Downlink scheduling with non-cooperative relays has been studied in [20]–[22]. The work in [20] considers an adaptive downlink system that uses either direct transmission to the MS or a two-hop transmission with relays, but does not consider relay cooperation. In [21], the authors propose a centralized downlink scheduling scheme in a cellular network that controls transmission over multiple hops of relays, resulting in higher spatial reuse. But, cooperation between relay nodes is not considered.
- **Cellular systems with symmetric relays:** Reference [23] considers the transmission of multiple messages via relays in a two-stage protocol. As before, relay cooperation is restricted

to frequency planning, and no asymmetric cooperation is considered.

• **Ad-hoc networks with cooperating relays:** A large number of papers explore cooperative two-hop relays (see, e.g., [12], [16], [22], [24]–[26]). However, they consider only the transmission of a single message via multiple, cooperating relays [16], [26] or multiple messages via a single relay [22]. Relaying with one or more multiple-antenna relays was considered in [22], [27], [28]. While multiple antennas can be interpreted as multiple (closely) cooperating relays, the co-location of the antennas precludes any asymmetry in message knowledge among antennas.

This paper focuses on the communication of multiple messages via multiple decode-and-forward relays. We provide a general formulation of the relay cooperation problem as one of optimizing the total throughput of the system, which takes into account the time spent in transmitting messages in both the BS-to-relays and relays-to-MSs transmission phases. Given its relative simplicity, practical feasibility, and good performance, we focus on linear precoding for relay cooperation. Each relay sends a linear combination of the messages it receives from the BS in the first phase, and these linear coefficients may be jointly designed. We demonstrate, through analysis and simulation, that the cases where the relays have asymmetric message knowledge are relevant and often arise when optimizing throughput. As we shall see, the extent to which this asymmetry is relevant depends on the throughput optimization criterion. We therefore optimize throughput for two diametrically opposed throughput criteria – *maximum throughput*, in which the sum throughput to all the MSs is maximized, and *extreme fairness*, in which each MS is required to be served with the same rate.[1] We prove results that analytically simplify the optimization significantly for both these fairness criteria. For analytical feasibility and simplicity of presentation, we assume that the number of messages, MSs, and relays, are all two. As we shall see, even this simple case is theoretically rich and relevant. Our results demonstrate that asymmetric cooperation is optimum in 20-30% of all cases and should thus be considered in future systems. Our contributions are the following:

- We introduce the notion of asymmetric relay cooperation, and set up a formulation of the optimum system parameters for cooperative relaying including asymmetry. This allows us to offer both quantitative and qualitative explanations why asymmetric cooperation can increase overall throughput.
- For two diametrically opposite fairness criteria, we simplify the generic formulation, which requires optimization over six parameters, to an optimization over two parameters.
- By numerical evaluation, we demonstrate that asymmetric cooperation is optimum in 20 to 30% of all cases and thus should be considered in future systems.

The remainder of the paper is structured as follows: Section II formulates the two-phase, linear precoding relay co-operation model and the throughput optimization problem for two fairness criteria. Section III categorizes the optimal parameters under the maximum throughput criterion. We see that both symmetric cooperation and asymmetric cooperation between the relays is a natural result of the optimization problem. Section IV does the same for the extreme fairness criterion. Section V numerically compares the performance of different cooperation scenarios. We conclude in Section VI.

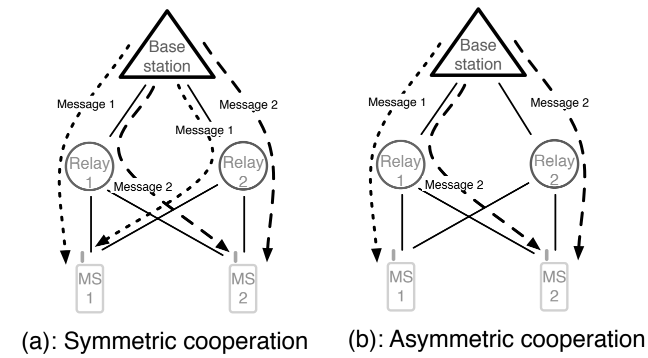


Fig. 1. An example of (a) Symmetric cooperation: relays both wish to transmit the same number of messages to the same MSs, and (b) Asymmetric cooperation: one relay wishes to transmit to two MSs, the other to one.

II. System Model and Problem Formulation

We consider the downlink in the cellular system outlined in Fig. 2 with single BS, two relays, and two mobiles, MS 1 and MS 2. We assume independent additive white Gaussian noise with zero mean, unit variance at the two relays and two MSs. We furthermore assume quasi-static fading channels [29], whose complex channel gains are all known to the BS, while other nodes require knowledge of the channels over which they are receiving. How these channel gains are obtained is beyond the scope of this work, but may, for example, be obtained through the use of feedback. The BS, relays, and MSs each have one transmit antenna. The gains of the channels between the BS and relay 1 and relay 2 are denoted as h_{BR_1} and $h_{BR_2} \in \mathbb{C}$, respectively. The channels between the two relays and the mobiles, MS 1 and MS 2 are given by $\mathbf{h}_1 = [h_{11}, h_{21}]$ and $\mathbf{h}_2 = [h_{12}, h_{22}] \in \mathbb{C}^2$, respectively, shown in Fig. 2. We now look at our two-phase model, study various message knowledge scenarios, and pose the overall optimization problem.

A. Two phase downlink communication

Transmission from the BS to the MSs takes place in two phases: during phase 1 the BS broadcasts the messages 1 and 2 sequentially in a TDMA fashion, as shown in Fig. 2. This involves the BS broadcasting message 1, of size n_1 bits, at a rate $R_1^{(1)}$ for time $t_1 = n_1/R_1^{(1)}$, followed by message 2, of size n_2 bits, at a rate $R_2^{(1)}$ for a (possibly different) time $t_2 = n_2/R_2^{(1)}$. Notice that given two of the three variables n_1, t_1, and $R_1^{(1)}$, the third may be determined (and likewise for n_2, t_2, and $R_2^{(1)}$). Messages 1 and 2 are encoded as the symbols[2]

[1]We use the term *extreme* since requiring that each user be served at the same rate each frame, regardless of channel conditions, is quite a stringent condition.

[2]In this work, we assume standard information theoretic random Gaussian codebooks and codewords [30].

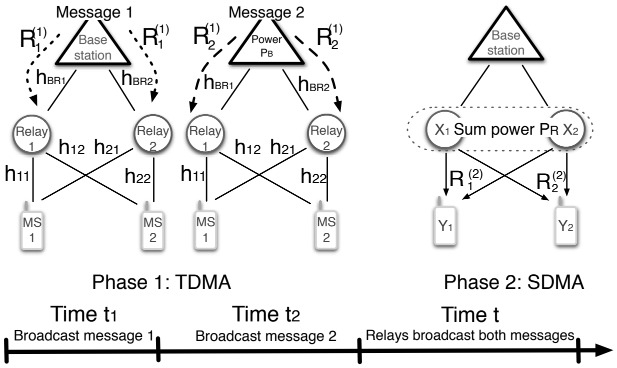


Fig. 2. Transmission takes place in two phases: in phase 1 the BS broadcasts messages in TDMA fashion: message message 1 for t_1 time units, then message 2 for t_2 time units. During phase 2 the relays simultaneously transmit all received messages to the MSs. Illustrated are the channel gains, power constraints, rates, and input-output variables.

U_1 and U_2. The BS's transmit power is limited to P_B and as we are maximizing throughput, it will always be optimal to transmit at this full power P_B. The relays are assumed to be of the decode-and-forward type – they either decode a message fully or not at all [10], [31], [32]. The TDMA structure of phase 1 exploits the broadcast nature of the wireless channel as every relay overhears the BS's transmission and may decode the transmitted messages if its rate is below the capacity of its BS-relay link. The link capacity, $\mathcal{C}_{BR_i}$, for a relay i, $i = 1$ or 2, is given by[3]

$$\mathcal{C}_{BR_i} = \log_2\left(1 + |h_{BR_i}|^2 P_B\right) \quad \text{bits/channel use.} \tag{1}$$

The TDMA structure is simple to implement, and is optimal, in terms of maximizing throughput, in the single-antenna scenario we consider [17].

The relays decode the messages in phase 1 and re-encode them for phase 2. Unlike phase 1 in which each message is transmitted in an interference-avoiding TDMA fashion from one transmitter, in phase 2, both relays simultaneously transmit linearly precoded versions of the messages they received in phase 1. While the simultaneous transmission increases data rates, it also leads to interference, which has to be mitigated through careful precoding.

Specifically, the phase 2 transmitted signal vector is denoted by $\mathbf{X} = [X_1\ X_2]'$, where X_i is the symbol transmitted by relay i. Here $\mathbf{A}'$ denotes the transpose of a matrix $\mathbf{A}$. $\mathbf{X}$ is a linear combination of the two messages, given by

$$\mathbf{X} = \mathbf{B}\mathbf{U}, \quad \text{where} \quad \mathbf{B} = \begin{bmatrix} b_{11} & b_{12} \\ b_{21} & b_{22} \end{bmatrix} \in \mathbb{C}^{2\times 2}.$$

The matrix $\mathbf{B}$ is the linear precoding matrix and $\mathbf{U} = [U_1\ U_2]'$ are the encoded messages 1 and 2, respectively. The signals Y_1 and Y_2 received at MS 1 and MS 2, respectively, are given by

$$\mathbf{Y} = \mathbf{H}\mathbf{B}\mathbf{U} + \mathbf{N} = \begin{bmatrix} h_{11} & h_{21} \\ h_{12} & h_{22} \end{bmatrix} \begin{bmatrix} b_{11} & b_{12} \\ b_{21} & b_{22} \end{bmatrix} \begin{bmatrix} U_1 \\ U_2 \end{bmatrix} + \begin{bmatrix} N_1 \\ N_2 \end{bmatrix}, \tag{2}$$

where $\mathbf{Y} = [Y_1\ \ Y_2]'$ and $\mathbf{N} = [N_1\ \ N_2]'$ is additive, white circularly symmetric Gaussian noise with covariance $E[\mathbf{N}\mathbf{N}'] = \mathbf{I}_2$, the 2×2 identity matrix. We assume, without loss of generality (w.l.o.g.), that $\mathbf{N}$ is zero-mean and $\mathbf{H}$ is invertible, which is fulfilled with probability 1 when its elements are random. The transmissions by the relays are subject to a total relay power sum constraint of P_R.[4] Assuming, w.l.o.g., that $E[\mathbf{U}\mathbf{U}'] = \mathbf{I}_2$, the sum-power constraint on the signals transmitted by the two relays becomes $|b_{11}|^2 + |b_{12}|^2 + |b_{21}|^2 + |b_{22}|^2 \leq P_R$.

In phase 2, we assume that each MS receiver decodes its respective desired signal by treating the undesired signal(s) as noise; no interference cancellation is assumed.[5] The signal to interference noise ratios (SINRs) at the two receivers, γ_1 and γ_2, and the corresponding information-theoretic phase 2 rates for the Gaussian noise channels, $R_1^{(2)}$ and $R_2^{(2)}$, are then given by

$$\gamma_1 = \frac{|h_{11}b_{11} + h_{21}b_{21}|^2}{|h_{11}b_{12} + h_{21}b_{22}|^2 + 1} \tag{3}$$

$$\Rightarrow \text{Phase 2 rate for message } 1 = R_1^{(2)} = \log_2(1 + \gamma_1) \tag{4}$$

$$\gamma_2 = \frac{|h_{12}b_{12} + h_{22}b_{22}|^2}{|h_{12}b_{11} + h_{22}b_{21}|^2 + 1} \tag{5}$$

$$\Rightarrow \text{Phase 2 rate for message } 2 = R_2^{(2)} = \log_2(1 + \gamma_2)\,. \tag{6}$$

B. Asymmetric cooperation: intuitive explanation

The system setup for phase 2 shown in (2) resembles standard space division multiple access (SDMA) with linear precoding schemes. However, there is one critical difference: depending on the channel gains h_{BR_1} and h_{BR_2} and the phase 1 transmission parameters $(n_i, t_i, R_i^{(1)})$, the relays may obtain different subsets of messages. This imposes constraints on some elements of the linear precoding matrix $\mathbf{B}$ because a relay cannot transmit a message it has not received. Assuming w.l.o.g. that $|h_{BR_1}| > |h_{BR_2}|$, we see that if relay 2 can decode the message from the BS, then relay 1 can as well. Four possible scenarios in which two messages are sent during phase 2 are illustrated in Fig. 4. We also illustrate the corresponding forms of the linear precoding matrices for phase 2; a zero element arises when a relay is unable to obtain a certain message. The figure also illustrates the phase 1 rates $(R_1^{(1)}, R_2^{(1)})$ that give rise to the four cases, which we outline next.

The four dual-message cases may be explained as follows. In Case 1, where one relay has decoded both messages, while the other has not decoded any, phase 2 corresponds to the classical broadcast channel [17], [30], [34] and two elements

[3]We drop the usual factor of 1/2 seen in the classical Shannon formula $\frac{1}{2}\log_2(1+\mathrm{SINR})$ as it a fixed scaling factor and plays no role in the overall throughput optimization that follows later on.

[4]A sum-power constraint allows for more flexibility than per-antenna power constraints as it allows a more opportunistic usage of the links and provides a more optimistic view of cooperation. However, in scenarios where the relay transmissions are limited by their power amplifiers, per-antenna power constraints may be more practical. For interesting aspects of per-antenna power constraints in downlink systems, which are beyond the scope of this paper, see [33].

[5]While this achieved rates that are pessimistic, it also ensures that they are practically and easily achievable.

of $\mathbf{B}$ are forced to 0. Case 4 corresponds to the classical SDMA problem in which the two relays jointly transmit two messages to the two MSs. In Case 2, relay 1 has both messages while relay 2 only has message 1. Therefore, the signal transmitted by relay 2, X_2, cannot contain message 2's encoding, U_2 and consequently b_{22} is forced to 0. Similarly, in Case 3, relay 1 has both messages while relay 2 has only message 2, which forces $b_{21} = 0$. We shall therefore refer to Cases 1 and 4 as *symmetric* and Cases 2 and 3 as *asymmetric*. The constraint that $\mathbf{B}$ must be triangular for the asymmetric cases changes the space of linear precoding matrices over which SDMA is optimized.

Note that the two phases are coupled. Phase 1, which determines the time needed to pass the messages from the BS to the relays, leads to different configurations of message knowledge for phase 2, and thus leads to correspondingly different transmission times required by the relays to forward the messages. In other words, complete message knowledge at the relays leads to the fastest message forwarding, but for some channel configurations the price (in terms of time required in phase 1 to achieve it) can be too high.

C. System throughput optimization

We now formulate the total throughput optimization problem. The overall throughput is the ratio of the total number of bits $n_1 + n_2$ to the total time taken to transmit them over both phases, as shown in (7).

$$\text{Total throughput} = \frac{n_1+n_2}{\frac{n_1}{R_1^{(1)}} + \frac{n_2}{R_2^{(1)}} + \max\left(\frac{n_1}{\log_2(1+\gamma_1)}, \frac{n_2}{\log_2(1+\gamma_2)}\right)}. \tag{7}$$

Phase 1 also impacts phase 2 by constraining the form the linear precoding matrix $\mathbf{B}$ can take. The overall sum throughput optimization problem may then be formulated as in (8)–(14) and involves determining the optimal rates $R_1^{(1)}$ and $R_2^{(1)}$, linear precoding matrix $\mathbf{B}$ and the number of bits n_1 and n_2. This is subject to the constraints in equations (11)–(14), which mandate that a decode-and-forward relay can only transmit a message that it has successfully decoded.

In order to minimize the time taken to transmit the messages, $R_1^{(1)}$ and $R_2^{(1)}$ must be either $\mathcal{C}_{BR_1}$ or $\mathcal{C}_{BR_2}$. Thus, there are only $2^2 = 4$ possible optimal values for the rate pairs $(R_1^{(1)}, R_2^{(1)})$ [6], which correspond to Cases 1–4 of Fig. 4. Given the unavoidable combinatorial nature of the constraints, the overall maximum is obtained by optimizing each of the four cases separately and choosing the one with the highest throughput. Here we note that determining the optimal transmit parameters, and, in particular, the optimal $\mathbf{B}$ requires knowledge of all fading coefficients. We assume this optimization is performed at the base station and that $\mathbf{B}$ is forwarded to the relays, enabling cooperation. The relays and MSs need only the channel gain parameters of the channels over which they are receiving.

[6] We deal with single message scenarios for which $R_i^{(1)} = 0$ separately.

$$\max_{\substack{\mathbf{B}, n_1, n_2\\ R_1^{(1)}, R_2^{(1)}}} \quad \frac{n_1+n_2}{\frac{n_1}{R_1^{(1)}} + \frac{n_2}{R_2^{(1)}} + \max\left(\frac{n_1}{\log_2(1+\gamma_1)}, \frac{n_2}{\log_2(1+\gamma_2)}\right)} \tag{8}$$

$$\text{s.t.} \qquad n_1, n_2 \geq 0 \tag{9}$$

$$|b_{11}|^2 + |b_{12}|^2 + |b_{21}|^2 + |b_{22}|^2 \leq P_R \tag{10}$$

$$\text{If } R_1^{(1)} \geq \log_2\left(1 + |h_{BR_1}|^2 P_B\right) \text{ then } b_{11} = 0 \tag{11}$$

$$\text{If } R_2^{(1)} \geq \log_2\left(1 + |h_{BR_1}|^2 P_B\right) \text{ then } b_{12} = 0 \tag{12}$$

$$\text{If } R_1^{(1)} \geq \log_2\left(1 + |h_{BR_2}|^2 P_B\right) \text{ then } b_{21} = 0 \tag{13}$$

$$\text{If } R_2^{(1)} \geq \log_2\left(1 + |h_{BR_2}|^2 P_B\right) \text{ then } b_{22} = 0. \tag{14}$$

D. Fairness metrics

The above formulation, which strives to maximize the total throughput, places no additional constraint on n_1 and n_2, and can even lead to one of them being zero in the optimal solution. This is the well-established *maximum throughput* metric. While it is certainly useful in a multi-user setting, it sacrifices fairness. For example, a user that consistently has a worse channel than the other user may be starved of data. We therefore also consider the diametrically opposite *extreme fairness* criterion, which mandates that the same amount of information ($n_1 = n_2$) be transmitted to each MS. The overall throughput optimization problem for this criterion is identical to that of the maximum throughput, except that (9) is replaced by $n_1 = n_2 > 0$.

III. Maximum throughput optimization

We now proceed to analytically solve the maximum throughput optimization problem. For this, we first prove two lemmas which apply to maximum throughput optimization in general. We then prove, for each of four cases, a series of lemmas/theorems that simplify the problem considerably.

Lemma 1: For a given $R_1^{(1)}$ and $R_2^{(1)}$, the optimal throughput occurs when the entire relay transmit power budget is consumed, that is, the inequality $|b_{11}|^2 + |b_{12}|^2 + |b_{21}|^2 + |b_{22}|^2 \leq P_R$ in (10) may be replaced by an equality.

Furthermore, the optimal number of bits n_1, n_2 in (8) is one of three cases. We note that since only the ratio of n_1 to n_2 is important, we can w.l.o.g. take either n_1 or n_2 to be 1.

Lemma 2: The optimal number of bits of messages 1 and 2, (n_1, n_2) is either: (i) $(n_1, n_2) = (1, 0)$, or (ii) $(n_1, n_2) = (0, 1)$, or (iii) $(n_1, n_2) = \left(\frac{\log_2(1+\gamma_1)}{\log_2(1+\gamma_2)}, 1\right)$.

We first consider and optimize the parameters for the two single-message solutions (i) $(n_1, n_2) = (1, 0)$, or (ii) $(n_1, n_2) = (0, 1)$ before turning to the more involved third dual-message solution in which n_1 and n_2 are both non-zero.

A. Single message solutions ($n_1 = 0$ or $n_2 = 0$)

As shown in Lemma 2, under the maximum throughput optimization criterion, it may be throughput-optimal to send a single message to only one of the two receivers. There are two possible optimal paths for message 1 to take, as shown in Fig. 3(a) and (b), and there are two possible optimal paths for message 2 to take, as shown in Fig. 3(c) and (d). Which of

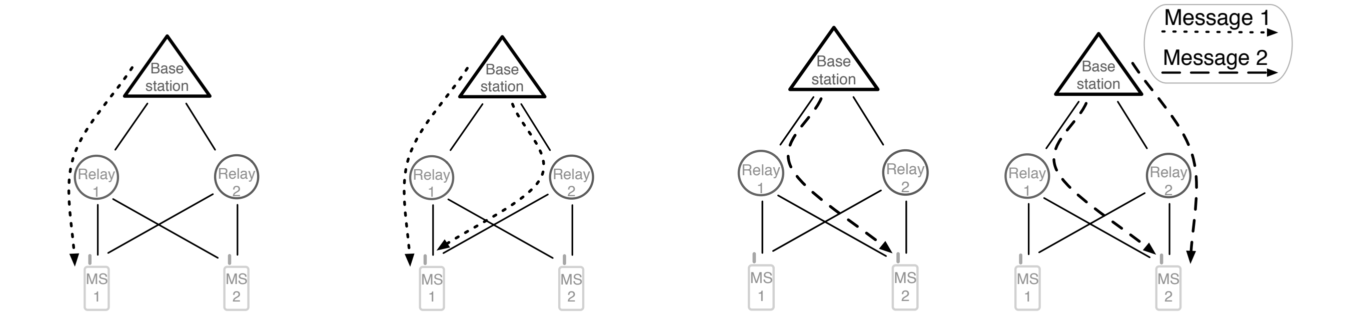


Fig. 3. Four message knowledge scenarios when $|h_{BR_1}| \geq |h_{BR_2}|$, and a single message is sent. The corresponding phase 1 and phase 2 rates are illustrated.

these four is best depends on the relative values of the channel gain amplitudes $|h_{BR_1}|$, $|h_{BR_2}|$, $|h_{11}|$, $|h_{12}|$, $|h_{21}|$ and $|h_{22}|$. When a single message is sent, the throughput is still given by (7) with the appropriate n_1 or n_2 set to 0, and $R_i^{(1)}, R_i^{(2)}$ as in Fig. 3. This yields

$$\text{Overall throughput, single message} = \frac{R_i^{(1)} R_i^{(2)}}{R_i^{(1)} + R_i^{(2)}}, \quad (15)$$

if $n_i \neq 0, i \in \{1, 2\}$.

B. Dual-message solutions

As shown in Lemma 2, it may alternatively be optimal for two messages to be sent, in which case the ratio of rates n_1/n_2 must equal the finite, positive value $x^* := \frac{\log_2(1+\gamma_1)}{\log_2(1+\gamma_2)}$. In this case, the simplified optimization problem is given by

$$\max_{\mathbf{B},\, R_1^{(1)}, R_2^{(1)}} \quad \frac{\log_2(1+\gamma_1) + \log_2(1+\gamma_2)}{\frac{1}{R_1^{(1)}}\log_2(1+\gamma_1) + \frac{1}{R_2^{(1)}}\log_2(1+\gamma_2) + 1} \quad (16)$$

$$\text{s.t.} \quad |b_{11}|^2 + |b_{12}|^2 + |b_{21}|^2 + |b_{22}|^2 = P_R. \quad (17)$$

It is also subject to the decode-and-forward conditions of (11)–(14), which are not repeated.

We now solve the optimization problem in (16)–(17) and (11)–(14) for the four cases (see Fig. 4) individually since each case imposes different constraints on $\mathbf{B}$.[7] As the novelty of this problem lies primarily in the consideration of the asymmetric cases, we devote most of our attention to Cases 2 and 3, with standard techniques applying to symmetric Cases 1 and 4. Our main result lies in Theorem 1, whose proof in the appendix makes use of a series of lemmas. For convenience, we define the function $C(x)$ as $C(x) \stackrel{\triangle}{=} \log_2(1+x)$.

1) Case 1 (Only one active relay): In Case 1, the relay with the better channel to the BS obtains both messages during phase 1, which it then transmits during phase 2 to the MSs. In this case, since $|h_{BR_1}| > |h_{BR_2}|$ as per our assumption, the phase 1 rates to relay 1 for MS 1 and MS 2 messages are $R_1^{(1)} = R_2^{(1)} = \log_2(1 + |h_{BR_1}|^2 P_B)$. Relay 2 is unable to obtain any of the messages; thus, during phase 2, we have $b_{21} = b_{22} = 0$. As the following Lemma shows, the optimization can now be simplified to a one-dimensional parameter search problem.

Lemma 3: In case 1 (only one active relay) in which w.l.o.g. $|h_{BR_1}| > |h_{BR_2}|$, the optimization problem reduces to the following single parameter search in t:

$$\max_{t \in [0, 2\pi]} \quad \frac{C\left(\frac{|h_{11}|^2 \cos^2(t) P_R}{|h_{11}|^2 \sin^2(t) P_R + 1}\right) + C\left(\frac{|h_{12}|^2 \sin^2(t) P_R}{|h_{12}|^2 \cos^2(t) P_R + 1}\right)}{\frac{C\left(\frac{|h_{11}|^2 \cos^2(t) P_R}{|h_{11}|^2 \sin^2(t) P_R + 1}\right)}{R_1^{(1)}} + \frac{C\left(\frac{|h_{12}|^2 \sin^2(t) P_R}{|h_{12}|^2 \cos^2(t) P_R + 1}\right)}{R_2^{(1)}} + 1}, \quad (18)$$

where $b_{11} = \sqrt{P_r}\cos(t)$ and $b_{12} = \sqrt{P_r}\sin(t)$, and $(n_1, n_2) = \left(\frac{C\left(\frac{|h_{11}|^2 \cos^2(t) P_R}{|h_{11}|^2 \sin^2(t) P_R + 1}\right)}{C\left(\frac{|h_{12}|^2 \sin^2(t) P_R}{|h_{12}|^2 \cos^2(t) P_R + 1}\right)}, 1 \right)$.

2) Asymmetric Cases 2 and 3: Cases 2 and 3 involve asymmetric message knowledge as relay 1 has both messages while relay 2 has only a single message. We solve the asymmetric Case 3, in which relay 1 has both message 1 and message 2, while relay 2 has only message 2, below. The optimization for Case 2 follows from it using an appropriate permutation of indices.

The linear precoding matrix $\mathbf{B}$ is thus of the form $B = \begin{bmatrix} b_{11} & b_{12} \\ 0 & b_{22} \end{bmatrix}$, which result in SINRs in phase 2 of

$$\gamma_1 = \frac{|h_{11} b_{11}|^2}{|h_{11} b_{12} + h_{21} b_{22}|^2 + 1}, \quad \gamma_2 = \frac{|h_{12} b_{12} + h_{22} b_{22}|^2}{|h_{12} b_{11}|^2 + 1}. \quad (19)$$

Consider the following change of variables from b_{11}, b_{12}, b_{22} to α, β, whose use will be made clear in Theorem 1:

$$\begin{bmatrix} \alpha \\ \beta \end{bmatrix} = \mathbf{H} \begin{bmatrix} b_{12}/b_{11} \\ b_{22}/b_{11} \end{bmatrix}, \begin{bmatrix} g_{11} & g_{12}/2 \\ g_{12}/2 & g_{22} \end{bmatrix} = (\mathbf{H}\mathbf{H}^\dagger)^{-1}. \quad (20)$$

Here $\mathbf{A}^\dagger$ denotes the conjugate transpose of the matrix $\mathbf{A}$. With this change of variables, the optimization is over six parameters: b_{11}, α, β (each is a complex number). In our main result, in Theorem 1, we derive a parametric representation for the optimization variables, which reduces the overall optimization problem from one over six variables to one over two variables.

[7]In general, if there are R relays and M mobile terminals (and, hence, M messages), M^R different cases need to be considered. Of these, only 2 are *symmetric*, while the rest are all *asymmetric*, with different relays having different, but overlapping, subsets of messages to transmit.

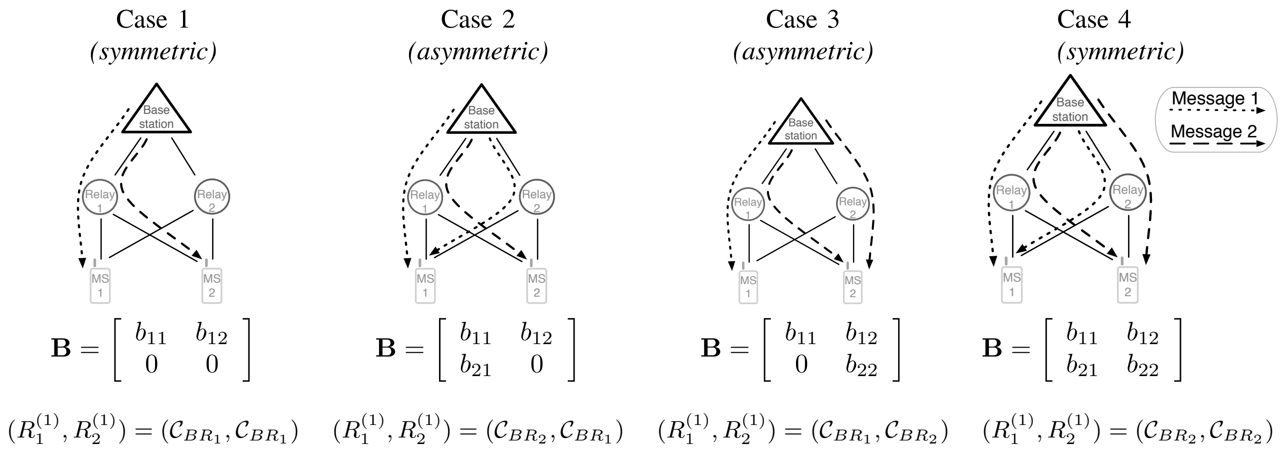


Fig. 4. Four dual-message knowledge scenarios when $|h_{BR_1}| \geq |h_{BR_2}|$, and the corresponding linear precoding matrices $\mathbf{B}$.

Theorem 1: The optimal values of $|\alpha|$ and $|\beta|$ lie on the ellipse which can be parameterized by the variables $t \in [0, 2\pi]$ and $|b_{11}|$ as

$$|\alpha(t, |b_{11}|)| = a' \cos(\phi)\cos(t) + b' \sin(\phi)\sin(t) \quad (21)$$

$$|\beta(t, |b_{11}|)| = -a' \sin(\phi)\cos(t) + b' \cos(\phi)\sin(t), \quad (22)$$

where $\phi = \frac{1}{2}\cot^{-1}\left(\frac{g_2 - g_1}{|g_{12}|}\right)$, and a' and b' are

$$(a')^2 = \frac{2(P_R/|b_{11}|^2 - 1)(g_1 g_2 - \frac{|g_{12}|^2}{4})}{(\frac{|g_{12}|^2}{4} - g_1 g_2)((g_2 - g_1)\sqrt{1 + \frac{|g_{12}|^2}{(g_1 - g_2)^2}} - g_2 - g_1)} \quad (23)$$

$$(b')^2 = \frac{2(P_R/|b_{11}|^2 - 1)(g_1 g_2 - \frac{|g_{12}|^2}{4})}{(\frac{|g_{12}|^2}{4} - g_1 g_2)((g_1 - g_2)\sqrt{1 + \frac{|g_{12}|^2}{(g_1 - g_2)^2}} - g_2 - g_1)}. \quad (24)$$

Consequently, the maximum throughput optimization problem reduces to the following two-parameter problem:

$$\max_{t, |b_{11}|} \quad \frac{C\left(\frac{|h_{11}|^2}{|\alpha(t,|b_{11}|)|^2 + 1/|b_{11}|^2}\right) + C\left(\frac{|\beta(t,|b_{11}|)|^2}{|h_{12}|^2 + 1/|b_{11}|^2}\right)}{\frac{C\left(\frac{|h_{11}|^2}{|\alpha(t,|b_{11}|)|^2 + 1/|b_{11}|^2}\right)}{R_1^{(1)}} + \frac{C\left(\frac{|\beta(t,|b_{11}|)|^2}{|h_{12}|^2 + 1/|b_{11}|^2}\right)}{R_2^{(1)}} + 1} \quad (25)$$

$$\text{s. t.} \quad |b_{11}|^2 \leq P_R, \quad t \in [0, 2\pi]. \quad (26)$$

Notice that while the symmetric Case 1 reduces to a single variable optimization problem, the asymmetric Cases 2 and 3 reduce to a two-variable optimizations.

3) Case 4 (Fully cooperating relays): In this case, both relays are able to decode both messages. This implies that the BS transmits at the worst relay's capacity in phase 1, i.e., $R_1^{(1)} = R_2^{(1)} = \log_2(1 + |h_{BR_2}|^2 P_B) = C(|h_{BR_2}|^2 P_B)$. Phase 2 then reduces to a classical 2 transmit antenna, 2 single receive antenna Gaussian multiple-input multiple-output broadcast channel under linear precoding constraints. Maximizing the sum-rate of the MIMO broadcast channel subject to linear precoding constraints is a non-convex problem whose closed-form solution remains an open and difficult problem [35], [36]. However, progress can be made along the lines of [35], [37]. We note that our problem is not equivalent to the maximum throughput problem, but is related. Since our focus is on the novel Cases 2 and 3, we simply state the Case 4 final optimization problem in (27)–(28), which is solvable using the results cited above.

$$\max_{\mathbf{B}} \quad \frac{C(\gamma_1) + C(\gamma_2)}{\frac{1}{R_1^{(1)}} C(\gamma_1) + \frac{1}{R_2^{(1)}} C(\gamma_2) + 1} \quad (27)$$

$$\text{s.t.} \quad |b_{11}|^2 + |b_{12}|^2 + |b_{21}|^2 + |b_{22}|^2 = P_R. \quad (28)$$

4) Summary of Maximum Throughput Optimization: In summary, the optimal two-phase transmission parameters $n_1, n_2, \mathbf{B}, R_1^{(1)}, R_2^{(1)}$ may be found as shown in the flowgraph in Fig. 5. The problem of linear precoding for broadcast channels is a notoriously difficult problem for which no general closed form solutions exist [35], [36]. We emphasize that our contributions lie in the introduction of the asymmetric cooperation scheme and in a significant reduction in the parameter optimization problem for this scheme.

IV. Extreme Fairness Optimization

As mentioned, the maximum throughput optimization criterion is unfair as it can lead to starvation of certain MSs. We thus optimize throughput subject to a very different *extreme fairness* criterion, which captures an alternate view of the benefits of asymmetric relay cooperation. Under extreme fairness, we require, at each time frame (or each new set of channel coefficients) that the same amount of information be sent to each MS, i.e., $n_1 = n_2$. We henceforth w.l.o.g. assume that $n_1 = n_2 = 1$. The overall throughput optimization problem is then given as follows:

$$\max_{\mathbf{B}, R_1^{(1)}, R_2^{(1)}} \quad \frac{2}{\frac{1}{R_1^{(1)}} + \frac{1}{R_2^{(1)}} + \max\left(\frac{1}{C(\gamma_1)}, \frac{1}{C(\gamma_2)}\right)} \quad (29)$$

$$\text{s. t.} \quad |b_{11}|^2 + |b_{12}|^2 + |b_{21}|^2 + |b_{22}|^2 = P_R. \quad (30)$$

As always, this is also subject to the decode-and-forward constraints of (11)–(14). It can be shown that Lemma 1, which states that the relays must transmit at the maximum allowed transmit power to maximize overall throughput, still applies. It is therefore reflected in the above formulation.

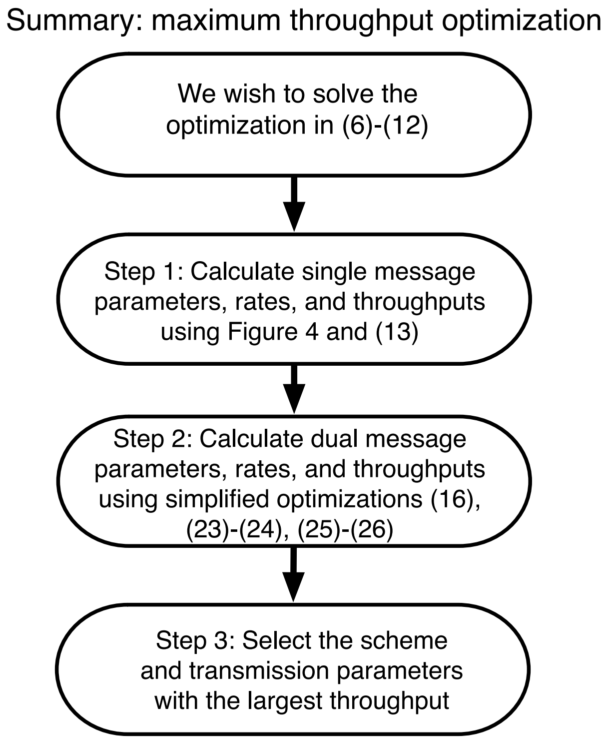


Fig. 5. Flowgraph of the maximum throughput optimization which could lead to asymmetric relay cooperation.

As in the maximum throughput criterion, four possible cases must again be considered. However, this time single-message scenarios are irrelevant, since they would lead to zero throughput as $n_1 = n_2$. Each case again corresponds to different fixed phase 1 rates. The following restatement of the optimization problem above sheds further light on the optimal parameter set:

$$\max_{R_1^{(1)}, R_2^{(1)}} \frac{2}{\frac{1}{R_1^{(1)}} + \frac{1}{R_2^{(1)}} + \min_{\mathbf{B}} \max\left(\frac{1}{C(\gamma_1)}, \frac{1}{C(\gamma_2)}\right)} \quad (31)$$

$$\text{s. t. } |b_{11}|^2 + |b_{12}|^2 + |b_{21}|^2 + |b_{22}|^2 = P_R. \quad (32)$$

The linear precoding matrix $\mathbf{B}$ only affects the denominator of (31) and may thus be considered independently. The following Lemma significantly narrows the search space of the optimal $\mathbf{B}$ for all four cases:

Lemma 4: If $f_1(t), f_2(t)$ are two continuous, differentiable functions over a compact set T, then the argmin of $\min_{t \in T} \max(1/f_1(t), 1/f_2(t))$ lies either:

1) At the boundary of T,
2) At point(s) t_X where $f_1(t_X) = f_2(t_X)$, if such point(s) exist, or
3) At a local minima of either $1/f_1(t)$ or $1/f_2(t)$.

The four final extreme fairness optimizations, are stated in their simplified form, as in (33)-(35). Cases 1, 2 and 3 are obtained using the substitutions and steps similar to Lemma 3 and Theorem 1 with $n_1 = n_2 = 1$ and the appropriate $\mathcal{C}_{BR_1}$ and $\mathcal{C}_{BR_2}$, while no simplifications are made for Case 4.

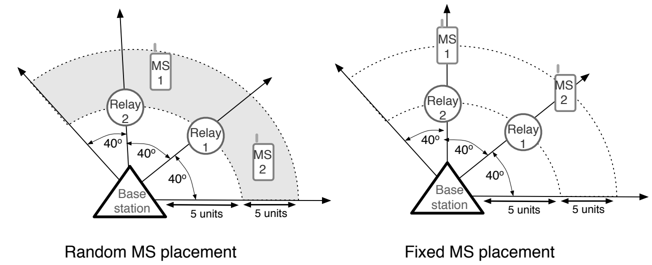


Fig. 6. Simulation setup: Relays are at equal distance from the base station and are spaced at angles 40° on an arc of radius 5 units. Two different models for mobile placement are simulated: MS positions are random in the shaded sector for the *random MS* placement, or at the fixed points in the *fixed MS* placement.

V. Simulation Results

In this section we use the analytical results of Secs. III and IV and present simulation results for the linear precoding schemes under the maximum throughput as well as the extreme fairness criteria. The throughput is evaluated over an ensemble of random channel realizations. We also track how often each of the four cases of Fig. 4 is responsible for this optimal throughput.

We assume channel gains h_{ij} between transmitter i and receiver j separated by distance d_{ij}, are given by $h_{ij} = f_{ij}/d_{ij}^2$, where f_{ij} models the fast fading of the link between i and j, which is assumed to be Rayleigh distributed and independently distributed for each link. The relays are placed along the arc of radius 5 units, and are separated by 40°, as shown in Fig. 6. We consider two geometric models for the placement of the MSs: in *random MS* placement (Figs. 7–9) the MSs are randomly placed in the shaded slice covering the distance between 5 and 10 from the BS. In *fixed MS* placement (Figs. 10–12) , the MSs are fixed on an arc of radius 10 units, and are equally spaced separated by 40°. The only randomness in the *fixed MS* placement model comes from the fading. The random MS placement will highlight average performance of the schemes (averaged over both fading and MS position), while the fixed MS placement averages only over Rayleigh fading and reflects the particular alignment of the MSs with the relays. For each of 2000 sets of geometric positions and fades, the throughputs for all four message knowledge cases are obtained by numerically solving the simplified optimization problem.

A. Comparison of symmetric and asymmetric cases

For the two geometric placements of the MSs and relays mentioned above, we compare the fraction of time each of the four message knowledge cases are selected. For the maximum throughput criterion, we also include the single message case in this comparison.

1) Random MS placement: Figure 7 demonstrates the fraction of the time each of the four cases of Fig. 4, as well as the single message case (for maximum throughput only) are optimal under the max throughput (black) and extreme fairness (grey) constraints. We can see that under the max throughput criterion, it is optimal to send a single message roughly half of the time. The symmetric Case 1 is never

Simplified Case 1:

$$\frac{2}{\frac{1}{\mathcal{C}_{BR_1}}+\frac{1}{\mathcal{C}_{BR_2}}+\min_{t\in[0,2\pi]}\max\left(\frac{1}{C\left(\frac{|h_{11}|^2\cos^2(t)P_R}{|h_{11}|^2\sin^2(t)P_R+1}\right)},\frac{1}{C\left(\frac{|h_{12}|^2\sin^2(t)P_R}{|h_{12}|^2\cos^2(t)P_R+1}\right)}\right)} \tag{33}$$

Simplified Case 3 (Case 2 analogous):

$$\frac{2}{\frac{1}{\mathcal{C}_{BR_1}}+\frac{1}{\mathcal{C}_{BR_2}}+\min_{t\in[0,2\pi],|b_{11}|\leq P_R}\max\left(\frac{1}{C\left(\frac{|h_{11}|^2}{|\alpha(t,|b_{11}|)|^2+1/|b_{11}|^2}\right)},\frac{1}{C\left(\frac{|\beta(t,|b_{11}|)|^2}{|h_{12}|^2+1/|b_{11}|^2}\right)}\right)} \tag{34}$$

Simplified Case 4:

$$\frac{2}{\frac{1}{\mathcal{C}_{BR_2}}+\frac{1}{\mathcal{C}_{BR_2}}+\min_{|b_{11}|^2+|b_{12}|^2+|b_{21}|^2+|b_{22}|^2=P_R}\max\left(\frac{1}{C\left(\frac{|h_{11}b_{11}+h_{21}b_{21}|^2}{|h_{11}b_{12}+h_{21}b_{22}|^2+1}\right)},\frac{1}{C\left(\frac{|h_{12}b_{12}+h_{22}b_{22}|^2}{|h_{12}b_{11}+h_{22}b_{21}|^2+1}\right)}\right)} \tag{35}$$

selected, while the asymmetric Cases 2 and 3 are each optimal roughly 20% of the time, and the fully symmetric Case 4 is optimal roughly 5% of the time. Thus, when it is optimal to transmit 2 messages, the asymmetric scenarios are almost always optimal. The grey bars in Fig. 7 correspond to the extreme fairness criterion. There, Case 1 is optimal 35% of the time, Cases 2 and 3 are each optimal for about 7%, and Case 4 is optimal for about 50% of the time. This indicates that full cooperation is desirable when two equal length messages must be transmitted. Thus, the selection of the best relay cooperation case is highly dependent on the criterion being optimized.

2) Fixed MS placement: Figure 10 demonstrates the fraction of time the 4 cases and the single message case are optimal when the MSs are placed at the fixed positions shown in Fig. 6. Note again that the single message case is possible only under the maximum throughput optimization criterion. Because of the geometry of the layout, where relay 1 is aligned with mobile 1 and relay 2 is aligned with mobile 2, the asymmetric Case 3 is optimal roughly 50% of the time under the max throughput criterion, in contrast to the 20% for Case 2. The single-message case is optimal about 30% of the time, while Case 4 is optimal about 3% of the time. Under the extreme fairness criterion, Case 1, 3 and 4 are optimal 25%, 5% and 70% of the time, respectively. Thus, the asymmetric case is particularly relevant for the maximum throughput optimization criterion.

We next compare the gains from relay cooperation with two conventional non-cooperative schemes. For this purpose, we show the cumulative distribution functions (CDFs) of the overall throughput. Plotting the CDF allows the visual and numerical comparison of the performance of different schemes which depend on random quantities (the MS placement and/or fading), and illustrates the spread the performance of a particular scheme is around its average. The first baseline scheme is the **round-robin with relay** scheme, in which the BS transmits to each MS in a round robin fashion, transmitting each message via the relay that has the best channel to the MS for which the message is intended. The second relevant baseline scheme that we consider is the **best two-hop** scheme in which the two-hop BS–relay j–MS i path, which takes the minimal time to transmit one unit of data, is chosen. For the extreme fairness criterion, one message is sent to *each* MS along the best 2 hop path to that MS, while for the maximum throughput criterion, only one message is sent along the best two-hop path of the best MS. The results are shown in Figs. 8, 9, 11, and 12. In both non-cooperative baselines, each MS performs maximum ratio combine (MRC) [29]) of the signals received from the BS and relay.[8]

We can see that relay cooperation, even after taking all the transmission phases into account, does result in higher sum-throughputs than the non-cooperative baselines. Despite not using any form of MRC, the relay cooperation schemes still perform better due to the spatial diversity offered by both symmetric as well as asymmetric forms of cooperation. The throughputs of the cooperative schemes are typically 20-30% higher than those of the baselines.

VI. CONCLUSION

We considered a two-hop downlink cellular system with cooperating relays. We introduced and motivated the study of *asymmetric cooperation* as a possible optimal transmission strategy in this downlink system. We set up an analytic framework for which the optimal transmission parameters are the solution of a notoriously difficult non-linear optimization problem. Through a series of algebraic manipulations and insights we reduced the optimization problem for the novel asymmetric cooperation case from one over 6 variables to one over 2 variables. Using our simplified expressions, we demonstrated through numerical simulation that the asymmetric cases are often optimal. The percentage of time that asymmetric cooperation outperforms symmetric cooperation depends highly on the fairness criteria used and the mobile placements. Thus, asymmetry should be considered when designing anything from standards to analytic frameworks involving cooperation. It is of interest to compare the rates obtained here with outer bounds, ideally the capacity, of this channel. Phase 2 itself may be a point-to-point MISO channel,

[8]Note that Cases 1-4 do not employ MRC combining, which will only improve the overall achieved throughputs. MRC was not considered so as to strictly compare the proposed cooperative schemes with other baselines without any other improvements.

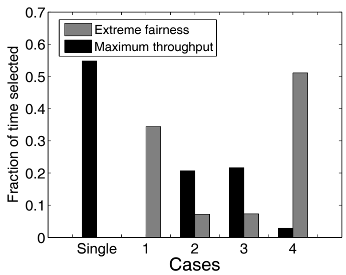


Fig. 7. Percentage of time the single message case and the 4 dual-message cases are chosen under *random MS* placement. $P_R = P_B = 1000$, radius=10 units.

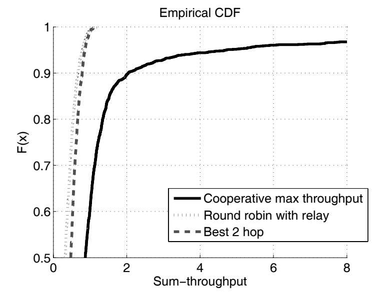


Fig. 8. CDF of sum throughput under the max throughput criterion, *random MS* placement.

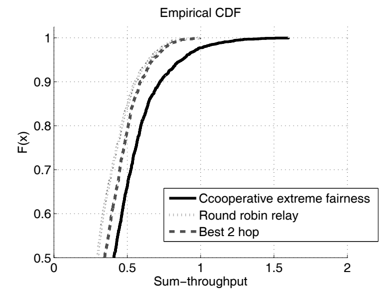


Fig. 9. CDF of sum throughput under the extreme fairness criterion, *random MS* placement.

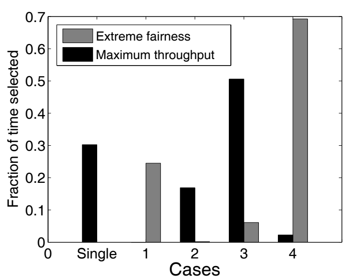


Fig. 10. Percentage of time the single message case and the 4 dual-message cases are chosen under *fixed MS* placement. $P_R = P_B = 1000$, radius=10 units.

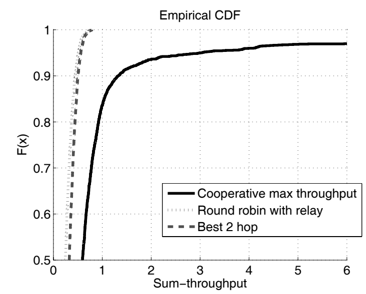


Fig. 11. CDF of sum throughput under the max throughput criterion, *fixed MS* placement.

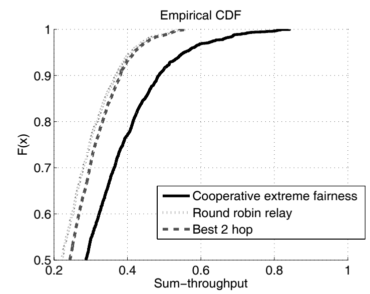


Fig. 12. CDF of sum throughput under the extreme fairness criterion, *fixed MS* placement.

an interference channel, or a MIMO broadcast channel, while phase 1 and 2 form a variant of the classical relay channel. As the capacity regions for the relay and interference channels are still in general unknown, and as our problem is more complex (and contains) these channels, we suspect that the capacity region is not trivial to obtain. Deriving tight outer bounds will thus be an interesting topic for future work.

Appendix

Proof of Lemma 1: Suppose for some $\eta > 1$ there exists an optimal solution for $b_{11}, b_{12}, b_{21}, b_{22}$ such that $\eta(|b_{11}|^2 + |b_{12}|^2 + |b_{21}|^2 + |b_{22}|^2) = P_R$ (where, depending on the case, some b_{ij} may be 0). We now show that this is impossible since we can improve upon it as follows. Consider the solution $\sqrt{\eta}b_{11}, \sqrt{\eta}b_{12}, \sqrt{\eta}b_{21}, \sqrt{\eta}b_{22}$, which still satisfies the equality constraint. Both the SINRs γ_1, γ_2 now increase, for a constant K, $\frac{\eta x}{\eta y + K} > \frac{x}{y+K}$ for $\eta > 1$. ■.

Proof of Lemma 2: Fix all variables in (8) except for n_1 and n_2. From (8)–(14), it follows that only the ratio of n_1 and n_2 is of importance. Let $x = n_1/n_2$, where it is understood that $n_2 = 0$ corresponds to $x \to \infty$. Let x^* be the point that equates the arguments of the $\max(\cdot,\cdot)$, i.e., $x^* = C(\gamma_1)/C(\gamma_2)$. The reduced objective function may be expressed as:

$$\max_{x \geq 0} \quad \frac{x+1}{\frac{x}{R_1^{(1)}} + \frac{1}{R_2^{(1)}} + \max\left(\frac{x}{C(\gamma_1)}, \frac{1}{C(\gamma_2)}\right)}$$
$$= \begin{cases} \frac{x+1}{\frac{x}{R_1^{(1)}} + \frac{1}{R_2^{(1)}} + \frac{1}{C(\gamma_2)}} & \text{for } x \in [0, x^*] \\ \frac{x+1}{\frac{x}{R_1^{(1)}} + \frac{1}{R_2^{(1)}} + \frac{x}{C(\gamma_1)}} & \text{for } x \in [x^*, \infty) \end{cases} .$$

In each interval, the objective function is of the form $f(x) \triangleq \frac{ax+b}{cx+d}$ for real constants a, b, c, d, for real x and non-negative. This function is either monotonically increasing or decreasing, since $\frac{df(x)}{dx} = \frac{ad-bc}{(cx+d)^2}$. The three optimal points are $x = \{0, x^*, \infty\}$, as captured by (ii), (iii) and (i). ■

Proof of Lemma 3: In case 1 (only one active relay), (16)–(17) may be reduced to

$$\max_{b_{11}, b_{12}} \quad \frac{C\left(\frac{|h_{11}b_{11}|^2}{|h_{11}b_{12}|^2+1}\right) + C\left(\frac{|h_{12}b_{12}|^2}{|h_{12}b_{11}|^2+1}\right)}{\frac{1}{R_1^{(1)}}C\left(\frac{|h_{11}b_{11}|^2}{|h_{11}b_{12}|^2+1}\right) + \frac{1}{R_2^{(1)}}C\left(\frac{|h_{12}b_{12}|^2}{|h_{12}b_{11}|^2+1}\right) + 1} \tag{36}$$

$$\text{s.t.} \qquad |b_{11}|^2 + |b_{12}|^2 = P_R. \tag{37}$$

From the form of (36) and (37), it is clear that the phases of b_{11} and b_{12} are irrelevant. Equation (37) is an ellipse, parameterized by, for $t \in [0, 2\pi]$, $b_{11} = \sqrt{P_R}\cos(t)$, $b_{12} = \sqrt{P_R}\sin(t)$. ■

Proof of Theorem 1: We prove Theorem 1 by making use of a series of lemmas.

Lemma 5: In Case 3, the maximum throughput optimization problem reduces to

$$\max_{b_{11}, \alpha, \beta} \quad \frac{C(\frac{|h_{11}|^2}{|\alpha|^2+1/|b_{11}|^2}) + C(\frac{|\beta|^2}{|h_{12}|^2+1/|b_{11}|^2})}{\frac{C(\frac{|h_{11}|^2}{|\alpha|^2+1/|b_{11}|^2})}{R_1^{(1)}} + \frac{C(\frac{|\beta|^2}{|h_{12}|^2+1/|b_{11}|^2})}{R_2^{(1)}} + 1} \tag{38}$$

$$\text{s.t.} \quad g_1|\alpha|^2 + 2|g_{12}||\alpha||\beta|\cos(\theta_G + \theta) + g_2|\beta|^2 = \frac{P_R}{|b_{11}|^2} - 1, \tag{39}$$

where $\theta_G = \angle g_{12}$ and $\theta = \angle \alpha\beta^*$, and the variables g_1, g_{12}, g_2, α, and β are explained below.

Proof: Setting $b_{21} = 0$ for Case 3 in the optimization problem (16)–(17) leads to SINRs as given by (19). It can be shown that the ellipsoid constraint $|b_{11}|^2+|b_{12}|^2+|b_{22}|^2 = P_R$ transforms to (39). Substituting the change of variables (20) into (19) we obtain the lemma. ■

We now have an optimization problem over six parameters b_{11}, α, β (each is a complex number). The following lemmas further reduce the number of optimization variables.

Lemma 6: The optimal solution is independent of the angle of b_{11} and the angle of α. Thus, we may w.l.o.g. assume b_{11} to be a real number in the range $[0, \sqrt{P_R}]$, and $\angle\alpha = 0$.

Proof: From the form of (38)–(39), the angle $\theta = \angle\alpha\beta^*$ between α and β is the only relevant variable and the angles of b_{11} and α are irrelevant and can be set to 0. ■

Lemma 7: The optimal θ^* satisfies $\theta_G + \theta^* = 0$ or π.

Proof: In the optimization (38)–(39) is non-convex, and thus the classical KKT-conditions are necessary but need not be sufficient for optimality. The Lagrangian is given by $\mathcal{L} = \frac{C(\frac{|h_{11}|^2}{|\alpha|^2+1/|b_{11}|^2})+C(\frac{|\beta|^2}{|h_{12}|^2+1/|b_{11}|^2})}{\frac{1}{R_1^{(1)}}C(\frac{|h_{11}|^2}{|\alpha|^2+1/|b_{11}|^2})+\frac{1}{R_2^{(1)}}C(\frac{|\beta|^2}{|h_{12}|^2+1/|b_{11}|^2})+1}$ $-\lambda(g_1|\alpha|^2+2|g_{12}||\alpha||\beta|\cos(\theta_G+\theta)+g_2|\beta|^2-P_R/|b_{11}|^2+1)$. Setting the derivative of the Lagrangian $\mathcal{L}$ with respect to the variable θ to zero yields $\frac{\partial\mathcal{L}}{\partial\theta} = \lambda\sin(\theta_G+\theta) = 0$. Since λ is not in general 0, we must have $\theta_G+\theta = k\pi$, $k \in \mathbb{Z}$. ■

We now derive a parametric representation for each ellipse for a given $|b_{11}|$, which we use to reduce the overall optimization problem to one of two variables, to complete the proof of Theorem 1. Equation (39), for fixed $|b_{11}|$ and θ is clearly the equation of an ellipse in $|\alpha|$ and $|\beta|$. A general form for an ellipse in the x-y plane is $ax^2+2bxy+cy^2+2dx+2fy+g = 0$, where a, b, c, d, f, and g are constants [38]. The center, (x_0, y_0), of this ellipse lies at $(x_0, y_0) = \left(\frac{cd-bf}{b^2-ac}, \frac{af-bd}{b^2-ac}\right)$. When $b \neq 0$, the major axis of the ellipse is oriented at an angle ϕ (in counterclockwise direction) that is given by $\phi = \frac{1}{2}\cot^{-1}\left(\frac{c-a}{2b}\right)$. The lengths of its semi-major and semi-minor axes are then:

$$(a')^2 = \frac{2(af^2+cd^2+gb^2-2bdf-acg)}{(b^2-ac)((c-a)\sqrt{1+\frac{4b^2}{(a-c)^2}}-(c+a))},$$

$$(b')^2 = \frac{2(af^2+cd^2+gb^2-2bdf-acg)}{(b^2-ac)((a-c)\sqrt{1+\frac{4b^2}{(a-c)^2}}-(c+a))}.$$

Equating coefficients, we see the ellipse in (39) has major and minor axes given by (23)–(24). From Lemma 4, its major axis is at an angle $\phi = \pm\frac{1}{2}\cot^{-1}\left(\frac{g_2-g_1}{|g_{12}|}\right)$. The $\pm$ sign does not affect the angle, but only the relative orientation about the x-axis. We may thus use only the positive orientation. The ellipse may then be parameterized in terms of $t \in [0, 2\pi]$ as in (21) and (22). ■

Proof of Lemma 4: Let t^* minimize $\max(1/f_1(t), 1/f_2(t))$. W.l.o.g., we may assume $1/f_1(t^*) \geq 1/f_2(t^*)$. Thus, the optimal objective value is $1/f_1(t^*)$. If t^* is at the boundary, or is a local minimum of $1/f_1(t^*)$ we are done. Thus suppose t^* lies strictly within the region and is not a local minimum of $1/f_1(t)$. Suppose $1/f_1(t^*) \neq 1/f_2(t^*)$. Clearly $(1/f(t))' \neq 0$ else it would be a local minimum. Thus, there exists a direction of descent, which may w.l.o.g. be assumed to be in the positive t direction. Since $1/f_1(t^*) \neq 1/f_2(t^*)$, let Δt be the distance from t^* to the closest point where $1/f_1(t^*+\Delta t) = 1/f_2(t^*+\Delta t)$. Then for $\delta t < \Delta t$, by continuity, we can find a smaller objective value, a contradiction. ■

REFERENCES

[1] I. Kang, W. Sheen, R. Chen, and S. L. C. Hsiao, "Throughput improvement with relay-augmented cellular architecture," Sept. 2005. [Online]. Available: http://www.wirelessman.org

[2] R. Pabst, B. H. Walke, D. C. Schultz, D. C. Herhold, H. Yanikomeroglu, S. Mukherjee, H. Viswanathan, M. Lott, W. Zirwas, M. Dohler, H. Aghvami, D. D. Falconer, and G. P. Fettweis, "Relay-based deployment concepts for wireless and mobile broadband radio," *IEEE Commun. Mag.*, vol. 42, no. 9, pp. 80–89, Sept. 2005.

[3] J. Zhao, I. Hammerström, M. Kuhn, A. Wittneben, M. Herdin, and G. Bauch, "Coverage analysis for cellular systems with multiple antennas using decode-and-forward relays," in *Proc. IEEE Vehicular Technology Conference (VTC)*, Apr. 2007.

[4] H. Luo, R. Ramjee, P. Sinha, L. Li, and S. Lu, "UCAN: a unified cellular and ad-hoc network architecture," in *Proc. ACM MobiCom*, 2003.

[5] I. Hammerstrom, M. Kuhn, and A. Wittenben, "Channel adaptive scheduling for cooperative relay networks," in *Proc. IEEE Vehicular Technology Conference (VTC)*, Sept. 2004.

[6] T.-Y. Ng and W. Yu, "Joint optimization of relay strategies and resource allocation in cooperative cellular networks," *IEEE J. Select. Areas Commun.*, vol. 25, no. 2, pp. 328–339, Feb. 2007.

[7] S. Mukherjee and H. Viswanathan, "Analysis of throughput gains from relays in cellular networks," in *Proc. GLOBECOM*, Nov. 2005, pp. 3471–3476.

[8] T. Liu, M. Rong, H. Shi, D. Yu, Y. Xue, and E. Schulz, "Reuse partitioning in fixed two-hop cellular relaying network," in *Proc. IEEE Wireless Communications and Networking Conference*, Apr. 2006.

[9] 802.16's Relay Task Group. [Online]. Available: http://www.ieee802.org/16/relay/

[10] G. Kramer, M. Gastpar, and P. Gupta, "Cooperative strategies and capacity theorems for relay networks," *IEEE Trans. Inform. Theory*, vol. 51, no. 9, pp. 3037–3063, Sept. 2005.

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

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

[13] A. Stefanov and E. Erkip, "Cooperative space-time coding for wireless networks," *IEEE Trans, Commun.*, vol. 53, no. 11, pp. 1804–1809, Nov. 2005.

[14] I. Hammerstrom, M. Kuhn, and A. Wittenben, "Distributed MIMO for cellular networks with multihop transmission protocols," in *Proc. Asilomar Conference on Signals, Systems and Computers*, Nov. 2006.

[15] K. E. nd J. Lee, M.-S. Do, and K. Jung, "Comparison of symmetric and asymmetric routing for fixed two-hop cellular relaying network," *IEEE Commun. Lett.*, vol. 11, no. 5, pp. 1820–1830, May 2007.

[16] M. Dohler, A. Gkelias, and H. Aghvami, "2-hop distributed MIMO communication system," *Electron. Lett.*, vol. 39, no. 18, pp. 1350–1351, Sept. 2003.

[17] H. Weingarten, Y. Steinberg, and S. Shamai, "The capacity region of the Gaussian MIMO broadcast channel," *IEEE Trans. Inform. Theory*, vol. 52, no. 9, pp. 3936–3964, Sept. 2006.

[18] N. Jindal and A. Goldsmith, "Dirty-paper coding versus TDMA for MIMO broadcast channels," *IEEE Trans. Inform. Theory*, pp. 1783–1794, May 2005.

[19] Y. Wei and R. Gitlin, "WWAN/WLAN two-hop-relay architecture for capacity enhancement," in *Proc. WNCG Wireless Networking Symposium*, Oct. 2003.

[20] N. Challa and H. Cam, "Cost-aware downlink scheduling of shared channels for cellular networks with relays," in *Proc. IEEE IPCCC*, Apr. 2004.

[21] H. Viswanathan and S. Mukherjee, "Performance of cellular networks with relays and centralized scheduling," *IEEE Trans. Wireless Commun.*, vol. 4, no. 5, pp. 2318–2323, Sept. 2005.

[22] C. Chae, T. Tang, J. R.W. Heath, and S. Cho, "MIMO relaying with linear processing for multiuser transmission in fixed relay networks," *IEEE Trans. Signal Processing*, vol. 9, no. 1, pp. 19–31, 2006.

[23] A. Agustin and J. Vidal, "TDMA cooperation using spatial reuse of the relay slotwith interfering power distribution information," in *Proc ICASSP*, May 2006.

[24] J. Zhang and T. Lok, "Multiple-source multiple-relay cooperation system," in *Proc. IEEE International Conference on Communications*, June 2006.

[25] A. Wittneben, "Coherent multiuser relaying with partial relay cooperation," in *Proc. IEEE Wireless Communications and Networking Conference*, Apr. 2006.

[26] R. Madan, N. Mehta, A. Molisch, and J. Zhang, "Energy-efficient cooperative relaying over fading channels with simple relay selection," *IEEE Trans Wireless Commun.*, vol. 7, no. 8, pp. 3013–3025, Aug. 2008.

[27] Y. Jing and B. Hassibi, "Cooperative diversity in wireless relay networks with multiple-antenna nodes," in *Proc. IEEE Int. Symp. Inf. Theory*, July 2005.

[28] S. Yiu, R. Schober, and L. Lampe, "Distributed space-time block coding for cooperative networks with multiple-antenna nodes," in *Proc. IEEE Int. Workshop on Computational Advances in Multi-Sensor Adaptive Proc.*, Dec. 2005.

[29] A. Molisch, *Wireless Communications*. Hoboken, NJ: John Wiley & Sons, 2005.

[30] T. Cover and J. Thomas, *Elements of Information Theory*. New York: John Wiley & Sons, 1991.

[31] J. Hu and N. Beaulieu, "Performance analysis of decode-and-forward relaying with selection combining," *IEEE Commun. Lett.*, vol. 11, no. 6, pp. 1820–1830, June 2007.

[32] B. Rankov and A. Wittneben, "Spectral efficient protocols for half-duplex fading relay channels," *IEEE J. Select. Areas Commun.*, vol. 25, no. 2, pp. 328–339, Feb. 2007.

[33] W. Yu and T. Lan, "Transmitter optimization for the multi-antenna downlink with per-antenna power constraints," *IEEE Trans. Signal Processing*, vol. 55, no. 6, pp. 161–176, June 2007.

[34] G. Caire and S. Shamai, "On the achievable throughput of a multi-antenna Gaussian broadcast channel," *IEEE Trans. Inform. Theory*, vol. 49, no. 7, pp. 1691–1705, July 2003.

[35] M. Stojnic, H. Vikalo, and B. Hassibi, "Rate maximization in multi-antenna broadcast channels with linear preprocessing," *IEEE Trans. Wireless Commun.*, vol. 5, no. 9, pp. 2338–2343, Sept. 2006.

[36] G. Caire, "MIMO downlink joint processing and scheduling: a survey of classical and recent results," in *Proc. Information Theory and its Applications Workshop*, Feb. 2006.

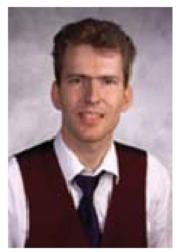
[37] A. Wiesel, Y. C. Eldar, and S. Shamai, "Linear precoding via conic optimization for fixed MIMO receivers," *IEEE Trans. Signal Processing*, vol. 54, no. 1, pp. 161–176, Jan. 2006.

[38] E. W. Weisstein, "Ellipse. from Mathworld–a Wolfram Web Resource." [Online]. Available: http://mathworld.wolfram.com/Ellipse.html

Natasha Devroye will join the University of Illinois at Chicago as an Assistant Professor in January 2009. From July 2007 until July 2008 she was a Lecturer at Harvard University. Natasha obtained her Ph.D in Engineering Sciences from the School of Engineering and Applied Sciences at Harvard University in 2007, an M.Sc from Harvard University in 2003 and a Honors B. Eng in Electrical Engineering from McGill University in 2001. She has worked as a research intern at Intel Corporation, Santa Clara, as well as Mitsubishi Electric Research Labs, in Cambridge, MA. Her interests include multi-user information theory, wireless communications, and scheduling. She in particularly focused on information theoretic limits of cognitive and cooperative communications.

Neelesh B. Mehta (S'98-M'01-SM'06) received his Bachelor of Technology degree in Electronics and Communications Engineering from the Indian Institute of Technology, Madras in 1996, and his M.S. and Ph.D. degrees in Electrical Engineering from the California Institute of Technology, Pasadena, CA in 1997 and 2001, respectively. He was a visiting graduate student researcher at Stanford University in 1999 and 2000.

He is now an Assistant Professor at the Dept. of Electrical Communication Engineering, Indian Institute of Science (IISc), Bangalore, India. Until 2002, he was a research scientist in the Wireless Systems Research group in AT&T Laboratories, Middletown, NJ. In 2002-2003, he was a Staff Scientist at Broadcom Corp., Matawan, NJ, and was involved in GPRS/EDGE cellular handset development. From 2003-2007, he was a Principal Member of Technical Staff at the Mitsubishi Electric Research Laboratories (MERL), Cambridge, MA, USA. His research includes work on link adaptation, multiple access protocols, WCDMA downlinks, system-level analysis and simulation of cellular systems, MIMO and antenna selection, and cooperative communications. He is also actively involved in radio access network physical layer (RAN1) standardization activities in 3GPP. He is a TPC co-chair for the Frontiers of Networking and Communications symposium of Chinacom 2008, and is a TPC member of Globecom 2007, Globecom 2008, WCNC 2008, and VTC 2008 (Spring).



Andreas F. Molisch (S'89, M'95, SM'00, F'05) received the Dipl. Ing., Dr. techn., and habilitation degrees from the Technical University Vienna (Austria) in 1990, 1994, and 1999, respectively. From 1991 to 2000, he was with the TU Vienna, becoming an associate professor there in 1999. From 2000-2002, he was with the Wireless Systems Research Department at AT&T (Bell) Laboratories Research in Middletown, NJ. Since then, he has been with Mitsubishi Electric Research Labs, Cambridge, MA, where he is now a Distinguished Member of Technical Staff and Chief Wireless Standards Architect. He is also professor and chairholder for radio systems at Lund University, Sweden.

Dr. Molisch has done research in the areas of SAW filters, radiative transfer in atomic vapors, atomic line filters, smart antennas, and wideband systems. His current research interests are measurement and modeling of mobile radio channels, UWB, cooperative communications, and MIMO systems. Dr. Molisch has authored, co-authored or edited four books, among them the textbook *Wireless Communications* (Wiley-IEEE Press), eleven book chapters, more than 100 journal papers, and numerous conference contributions, as well as more than 70 patents.

Dr. Molisch is an editor of the IEEE TRANSACTIONS WIRELESS COMMUNICATIONS, co-editor of special issues of several journals. He has been member of numerous TPCs, vice chair of the TPC of VTC 2005 spring, general chair of ICUWB 2006, TPC co-chair of the wireless symposium of Globecomm 2007, TPC chair of Chinacom2007, and general chair of Chinacom 2008. He has participated in the European research initiatives "COST 231", "COST 259", and "COST273", where he was chairman of the MIMO channel working group, he was chairman of the IEEE 802.15.4a channel model standardization group, and was also chairman of Commission C (signals and systems) of URSI (International Union of Radio Scientists). Dr. Molisch is a Fellow of the IEEE, a Fellow of the IET, an IEEE Distinguished Lecturer, and recipient of several awards.