An Optimization Framework for XOR-Assisted Cooperative Relaying in Cellular Networks

Hong Xu, *Student Member, IEEE*, Baochun Li, *Senior Member, IEEE*

Abstract—This work seeks to address two questions in cooperative OFDMA networks: First, how network coding based cooperative diversity can be exploited effectively when overhearing is not readily available. Second, how to realize various forms of gains available, including multi-user diversity, cooperative diversity, and network coding. The main contribution of this paper is an unifying network utility maximization framework that jointly considers relay assignment, relay strategy selection, channel assignment and power allocation. We formulate the optimization problem both with and without XOR-CD, a simple XOR-assisted cooperative diversity scheme. We show that the optimization of physical layer resource allocation with XOR-CD is equivalent to a weighted 3-set packing problem, which is NP-complete, and can be efficiently solved with provably the best approximation factor. Without XOR-CD, the problem reduces to a weighted bipartite matching problem which can be optimally solved.

Index Terms—Cooperative communication, relays, network coding, resource allocation, OFDMA, cellular networks.

1 Introduction

Network coding, a technique to allow coding capability in exchange for network capacity gain, has been utilized to improve performance in wireless networks in general [2], [3]. In the context of cooperative diversity [4]–[6], network coding has been leveraged at the relay to mix packets from different cooperative sessions, provided that the relay overhears and successfully decodes multiple transmissions and these transmissions share a common destination [7]–[9].

In this paper, we investigate the use of network coding in cooperative diversity from a new perspective of multi-channel networks. We assume the context of OFDMA [10] based cellular networks, which impose unique challenges since overhearing is no longer naturally available as in previous work. Users cannot hear each other unless tuned to the same channel. Coding opportunities are therefore to be carefully *invented* and *engineered*, rather than opportunistically *harvested*. Moreover, network coding entails that the broadcast rate is confined to the worst rate among all links involved, aggravating the task of finding profitable coding opportunities.

In light of these challenges, we propose a simple XOR-assisted cooperative diversity scheme called XOR-CD. It exploits coding opportunities on bi-directional traffic on the uplink and downlink of a mobile station (MS). Bi-directional traffic is profoundly available in cellular networks, providing abundant network coding opportunities. Fig. 1 illustrates an example to show the basic idea of XOR-assisted cooperative diversity (XOR-CD). Bi-directional traffic exists between MS and the base station (BS). The relay station (RS) performs cooperative relaying using orthogonal channels. XOR network coding can be used here to mix packets A and B at the RS and multicast a re-encoded packet $(A \oplus B)'$ using only one subchannel. Assume that channel coding and modulation are linear, $(A \oplus B)' = A' \oplus B'$. The MS and BS can still receive the intended information by XORing the coded packet with one that is known *a priori* to itself. Therefore, cooperative diversity can still be capitalized.

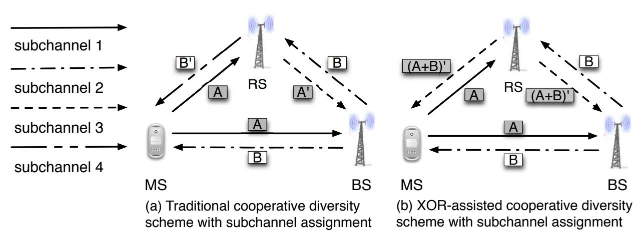


Fig. 1. The motivating scenario for XOR-CD in OFDMA networks. $(\cdot)'$ represents modulation and channel coding. With XOR-CD, only 3 subchannels are needed instead of 4 for conventional DF.

The benefits of XOR-CD are intuitive. In the ideal case where channels are symmetric, and BS-RS and MS-RS channel qualities are identical, XOR-CD achieves the same transmission rate for both cooperative sessions involved, with a saving of one subchannel and the power of one transmission compared

- *This work has been presented in part at IEEE INFOCOM, Rio de Janeiro, Brazil, April 2009 [1].*
- *The authors are with The Edward S. Rogers Sr. Department of Electrical and Computer Engineering, University of Toronto, Toronto, Ontario M5S 3G4, Canada. Email:* {henryxu, bli}*@eecg.toronto.edu. Hong Xu will join the Department of Computer Science, City University of Hong Kong as an Assistant Professor starting from August 2013.*

to the conventional Decode-and-Forward. The saved subchannel and power can be used to accommodate more cooperative sessions, thereby further improving the network throughput.

The question then becomes how to effectively reap the promising gains of XOR-CD in OFDMA networks. Three kinds of gains can be exploited here: (i) *multi-user diversity gain*: for a given data/relay subchannel, different MS experience independent fading, allowing us to assign a subchannel to the MS with the largest channel gain; (ii) *cooperative diversity gain*: The RS helps the intended receiver to combat fading and improve SNR through cooperative relaying; (iii) *network coding gain*: bi-directional traffic is amenable to network coding which is utilized at RS to make relaying more resource efficient, increasing the network capacity.

Our main contribution in this paper is a unifying network utility maximization framework (NUM) to tackle the above problem. It jointly considers the following dimensions of resource allocation: **r**elay assignment, relay-**s**trategy selection, and **s**ubchannel assignment for both MS and RS in a cell, which is referred to as the RSS-XOR problem. Through dual decoupling, we show that the cross-layer problem can be decoupled into two subproblems: an application layer rate adaptation problem that is trivial to solve, and a physical layer resource allocation problem which is much more difficult.

Specifically, we prove that the RSS-XOR physical layer resource allocation problem is NP-complete by transforming it into a weighted 3-set packing problem. We propose a polynomial-time algorithm to solve it with the best known constant approximation factor, based on an algorithm for the weighted independent set problem. We also formulate the optimization problem with only conventional Decode-and-Forward cooperative diversity, which is referred to as the NO-XOR problem. Using the same decoupling technique, we design an efficient algorithm that optimally solves the NO-XOR physical layer resource allocation problem as a weighted bipartite matching problem. Finally, we extend to consider relay power allocation among cooperative sessions for both RSS-XOR and NO-XOR, and propose subgradient-based algorithms to solve them in the dual domain.

The remainder of this paper is structured as follows. Sec. 2 summarizes related work, and Sec. 3 introduces our system models. In Sec. 4 we formally present our NUM framework, the RSS-XOR optimization problem and its counterpart NO-XOR problem, and extend both models for power allocation. In Sec. 5, we present efficient algorithms to solve the difficult physical layer resource allocation problems. We conduct extensive simulations to verify the effectiveness of algorithms in Sec. 6. Finally we give concluding remarks in Sec. 7.

2 Related Work

This paper builds upon prior work on cooperative diversity, whose roots can be traced back to the relay channel model studied in [11]. The popularity of cooperative diversity is owed to [4]–[6] where different relay strategies are developed. Recent research aims to exploit distributed antennas on neighboring nodes in the network, and has resulted in many protocols at both the physical layer [12], [13] and the network layer [14]. We also build on work on network coding introduced in [15]. It is shown in [2], [16]–[19] that network coding combined with routing and scheduling can greatly improve throughput in wireless multi-hop networks. A similar conclusion is made for the information exchange paradigm in which two nodes exchange data via a relay [3], a scenario similar to ours. However, the distinction is clear: in these prior work, cooperative diversity is not leveraged as a mechanism to combat fading, since the two nodes cannot directly communicate without the relay.

Recently there are studies that incorporate network coding into cooperative communication. [8] is arguably the first work that studies the diversity gain of network coding. Adaptive network coded cooperation, the idea of which is to match network-on-graph with code-on-graph to construct efficient network codes accounting for changing topology and lossy nature of wireless networks, is studied in [20]. In [9], a network coding based cooperative diversity scheme is proposed. The focus of these work is on the analysis of diversity-multiplexing tradeoff of network coded cooperative diversity. They rely on the overhearing assumption with the single shared channel model, while XOR-CD assumes a multi-channel setting and focuses on the resource allocation problem to realize the promise of network coding. In this regard our work also differs from studies of joint network and channel coding in two-way relay channel [21], [22].

Resource allocation in cooperative networks has been extensively studied in previous work as well [23]–[26]. [23] considers power allocation for a simple triangle network with one pair of source-destination and one relay. [24] considers multi-hop ad hoc networks and proposes a framework that jointly considers routing, relay selection and power allocation. [25], [26] considers OFDMA based cellular networks and are most related to our work. [25] studies channel assignment and power allocation for multi-hop OFDMA networks. [26] proposes solutions for joint optimization of channel assignment, relay strategy selection and power allocation in OFDMA cellular networks based on conventional Amplify-and-Forward and Decode-and-Forward.

Different from these efforts (summarized in Table 1), this work represents an early attempt to study cooperative relaying in multi-channel networks with the use of network coding. We propose a novel di-

TABLE 1
Related studies to this paper.

	Coding diversity	Channel assignment	Relay strategy	Power allocation
This paper	√	√	√	√
[8], [9], [20]	√	×	×	×
[23]	×	×	×	√
[24]	×	×	√	√
[25]	×	√	×	√
[26]	×	√	√	√

TABLE 2
Key notations used in this paper.

ζ	set of data subchannels
ψ	set of relay subchannels
Ω	set of mobile stations
Φ	set of relay stations
$\mathcal{L}$	set of links
$M(l)$	MS corresponding to link l
c_i	data subchannel $c_i \in \zeta$
c_r	relay subchannel $c_r \in \psi$
r	relay station $r \in \Phi$
s	mobile station $s \in \Omega$
$\sigma_{l,c}$	channel gain to noise ratio on link l channel c
P	fixed power budget for direct transmission
p_l^{r,c_i,c_r}	conventional CD power allocation on link l
p_s^{r,c_i,c_j,c_r}	XOR-CD power allocation on mobile station s

versity scheme with XOR that improves the resource efficiency of relaying and thereby boosts throughput performance. More importantly, we present a cross-layer optimization framework to address the resource allocation problem for the cooperative network. Our framework jointly considers network coding, channel assignment, relay strategy selection and power allocation, which has not yet been discussed to our knowledge.

Finally, our conference version [1] focuses on solving the physical layer resource allocation problem with a specific utility function. In this work, we adopt a network utility maximization framework that generalizes to encompass many possible choices of utility functions at the upper layer. We also show that the upper layer rate allocation problem and the lower layer resource allocation problem can be decoupled and solved independently through a subgradient method, which is not present in [1].

3 System Models

In this section, we introduce the underlying system models for our optimization framework.

3.1 Network Models

We consider a single-cell OFDMA network. The BS is communicating with each MS with bi-directional traffic. The system operates in FDD mode, meaning that the uplink and downlink of a MS are assigned orthogonal sets of subchannels. A small number of RS are employed in the cell to provide cooperative diversity. They may help some MS for transmissions on their *data subchannels*, using *relay subchannels* from a relay channel pool orthogonal to the data channel pool. One relay subchannel is used to support only one data subchannel of a MS in conventional cooperative diversity (CD). In the case of XOR-CD, one relay subchannel is used to support two data subchannels, one on the uplink and one on the downlink, as we illustrated in Sec. 1. We further assume that the BS and MS have infinite backlog of traffic. It is then the case that cooperative transmissions progress concurrently with direct data transmissions, which we shall discuss in more detail in Sec. 3.4.2. Decode-and-Forward (DF) is used as the conventional CD scheme.

3.2 Channel Models

We model the wireless fading environment by large scale path loss and shadowing, along with small scale frequency-selective Rayleigh fading. Fading between different subchannels are independent. We assume the network operates in a slow fading environment, so that channel estimation is possible and full channel side-information (CSI) is available, which makes the optimization feasible. In a practical system, channel estimation is generally done at the receiver end and fed back to the base station, which then solves the optimization and informs all MS and RS the channel assignment, power levels, and relay strategies. Such assumptions about the fading environment are commonly used as in [23], [25], [26]. Also note that when the environment has fast fading components, optimization may be done in a statistical sense.

In practical systems, it is usually not feasible to assign arbitrary subchannels for the uplink and downlink transmissions due to self-interference. Usually separate chunks of frequency bands are allocated in FDD mode. Such a practical constraint does not contradict our channel model, however, as we can view any interference-free chunk of frequency bands as a subchannel, which is the basic unit of channel allocation in our problem. This constraint limits the flexibility of channel allocation, and negatively impacts the throughput performance of cooperative and direct transmissions. Nevertheless, it does not affect the relative performance improvement of XOR-CD over comparative schemes, since the same constraint exists for all types of transmissions.

An equal amount of power P is allocated for both direct and relay transmissions across all data and relay subchannels. In the extended models with relay power allocation, however, RS can adjust the power level for each of the relay subchannels they use in order to confine themselves to their power budget.

3.3 Notations

Denote ζ, ψ, Ω and Φ as the set of data subchannels, relay subchannels, MS, and RS, respectively. $s \in \Omega$

denotes a MS and $r \in \Phi$ denotes a RS r, respectively. $l \in \mathcal{L}$ denotes a directed link from the source $S(l)$ to the destination $D(l)$ where $\mathcal{L}$ is the set of links. Each link, being an uplink or downlink, has a corresponding MS s such that $S(l) = s$ or $D(l) = s$. Let $M(l) = s$ denote this relationship between l and s. Each link can operate in one and *only* one of three modes, namely the direct transmission mode, conventional CD mode and XOR-CD mode, depending on the choice of relay strategy.

Define the function $R(c_i, l)$ as the achievable rate of direct transmission on link l when it is assigned with subchannel c_i. For conventional CD, $R(c_i, c_r, r, p_l^{r,c_i,c_r}, l)$ is the achievable rate function of l, when RS r is assigned to be the relay for transmission on data subchannel c_i, with allocated power p_l^{r,c_i,c_r} on relay subchannel c_r. For XOR-CD, $R(c_i, c_j, c_r, r, p_s^{r,c_i,c_j,c_r}, s)$ denotes the achievable rate function if r is the relay of s for its uplink transmission on c_i and downlink transmission on c_j, with allocated power p_s^{r,c_i,c_j,c_r} on relay subchannel c_r.

To assist the understanding of the analysis, we summarize the key notations used throughout the paper in Table. 2.

3.4 An Information Theoretic Analysis

We first provide an information theoretical analysis in order to derive the rate functions for three transmission modes, especially XOR-CD. The setup, including the complex channel gains on different links, is shown in Fig. 2. Noises are modeled as i.i.d. circularly symmetric complex Gaussian noises $\mathcal{CN}(0, N_0 W)$.

3.4.1 Direct Transmission

For direct transmission of say link AB, the achievable rate is found using the well-known formula (in b/s/Hz):

$$R(AB, c_1) = \log_2\Big(1 + \frac{P \cdot |h_{AB,c_1}|^2}{\Gamma N_0 W}\Big), \qquad (3.1)$$

where Γ is the gap to capacity and P denotes the direct transmission power. For notational convenience we denote $\frac{|h_{AB,c_1}|^2}{\Gamma N_0 W}$ as σ_{AB,c_1}, where σ represents the channel gain-to-noise ratio. Then the rate function can be simply expressed as:

$$R(AB, c_1) = \log_2(1 + P \cdot \sigma_{AB,c_1}). \qquad (3.2)$$

3.4.2 Conventional CD

For the DF relay transmission for the traffic A to B, assuming A is a MS and B is the BS, first R_0 attempts to decode A's message. If decoding is successful, R_0 transmits to B with power $p_{AB}^{R_0,c_1,c_4}$ using relay subchannel c_4 as depicted in Fig. 2. Therefore, the maximum rate for this mode can be found to be

$$R(c_1, c_4, R_0, p_{AB}^{R_0,c_1,c_4}, AB) = \min\{\log_2(1 + P \cdot \sigma_{AR_0,c_1}), \log_2(1 + P \cdot \sigma_{AB,c_1} + p_{AB}^{R_0,c_1,c_4} \cdot \sigma_{R_0B,c_4})\}. \qquad (3.3)$$

Compared to the result in [6], ours does not have $\frac{1}{2}$ before the expression. The reason is that a two-slot implementation is assumed in [6] with a shared channel, whereas in our OFDMA-based multi-channel system, relay transmissions progress concurrently with direct transmissions on orthogonal channels. At any time slot, the concurrent relay transmission carries the message it received from the direct transmission in the previous time slot. For a reasonably long time period, say hundreds of time slots, the one-time-slot lead time can be safely ignored.

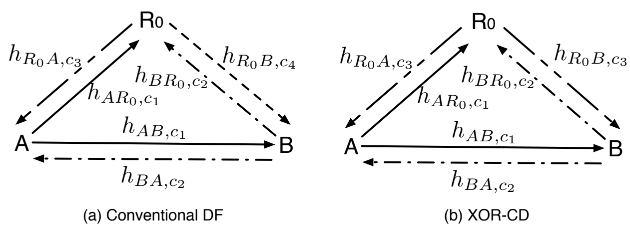


Fig. 2. The channel models for Decode-and-Forward and XOR-CD, where $h_{l,c}$ denotes channel gain of link l when it is assigned subchannel c.

Inspecting the rate function, we can see that increasing the relay power will first increase the rate, but not any more after reaching a threshold, since the relay cannot deliver more information than what it can decode. Thus, the threshold value of the relay power is such that

$$R(c_1, c_4, R_0, \tilde{p}_{AB}^{R_0,c_1,c_4}, AB) = \log_2(1 + P \cdot \sigma_{AR_0,c_1}) = \log_2(1 + P \cdot \sigma_{AB,c_1} + \tilde{p}_{AB}^{R_0,c_1,c_4} \cdot \sigma_{R_0B,c_4}), \qquad (3.4)$$

which gives us

$$\tilde{p}_{AB}^{R_0,c_1,c_4} = \frac{\sigma_{AR_0,c_1} - \sigma_{AB,c_1}}{\sigma_{R_0B,c_4}} P. \qquad (3.5)$$

Similar analysis can be carried out for $R(c_2, c_3, R_0, p_{BA}^{R_0,c_2,c_3}, BA)$ and the threshold power $\tilde{p}_{BA}^{R_0,c_2,c_3}$. These calculations will be used later for the power control problem in Sec. 4.4.

3.4.3 XOR-CD

The relay transmissions from R_0 to a MS A and the BS B are done by performing XOR over the two messages and multicasting using a single relay subchannel c_3. Therefore, the rate of this mode, for each of the two links involved, can be shown to be

$$R(c_1, c_2, c_3, R_0, p_A^{R_0,c_1,c_2,c_3}, A) = \min\{\log_2(1 + P \cdot \sigma_{AR_0,c_1}), \log_2(1 + P \cdot \sigma_{BR_0,c_2}), \log_2(1 + P \cdot \sigma_{AB,c_1} + p_A^{R_0,c_1,c_2,c_3} \cdot \sigma_{R_0B,c_3}), \log_2(1 + P \cdot \sigma_{BA,c_2} + p_A^{R_0,c_1,c_2,c_3} \cdot \sigma_{R_0A,c_3})\}, \qquad (3.6)$$

The first two terms in (3.6) represent the maximum rate at which the relay can reliably decode the source messages from both A and B, while the last two terms represent the maximum rate at which A and B can reliably decode their intended message given repeated transmissions from R_0's multicast, respectively. Note that the uplink and downlink flows have the same rate given by (3.6).

Again the threshold value of relay power at R_0 is such that

$$R(c_1, c_2, c_3, R_0, \tilde{p}_A^{R_0,c_1,c_2,c_3}, A) = \min\{\log_2(1 + P \cdot \sigma_{AR_0,c_1}), \log_2(1 + P \cdot \sigma_{BR_0,c_2})\}.$$

The detailed expression for $\tilde{p}_A^{R_0,c_1,c_2,c_3}$ can then be derived.

4 An Optimization Framework

We present our optimization framework in this section. After introducing the network utility maximization framework, we first present RSS-XOR, our formulation for the joint channel assignment, relay assignment, and relay strategy selection problem with XOR-CD. We then provide another formulation with conventional diversity schemes only, i.e. the NO-XOR problem. We extend both formulations by considering relay power allocation. Finally we demonstrate that both problems under the NUM framework can be solved in the dual domain as a cross-layer optimization problem.

4.1 The Network Utility Maximization Framework

We adopt a network utility maximization framework where each data stream of a particular link has a utility function, and the overall objective is to maximize the total utility of the network. The network utility maximization framework (NUM) is originated from the seminal work of Kelly [27], and has been extensively applied to cross-layer design problems in wireless networking [28]. A utility function is a concave and increasing function of the link throughput that reflects a MS's satisfaction. Depending on the application the traffic is serving (e.g. voice, data), the utility function can take on different shapes.

Denote the throughput of link l as d_l, then utility function can be denoted as $U_l(d_l)$. The objective of the optimization can be expressed as:

$$\max_{d_l} \sum_{l \in \mathcal{L}} U_l(d_l) \qquad (4.1)$$

4.2 The RSS-XOR Problem

Our goal is to optimize the strategies of assigning appropriate relay subchannels to RS and data subchannels to MS, and pairing RS to the data subchannels of MS with different choices of relay strategies, in order to maximize the aggregated utility. We now present the optimization constraints that reflect these considerations in the following.

For both uplink and downlink, traffic falls into three classes corresponding to the three transmission modes, namely direct traffic, conventional CD traffic, and XOR-CD traffic. Introduce three 0–1 decision variables $x_l^{c_i}$, y_l^{r,c_i,c_r}, and z_s^{r,c_i,c_j,c_r}. $x_l^{c_i}$ indicates whether link l on data subchannel c_i is performing direct transmission. y_l^{r,c_i,c_r} indicates whether link l is operating in conventional CD mode with RS r and data-relay subchannel pair (c_i, c_r). Each MS may be assigned multiple such channel pairs depending on the instantaneous channel condition. z_s^{r,c_i,c_j,c_r} indicates whether MS s is assigned with RS r and relay subchannel c_r for its uplink on data subchannel c_i and downlink on c_j for XOR-CD.

Since an equal amount of power P is used for each direct and relay transmission, throughput of link l can be characterized as follows:

$$d_l = \sum_{c_i \in \zeta} R(c_i, l) x_l^{c_i} + \sum_{c_i \in \zeta, c_r \in \psi, r \in \Phi} R\big(c_i, c_r, r, P, l\big) y_l^{r,c_i,c_r} + \sum_{c_i, c_j \in \zeta, c_r \in \psi, r \in \Phi} R(c_i, c_j, c_r, r, P, s) z_s^{r,c_i,c_j,c_r},$$
$$\text{where } s = M(l), \ \forall l \in \mathcal{L}. \qquad (4.2)$$

Recall that each data subchannel can only be assigned to one link which operates in one of the three modes. Therefore,

$$\sum_{l \in \mathcal{L}} \Big(x_l^{c_i} + \sum_{c_r \in \psi, r \in \Phi} y_l^{r,c_i,c_r} \Big) + \sum_{s \in \Omega, r \in \Phi, c_j \in \zeta, c_r \in \psi} \Big(z_s^{r,c_i,c_j,c_r} + z_s^{r,c_j,c_i,c_r} \Big) \leq 1, \forall c_i \in \zeta, \qquad (4.3)$$

where the first term accounts for the possibility that c_i is assigned for direct and conventional CD modes, and the second term accounts for the possibility of XOR-CD. Notice that this constraint also implicitly takes into consideration that each link can only operate in one of the three modes.

Similarly, each relay subchannel can be assigned to only one cooperative session, be it conventional CD session or XOR-CD session.

$$\sum_{l \in \mathcal{L}} \sum_{r \in \Phi} \sum_{c_i \in \zeta} y_l^{r,c_i,c_r} + \sum_{s \in \Omega} \sum_{r \in \Phi} \sum_{c_i \in \zeta} \sum_{c_j \in \zeta} z_s^{r,c_i,c_j,c_r} \leq 1, \quad \forall c_r \in \psi. \qquad (4.4)$$

Consequently, the RSS-XOR problem becomes an integer program, with the objective (4.1) subject to constraints (4.2), (4.3), and (4.4). For ease of presentation, we use $\boldsymbol{x}, \boldsymbol{y}$ and $\boldsymbol{z}$ to represent all the $x_l^{c_i}$s, y_l^{r,c_i,c_r}s, and z_s^{r,c_i,c_j,c_r}s as optimizing variables.

$$\text{RSS-XOR: } \max_{\boldsymbol{x},\boldsymbol{y},\boldsymbol{z}} \sum_{l \in \mathcal{L}} U_l(d_l) \quad \text{s.t. } (4.2), (4.3), \text{ and } (4.4). \qquad (4.5)$$

4.3 The NO-XOR Problem

We also provide the optimization formulation under NUM with only conventional cooperative diversity, *i.e.*, the NO-XOR problem, which is studied as a baseline comparison. It can be readily formulated in a similar way as the RSS-XOR problem, with z_s^{r,c_i,c_j,c_r} equal to zero for any $c_i, c_j \in \zeta$, $c_r \in \psi$, $s \in \Omega$, $r \in \Phi$. Formally,

$$\begin{aligned}\text{NO-XOR: } \max_{\boldsymbol{x},\boldsymbol{y}} & \sum_{l\in\mathcal{L}} U_l(d_l) \\ \text{s.t. } d_l = & \sum_{c_i\in\zeta} R(c_i,l)x_l^{c_i} \\ & + \sum_{c_i\in\zeta}\sum_{c_r\in\psi}\sum_{r\in\Phi} R(c_i,c_r,r,P,l)y_l^{r,c_i,c_r}, \\ & \sum_{l\in\mathcal{L}} x_l^{c_i} + \sum_{l\in\mathcal{L}}\sum_{r\in\Phi}\sum_{c_r\in\psi} y_l^{r,c_i,c_r} \le 1, \forall c_i \in \zeta, \\ & \sum_{l\in\mathcal{L}}\sum_{r\in\Phi}\sum_{c_i\in\zeta} y_l^{r,c_i,c_r} \le 1, \forall c_r \in \psi. \end{aligned} \quad (4.6)$$

4.4 Power Allocation

We can extend the two models by incorporating an additional constraint that each RS has a limited power budget. A RS then has to allocate the right amount of power across all the cooperative sessions it supports in order to maximize the total utility. Mathematically, the throughput constraints of both problems are updated by using $R(c_i, c_r, r, p_l^{r,c,c_r}, l)$ to replace $R(c_i, c_r, r, P, l)$ in (4.2), and using $R(c_i, c_j, c_r, r, p_s^{r,c_i,c_j,c_r}, s)$ to replace $R(c_i, c_j, c_r, r, P, s)$ in (4.6).

The constraint that the total power of RS cannot exceed its budget can be expressed as follows for the RSS-XOR problem:

$$\sum_{l\in\mathcal{L}}\sum_{c_i\in\zeta}\sum_{c_r\in\psi} p_l^{r,c_i,c_r} + \sum_{s\in\Omega}\sum_{c_i\in\zeta}\sum_{c_j\in\zeta}\sum_{c_r\in\psi} p_s^{r,c_i,c_j,c_r} \le P_r, \forall r. \quad (4.7)$$

where P_r denotes the power budget of RS r. RSS-XOR with power allocation can be formulated by adding constraint (4.7) into the original formulation.

For NO-XOR, the power constraint is simply:

$$\sum_{l\in\mathcal{L}}\sum_{c_i\in\zeta}\sum_{c_r\in\psi} p_l^{r,c_i,c_r} \le P_r, \forall r \in \Phi. \quad (4.8)$$

The power allocation version of NO-XOR is similarly formulated by adding constraint (4.8) into (4.6).

4.5 Cross-layer Optimization in the Dual Domain

Both RSS-XOR and NO-XOR are non-convex problems because of the integer constraints $\boldsymbol{x}$, $\boldsymbol{y}$ and $\boldsymbol{z}$. Duality gap for non-convex problems is non-zero in general. However, in an OFDMA system with many narrow subchannels, the optimal solutions of RSS-XOR and NO-XOR are always convex functions of P, because if two sets of throughput using two different channel-RS-link assignments and relay strategies are achievable individually, their linear combination is also achievable by a frequency-division multiplexing of the two sets of strategies. This idea for non-convex problems of multi-carrier systems is discussed in [29]. In particular, using the duality theory of [29], the following is true:

Proposition 1: The RSS-XOR and NO-XOR problems, with the discrete selection of channels, RS and relay strategies, have zero duality gap in the limit as the number of OFDM subchannels goes to infinity.

A detailed proof can be constructed along the same line of argument as in [29]. This proposition allows us to solve non-convex problems in their dual domain. Although it requires number of channels to go to infinity, in reality the duality gap is very close to zero as long as number of channels is large [26].

With this proposition, we show that the RSS-XOR and NO-XOR problems can be decoupled into an application layer rate adaption problem and a physical layer resource allocation problem, and be solved by solving these two problems separately. Our technique is reminiscent of that in [26]. We focus on the basic RSS-XOR problem, while the technique can be easily applied to the NO-XOR problem and their power allocation extensions as well.

First, introduce a new variable $\boldsymbol{t} = [t_1, \ldots, t_l, \ldots, t_{|\mathcal{L}|}]$, and rewrite the RSS-XOR problem as follows:

$$\begin{aligned}\max_{\boldsymbol{t},\boldsymbol{x},\boldsymbol{y},\boldsymbol{z}} & \sum_{l\in\mathcal{L}} U_l(t_l) \\ \text{s.t. } & d_l \ge t_l, \forall l \in \mathcal{L}, \\ & (4.2), (4.3), \text{ and } (4.4). \end{aligned} \quad (4.9)$$

Because U_l is an increasing function, when the objective of (4.9) is maximized, t_l must be equal to d_l. Thus (4.5) and (4.9) must have the same solution. The key step to decompose the problem is to relax the new constraint $d_l \ge t_l$. The Lagrangian becomes

$$L(\boldsymbol{\lambda}, \boldsymbol{t}, \boldsymbol{x}, \boldsymbol{y}, \boldsymbol{z}) = \sum_{l\in\mathcal{L}} \left(U_l(t_l) + \lambda_l(d_l - t_l)\right). \quad (4.10)$$

where λ_l being a dual variable corresponding to link l. Observe the dual function

$$g(\boldsymbol{\lambda}) = \begin{cases} \max_{\boldsymbol{t},\boldsymbol{x},\boldsymbol{y},\boldsymbol{z}} & L(\boldsymbol{\lambda}, \boldsymbol{t}, \boldsymbol{x}, \boldsymbol{y}, \boldsymbol{z}) \\ \text{s.t.} & (4.2), (4.3), \text{ and } (4.4). \end{cases} \quad (4.11)$$

now consists of two sets of variables: application layer variable $\boldsymbol{t}$, and physical layer variables $\boldsymbol{x}, \boldsymbol{y}, \boldsymbol{z}$. It can be readily separated into two maximization subproblems, one rate adaptation problem in the application layer,

$$g_{app}(\boldsymbol{\lambda}) = \max_{\boldsymbol{t}} \sum_{l\in\mathcal{L}} (U_l(t_l) - \lambda_l t_l), \quad (4.12)$$

and a resource allocation problem in the physical layer,

$$g_{phy}(\boldsymbol{\lambda}) = \begin{cases} \max\limits_{\boldsymbol{x},\boldsymbol{y},\boldsymbol{z}} & \sum\limits_{l\in\mathcal{L}} \lambda_l d_l \\ \text{s.t.} & (4.2), (4.3), \text{ and } (4.4). \end{cases} \quad (4.13)$$

We can see that the optimization formulation provides a layered approach to the network utility maximization problem. The use of dual variable $\boldsymbol{\lambda}$ controls the interaction between the layers. It can be interpreted as a price signal that coordinates the throughput supply and demand relationship between the physical and application layer. The physical layer attempts to maximize the total revenue given the per-link rate price λ_l. A higher value of λ_l attracts the physical layer to allocate more resources to l. The application layer, on the other hand, tries to maximize the total net utility, given the per-link rate cost λ_l. A higher value of λ_l causes the application layer to reduce its demand for throughput.

Finally, since the network utility maximization problem (4.9) has zero duality gap, it can be solved by minimizing the dual objective:

$$\begin{aligned} &\min\ g(\boldsymbol{\lambda}) \\ &\text{s.t. } \boldsymbol{\lambda} \succeq \mathbf{0}. \end{aligned} \quad (4.14)$$

One way to solve the dual problem is to use a subgradient method that updates $\boldsymbol{\lambda}$ iteratively as shown in Algorithm 1.

Algorithm 1 *Subgradient method for solving* (4.9).

1. Initialize $\boldsymbol{\lambda^{(0)}}$.
2. Given $\boldsymbol{\lambda^{(k)}}$, solve the application layer and physical layer subproblems (4.12) and (4.13), respectively. Obtain the optimal values $\boldsymbol{t^*}$, $\boldsymbol{x^*}$, $\boldsymbol{y^*}$ and $\boldsymbol{z^*}$, and thus $\boldsymbol{d^*}$.
3. Perform a subgradient update for $\boldsymbol{\lambda}$, where $\boldsymbol{\nu^{(k)}}$ follows a diminishing step size rule:
$$\boldsymbol{\lambda_r^{(k+1)}} = \left[\boldsymbol{\lambda^{(k)}} + \left(\boldsymbol{\nu^{(k)}}\right)^T \left(\boldsymbol{t^*} - \boldsymbol{d^*}\right)\right]^+$$
4. Return to step 2 until convergence.

Following a diminishing step size rule for choosing $\boldsymbol{\nu^{(k)}}$, the subgradient method above is guaranteed to converge to the optimal dual variables [30]. Careful readers may be concerned with the slow convergence of the subgradient updates, especially when the problem scales up. Computational experiences suggest that the complexity of subgradient updates is polynomial in the dimension of the dual problem, which is $|\mathcal{L}|$ for $g(\boldsymbol{\lambda})$ [26].

Solving the application layer subproblem is straightforward. It can be readily seen that the objective of (4.12) is maximized by maximizing each term in the summation separately. $U_l(t_l) - \lambda_l t_l$ is concave since by assumption U_l is a concave function of t_l. Thus, the optimal throughput demand in the application layer t_l^* can be found by simply taking the derivative of $U_l(t_l) - \lambda_l t_l$ with respect to t_l and setting it to zero. The BS searches for t_l^* for each link l following this procedure.

Many different choices of utility functions are possible. Since we assume infinite backlog of data, we use the following utility function definition for every link throughout the rest of the paper:

$$U_l(t_l) = \ln t_l \quad (4.15)$$

This corresponds to the well-known proportional fairness utility model [27]. Its merits include the ability to strike a good balance between throughput and fairness, and robustness with respect to changes in topology and power constraints [28]. We let each link shares the same utility function definition here for ease of illustration. The optimization however does not depend on this assumption to work. With this simple utility function, t_l^* can be readily found as follows:

$$t_l^* = \frac{1}{\lambda_l}. \quad (4.16)$$

5 Solution Algorithms for the Physical Layer Resource Allocation Problem

We demonstrated that our network utility maximization problems can be solved in their dual domain by solving the application layer and physical layer subproblems separately. We also showed that the application layer subproblem is easy to solve. In this section, we tackle the more difficult physical layer subproblem.

The physical layer resource allocation problem is essentially an integer program. Conventional approaches, such as branch and bound [31], are computationally expensive. Our solution algorithms need to be run frequently at each scheduling epoch, making the task of deriving efficient heuristic algorithms imperative. Here we design efficient algorithms that solve the resource allocation for both the RSS-XOR and NO-XOR problems. Specifically, we first prove that RSS-XOR resource allocation is NP-complete and can be solved in polynomial-time with an approximation ratio of 1.5 using our algorithm. We then show that NO-XOR resource allocation can be optimally solved by transforming to weighted bipartite matching. Finally we design a subgradient algorithm to solve power allocation of the two problems in the dual domain.

5.1 A Set Packing Algorithm for RSS-XOR Resource Allocation

Solving the seemingly prohibitive RSS-XOR resource allocation problem (4.13) hinges on transforming to a weighted set packing problem. We first establish the equivalence and prove the hardness of the problem. We then propose our algorithm with a constant approximation factor.

Proposition 2: The RSS-XOR resource allocation problem is equivalent to a maximum weighted 3-set packing problem, and is NP-complete.

Proof: Construct a collection of channel sets C from a base set $\zeta \cup \psi$ as shown in Fig. 3. There are three kinds of channel sets, representing three transmission modes respectively. (c_i), where $c_i \in \zeta$ represents all the available channel sets for the direct transmission mode. (c_i, c_r) where $c_i \in \zeta, c_r \in \psi$ corresponds to all the available data-relay channel combinations for the conventional CD mode, with data subchannel c_i and relay subchannel c_r. The third kind, (c_i, c_j, c_r) corresponds to all the channel sets for the XOR-CD mode with data subchannel pair (c_i, c_j) and relay subchannel c_r, where $c_i, c_j \in \zeta, c_r \in \psi$. Sets intersect if they share at least one common element, and are otherwise said to be disjoint.

Each set has a corresponding weight, denoting the maximum objective value found across all possible assignments of this channel set to different combinations of RS and links. Specifically,

$$w_{(c_i)} = \max_{l} \; \lambda_l R(c_i, l), \tag{5.1}$$

$$w_{(c_i,c_r)} = \max_{l,r} \; \lambda_l R(c_i, c_r, r, P, l). \tag{5.2}$$

For set (c_i, c_j, c_r), its weight is found over all possible assignments of this set to combinations of RS and uplink-downlink of a MS, since it can only be assigned to one MS. Formally,

$$w_{(c_i,c_j,c_r)} = \max_{s,r} \sum_{l:s=M(l)} \lambda_l R(c_i, c_j, c_r, r, P, s). \tag{5.3}$$

$w_{(c_i,c_j,c_r)}$ essentially sums up uplink and downlink rates of s since one XOR-CD session incorporates two cooperative transmissions.

The optimization (4.13) is to find the optimal strategy to choose the transmission mode and assign RS and channels to each link in order to maximize the aggregated utility. The maximization is done over all links. Equivalently, we can interpret it as to find the optimal strategy to select disjoint channel combinations and assign RS and links to them so as to maximize the objective. This is simply a change of the order of summation in the objective of (4.13), when we substitute (4.2) into it. In this alternative interpretation, the maximization is done over all possible channel sets by matching them to the best possible links and RS without duplicate use of channels. The solution found must exhaust all subchannels since we can always improve the total weight by adding sets corresponding to unassigned data and relay subchannels. The number of elements in a set is at most 3, therefore the problem is equivalent to weighted 3-set packing [32], which is NP-complete. □

All the sets and their corresponding set weight are recorded in a table T_{assign}. We see that, for sets (c_i), the size of weight search space is $|\mathcal{L}|$; for sets (c_i, c_r) and (c_i, c_j, c_r), the search space size is $|\Phi||\mathcal{L}|$. Thus, the weight construction process is of polynomial time complexity, given the number of three kinds of sets are also polynomials of $|\zeta|$ and $|\psi|$.

To propose a good approximation algorithm with reasonable time complexity, first we construct an intersection graph G_C of the set system C with the set of vertices V_C and the set of undirected edges E_C as shown in Fig. 3. Weighted set packing then can be generalized as a weighted independent set problem, the objective of which is to find a maximum weighted subset of mutually non-adjacent vertices in G_C [33]. The size of sets is at most 3, therefore G_C is 3-claw free[1]. The best known approximation for the weighted independent set problem in a claw-free graph is proposed in [33] and then acknowledged in [32], which we extend to form our algorithm.

First we introduce a greedy algorithm, called *Greedy* that prepares the groundwork. Define $N(K, L)$ to be the set of vertices in L that intersect with vertices in K, *i.e.* $N(K, L) = \{u \in L : \exists v \in K$ such that $\{u, v\} \in E$ or $u = v\}$. *Greedy* is a natural heuristic that repeatedly picks the heaviest vertex from among the remaining vertices and eliminate it and the adjacent vertices as shown in Algorithm 2.

Algorithm 2 *Greedy.*

1. $S \leftarrow \varnothing$
2. **while** $V_C - N(S, V_C) \neq \varnothing$ **do**
3. choose $u \in V_C - N(S, V_C)$ with the maximum weight
4. $S \leftarrow S \cup \{u\}$
5. **end while**

Algorithm 3 *Approximation algorithm for solving* (4.13).

1. Construct the collection of weighted sets C and transform them into the weighted undirected graph G_C.
2. Obtain a maximal independent set S using *Greedy*.
3. **while** $\exists$ claw c such that T_c improves $w^2(S)$ **do**
4. $S \leftarrow S - N(T_c, S) \cup T_c$
5. **end while**
6. Assign channels to RS and links by searching the entries in T_{assign} corresponding to the sets present in S.

A natural thought to improve on the maximal independent set found by *Greedy* is to do local search and replace a set with its claw with larger weight, which motivates our approximation algorithm summarized

1. Here a d-claw c is an induced subgraph that consists of an independent set T_c of d nodes.

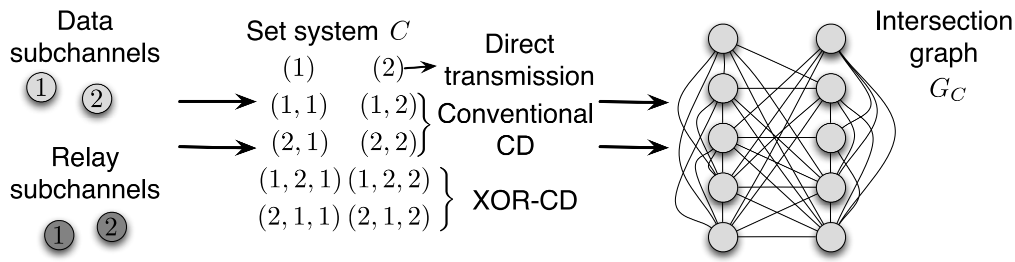


Fig. 3. Channel set construction and transformation into an intersection graph with 2 data subchannels and 2 relay subchannels. Vertices in G_C correspond to sets in C. Edges are added between vertices whose corresponding sets intersect.

as in Algorithm 3. From [33], local improvements on the square of total weights solve the weighted independent set problem with a constant approximation factor of 1.5, which is the best result known so far [32]. Therefore we have the following:

Proposition 3: Algorithm 3 provides at least $\frac{2}{3}$ of the optimum of RSS-XOR physical layer resource allocation problem (4.13). This is the *best* performance guarantee one can have unless a better algorithm can be found for the weighted independent set problem.

5.2 A Matching Algorithm for NO-XOR Resource Allocation

We now consider the NO-XOR physical layer subproblem. Using the same technique presented in Sec. 4.5, the problem can be shown in the following form:

$$\begin{aligned}
\max_{\boldsymbol{x},\boldsymbol{y}} \quad & \sum_{l\in\mathcal{L}} \lambda_l d_l \\
\text{s.t.} \quad & d_l = \sum_{c_i\in\zeta} R(c_i,l)x_l^{c_i} \\
& \quad + \sum_{c_i\in\zeta}\sum_{c_r\in\psi}\sum_{r\in\Phi} R\big(c_i,c_r,r,P,l\big) y_l^{r,c_i,c_r}, \\
& \sum_{l\in\mathcal{L}} x_l^{c_i} + \sum_{l\in\mathcal{L}}\sum_{r\in\Phi}\sum_{c_r\in\psi} y_l^{r,c_i,c_r} \le 1, \forall c_i\in\zeta, \\
& \sum_{l\in\mathcal{L}}\sum_{r\in\Phi}\sum_{c_i\in\zeta} y_l^{r,c_i,c_r} \le 1, \forall c_r\in\psi.
\end{aligned} \tag{5.4}$$

Surprisingly, we find that it can be optimally solved in polynomial time. Specifically,

Proposition 4: The NO-XOR physical layer resource allocation problem is equivalent to weighted bipartite matching over all data and relay subchannels, and thus can be solved optimally.

Proof: Construct a bipartite graph $A = (V_1 \times V_2, E)$ where V_1 and V_2 correspond to the set of data subchannels ζ and relay subchannels ψ respectively, as shown in Fig. 4. We patch void vertices to V_2 to make $|V_2| = |V_1| = |\zeta|$. The edge set E corresponds to $|\zeta|^2$ edges connecting all possible pairs of channels in two vertex sets. Each edge (k,j) carries three attributes, (w_{kj}, l_{kj}, r_{kj}), where

$$\begin{aligned}
w_{kj} &= \max_{l,r} \ \lambda_l R(k,j,r,P,l), \\
(l_{kj}, r_{kj}) &= \arg\max_{l,r} \ \lambda_l R(k,j,r,P,l).
\end{aligned} \tag{5.5}$$

For edges connecting data subchannels to void relay subchannels that we patched, the edge weights become $(w_{kj}, l_{kj}, 0)$ where l_{kj} is the link providing maximum objective value if data subchannel k is used. This essentially captures the maximum objective value given by direct transmission.

Observe that A is bipartite, we can see the NO-XOR resource allocation problem (5.4) is equivalent to finding the maximum weighted bipartite matching on A. The second attribute of an edge (k,j) in the maximum matching represents the link assigned with this data-relay subchannel pair (k,j), while the third attribute dictates the transmission mode or the corresponding RS. A 0 in the third attribute simply means the link should work in direct transmission mode. Hence, the maximum matching found represents the comprehensive assignment of RS, data and relay subchannels, as well as the transmission strategy decision, therefore optimally solves the problem. □

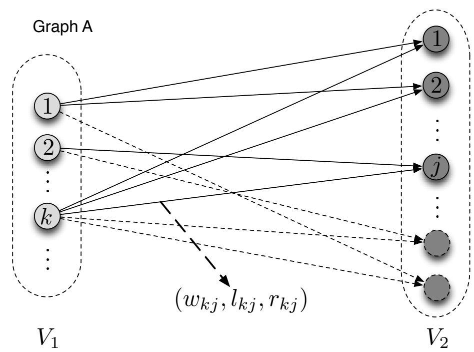


Fig. 4. The graphical model to show the equivalence of NO-XOR resource allocation problem (5.4) and weighted bipartite matching. Dotted vertices are void vertices patched. Not all links are shown here.

Several good polynomial-time algorithms exist for solving the bipartite matching problem, of which the

Hungarian algorithm [34] is a popular choice. Since the graph construction is $O(|\zeta|^2 \cdot |\mathcal{L}| \cdot |\psi|)$, the entire algorithm is polynomial time.

5.3 A Power Allocation Algorithm

Finally, we turn our focus to the power allocation problem. Recall that the power limited versions of RSS-XOR and NO-XOR are proposed in Sec. 4.4. Consider the power limited RSS-XOR problem. Readily we can see that it can be decoupled into the same application layer subproblem as (4.12), and a physical layer resource allocation of the following form, which includes now an additional power constraint (4.7):

$$\max_{\boldsymbol{x},\boldsymbol{y},\boldsymbol{z}} \sum_{l\in\mathcal{L}} \lambda_l d_l \quad \text{s.t. } (4.2), (4.3), (4.4), \text{ and } (4.7) \tag{5.6}$$

Therefore Algorithm 1 can be used to solve it, provided that we can solve the physical layer subproblem (5.6). In this section, we develop a dual method to solve (5.6) efficiently.

Introduce an Lagrangian multiplier vector $\boldsymbol{\mu}$ to the power constraint (4.7) and the dual problem becomes

$$\min_{\boldsymbol{\mu}\geq 0} \ g(\boldsymbol{\mu}) \tag{5.7}$$

where

$$g(\boldsymbol{\mu}) = \max \sum_{l\in\mathcal{L}} \lambda_l d_l + \sum_{r\in\psi} \mu_r \Big(P_r - \sum_{l,c,c_r} p_l^{r,c,c_r} - \sum_{s,c_i,c_j,c_r} p_s^{r,c_i,c_j,c_r} \Big) \quad \text{s.t. } (4.2)\text{–}(4.4). \tag{5.8}$$

Since each MS s corresponds to two links, we can equally split power used for XOR-CD p_s^{r,c_i,c_j,c_r} to these two links without violating the power constraint. Mathematically, we let

$$p_l^{r,c_i,c_j,c_r} = \begin{cases} \frac{1}{2} p_s^{r,c_i,c_j,c_r} & \text{if } s = M(l); \\ 0 & \text{otherwise.} \end{cases} \tag{5.9}$$

The objective function of (5.8) can then be written as

$$\max \sum_{l} \Big(\lambda_l d_l - \sum_{r,c,c_r} \mu_r p_l^{r,c,c_r} - \sum_{r,c_i,c_j,c_r} \mu_r p_l^{r,c_i,c_j,c_r} \Big) + \sum_r \mu_r P_r. \tag{5.10}$$

$\sum_r \mu_r P_r$ is constant in (5.10) since $\boldsymbol{\mu}$ is given for each instance of $g(\boldsymbol{\mu})$. So solving the optimization (5.8) is equivalent to solving the following:

$$\max \sum_{l} \Big(\lambda_l d_l - \sum_{r,c,c_r} \mu_r p_l^{r,c,c_r} - \sum_{r,c_i,c_j,c_r} \mu_r p_l^{r,c_i,c_j,c_r} \Big) \quad \text{s.t. } (4.2)\text{–}(4.4). \tag{5.11}$$

Compared with the original RSS-XOR physical layer subproblem (4.13), the only difference is the objective function which now includes the cost of power. $\boldsymbol{\mu}$ can be interpreted as a pricing variable vector for relay power. (5.11) can be thought of as maximizing the total throughput revenue minus the total cost of relay power used, given the current prices of power at RS. This is easily decomposed into maximization over every possible set of data and/or relay channels. Therefore, it can be solved using the Algorithm 3 in Sec. 5.1, with the weights being the maximum throughput revenue discounted by power cost instead of the maximum revenue only. For ease of exploration, we dictate that the relay power for each cooperative session is set to the threshold value as derived in Sec. 3.4.2 and 3.4.3. Then,

$$w_{(c_i,c_r)} = \max_{l,r} \ \lambda_l R(c_i, c_r, r, p_l^{r,c_i,c_r}, l) - \mu_r p_l^{r,c_i,c_r},$$

$$w_{(c_i,c_j,c_r)} = \max_{s,r} \sum_{l:s=M(l)} \lambda_l R(c_i, c_j, c_r, r, p_s^{r,c_i,c_j,c_r}, s) - \mu_r p_s^{r,c_i,c_j,c_r}.$$

After solving (5.11), the dual problem (5.7) can be readily solved via a subgradient method which repeatedly updates the power prices according to the demand and supply relationship at RS to regulate the power consumption. To summarize, the algorithm for solving the power limited RSS-XOR problem works as shown in Algorithm 4. Notice that the subgradient algorithm is suitable for distributed implementation across RS. Each RS is able to verify its power consumption, and update its own relay power price autonomously according to $\boldsymbol{\nu}^{(k)}$ informed by BS. The updated prices can be transmitted to the BS with a negligible amount of overhead. The dual method for power limited NO-XOR problem can be developed in a similar way.

Algorithm 4 *Algorithm for solving* (5.6).

---

1. Initialize $\boldsymbol{\mu}^{(0)}$.
2. Given $\boldsymbol{\mu}^{(k)}$, solve the maximization problem (5.11) using Algorithm 3. Obtain the solution values $\hat{p}_l^{r,c,c_r}$ and $\hat{p}_l^{r,c_i,c_j,c_r}$, and the maximal independent set $\hat{S}$.
3. Perform a subgradient update for $\boldsymbol{\mu}$, where $\boldsymbol{\nu}^{(k)}$ follows a diminishing step size rule:
$$\mu_r^{(k+1)} = \Big[\mu_r^{(k)} - \nu_r^{(k)} \Big(P_r - \sum_{l,c,c_r} \hat{p}_l^{r,c,c_r} - \sum_{l,c_i,c_j,c_r} \hat{p}_l^{r,c_i,c_j,c_r} \Big) \Big]^+$$
4. Return to step 2 until convergence.
5. Assign channels to RS and links by searching the entries in T_{assign} corresponding to the sets present in $\hat{S}$ found in step 4.

---

6 Performance Evaluation

We dedicate this section to evaluating the performance of our algorithms using simulations. Recall that we use the proportional fairness utility function $U_l(d_l) = \ln d_l$ for each link l as mentioned in Sec. 4.1.

6.1 Simulation Setup

We first introduce the simulation setup. The key of our experiment settings is to derive the achievable data rate of a subchannel when it is allocated to a particular MS, which requires computing the SNR value. We rely on the wireless channel simulator called Chsim [35] to generate realistic SNR results under different channel and mobility models, and adopt empirical parameters explained below to model the wireless fading environment.

The subchannel bandwidth is set to be 312.5 kHz. Data subchannels are centered at 5 GHz, while relay subchannels are centered at 2.5 GHz. The channel gain between two nodes at each subchannel can be decomposed into a small-scale Rayleigh fading component and a large-scale log normal shadowing with standard deviation of 5.8 and path loss exponent of 4. The inherent frequency selective property is characterized by an exponential power delay profile with delay spread 15 μs. The time selective nature is captured by Doppler spread, which depends on the MS's speed. Throughout the simulation the MS are moving with speed uniformly distributed from 1 to 5 m/s according to random waypoint model with zero-second pause period.

Without loss of generality, the gap to capacity Γ is set to 1, meaning that for a given instantaneous channel gain, the physical layer codewords adaptively operate at the instantaneous achievable rate of the relay protocol. This implies that an ideal adaptive modulation and coding scheme (AMC) is implemented. Γ can also be set to a value larger than 1 to reflect the gap between physical layer implementation and the theoretical result. The power constraint for each transmission is such that $\frac{P}{N_0 W}$ =23 dB. This corresponds to a medium SNR environment. Such an experimental setup is commonly used in related studies [25], [26].

6.2 Performance of XOR-CD

We first evaluate the performance of XOR-CD using Algorithm 1 with Algorithm 3. We compare it to conventional CD using Algorithm 1 with the Hungarian algorithm [34]. We focus on the scenario where 10 MS are randomly located in a cell with a 100-meter radius. We set the number of data subchannels to be 100, and that of relay subchannels to be 30. For fairness 130 subchannels are used for the simulation of direct transmission. 1 RS is deployed in the cell.

Fig. 5 plots the network throughput for one second, with a sampling period of 5 ms. The optimization therefore is done for 200 times. We can clearly see that XOR-CD outperforms conventional DF cooperative diversity by around 30%. This is as expected, because XOR-CD conserves relay channels that can be utilized to support more cooperative sessions. To further illustrate its superiority in this aspect, we study XOR-CD's relaying resource efficiency. We evaluate *power efficiency* defined as the ratio of throughput improvement over direct transmission and the amount of relay power used. Since we assume a fixed power for each relay subchannel without power allocation, this amounts to

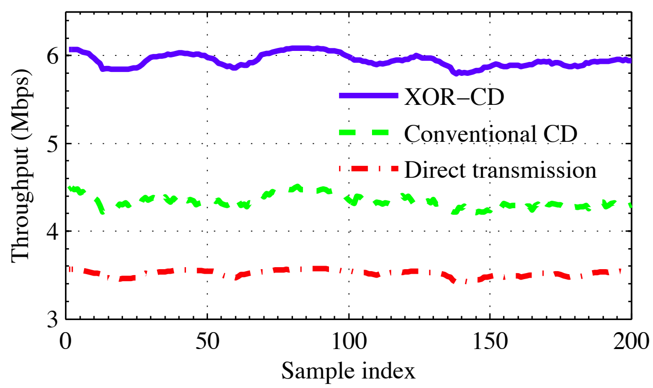


Fig. 5. Throughput comparison of XOR-CD against conventional CD and direct transmission.

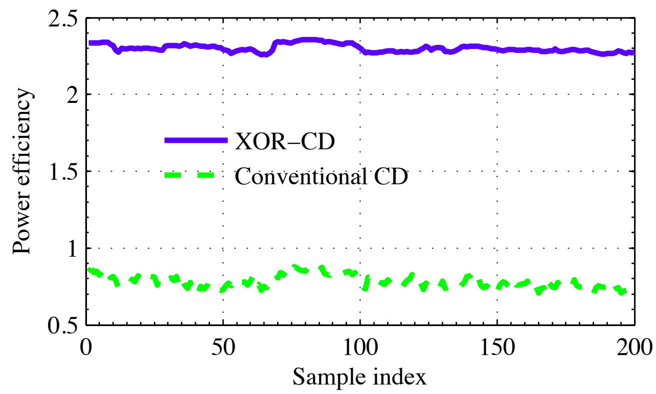


Fig. 6. Power efficiency comparison of XOR-CD against conventional CD.

$$\text{power efficiency} = \frac{\text{throughput improvement (\%)}}{\text{number of relay subchannels}}.$$

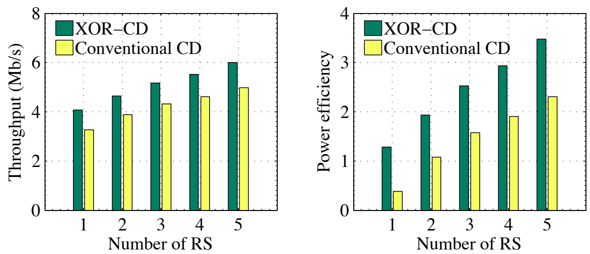
As seen in Fig. 6, XOR-CD's power efficiency is *significantly* better than that of conventional CD.

Finally, we notice that the conventional diversity scheme alone provides over 20% improvement compared with simple direct transmission. This diversity gain is similar to the network coding gain, which further confirms the advantage of XOR-CD to "double" the diversity gain without significant costs.

6.3 Effects of Relaying Resources

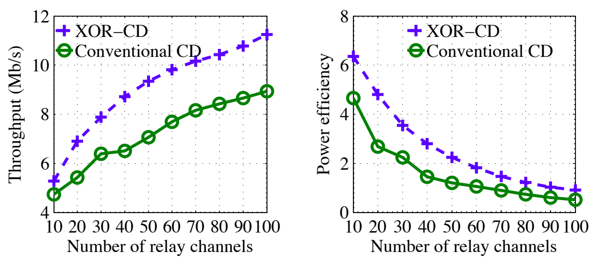
Next, we study the effects of relaying resources on the performance of XOR-CD. We focus on two kinds of resources: RS and relay subchannels. Fig. 7 and Fig. 8 show the results, respectively. Intuitively, more RS provide better chances for MS to find a nearby

RS with better relay channel qualities. More relay subchannels also enable more cooperative sessions to take place. Clearly, these two factors contribute to the increasing throughput reflected in Fig. 7(a) and Fig. 8(a). XOR-CD consistently maintains a 20%–30% gain over conventional CD.



(a) Throughput comparison. (b) Power efficiency comparison.

Fig. 7. Effects of the number of RS. Number of relay subchannels is 20, number of data channels is 100, number of MS is 10, and cell radius is set to be 100m.



(a) Throughput comparison. (b) Power efficiency comparison.

Fig. 8. Effects of the number of relay channels. Number of RS is 3, number of data channels is 100, number of MS is 10, and cell radius is set to be 100m.

Interestingly, when it comes to the power efficiency as defined in Sec. 6.2, we observe a clear discrepancy between the two kinds of resources. When we increase the number of RS, the total amount of relay power used is unchanged since the total number of relay subchannels, and thus the total number of cooperative sessions, is unchanged. Thus, the power efficiency is improved as shown in Fig. 7(b). However, when we increase the number of relay subchannels, the total amount of relay power used is proportionally increased, and efficiency is actually decreasing in Fig. 8(b). This is because the optimization always tries to exploit the largest performance gains first, which leads to the diminishing marginal gain of using more relay subchannels and thus relay power. This observation suggests that we could use a small amount of relaying resources to obtain a reasonably satisfactory improvement.

6.4 Effects of Path Loss

Intuitively, path loss increases when we increase the cell radius, and relaying becomes more beneficial for improving the throughput. This intuition is confirmed in Fig. 9, which plots the total throughput improvement of XOR-CD over conventional CD with different values of cell radius. We observe that as cell radius increases from 100m to 300m, the throughput improvement also increases from around 40% to around 70%. We have conducted simulations with conventional CD and observed the same trend. This observation suggests that relaying in general is more helpful for networks that are limited by path loss.

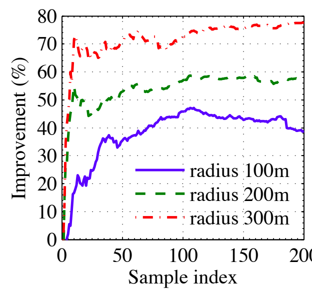


Fig. 9. Effects of path loss for 100 data subchannels, 30 relay subchannels, 10 MS and 3 RS.

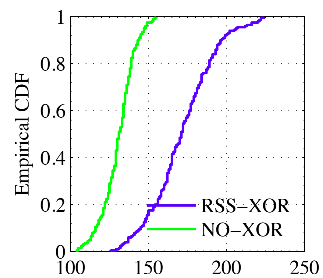


Fig. 10. Empirical CDF of convergence iterations with 100 data subchannels and 30 relay subchannels.

6.5 Convergence and Duality Gap

It is necessary to examine the convergence speed of our algorithms based on the subgradient methods. Fig. 10 plots the empirical CDF of convergence iterations of Algorithm 1 for both RSS-XOR and NO-XOR problems. We observe that it takes on average around 180 iterations to solve the RSS-XOR problem, and around 130 iterations to solve the NO-XOR problem. No more than 220 and 157 iterations are needed for solving the RSS-XOR and NO-XOR problems, respectively. We believe that such a computation burden is affordable for typical base stations with ever-increasing computational power.

We also examine the duality gap of our problem as a result of the finite number of OFDM subchannels. Since the RSS-XOR problem (4.5) is an NP-complete integer program with a non-linear objective function, we estimate its primal optimum by using the CVX solver [36] to solve (4.5). We plot the performance gap between CVX and our subgradient method with Algorithm 3 that provides the dual optimum. The number of data subchannels is fixed at 100, and that of relay subchannels is set to 30, 60, and 90. Table. 3 shows the estimated duality gap. We observe the gap is indeed small and decreasing as the number of channels increases. This demonstrates the validity of solving the optimization in the dual domain. Note that CVX is significantly slower than our algorithms.

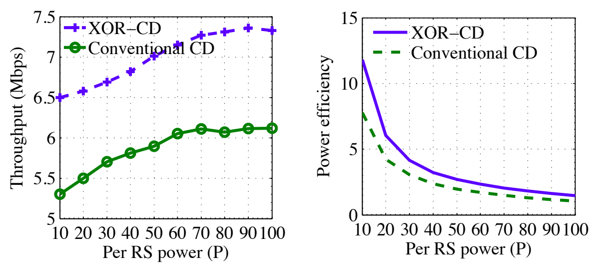
6.6 Performance of Power Allocation

Finally, we evaluate the performance of our power allocation algorithms. We implement the subgradient

TABLE 3
Estimated duality gap.

# of relay subchannels	30	60	90
Performance gap	5.46%	4.80%	4.24%

based Algorithm 4 for the RSS-XOR problem, as well as its counterpart for the NO-XOR problem. We enforce a uniform power constraint across RS. Each RS has relay power NP, where P is the power used for one direct transmission as in Sec. 3.4.1. We vary N from 10 to 100 and obtain Fig. 11. We observe from Fig. 11(a) that XOR-CD is *less* sensitive to power constraints as reflected by the marginal improvement compared to conventional CD. This is because XOR-CD utilizes power more efficiently, resulting in a lower power demand at RS. Therefore, pumping more power does not improve its performance substantially.



(a) Throughput comparison. (b) Power efficiency comparison.

Fig. 11. Effects of power constraints for a 100m cell with 100 data subchannels, 30 relay subchannels, 10 MS, 3 RS.

We also evaluate the effect of power allocation on relay power efficiency, which now is calculated by dividing the throughput improvement (in %) by N, the per RS relay power. The result is shown in Fig. 11(b). We observe that, expectedly, relay power efficiency is decreasing due to the intelligent power allocation that always maximizes the throughput gain under a fixed power budget. XOR-CD consistently provides better power efficiency up to the point that relay power is abundant for the conventional CD to achieve the same throughput gain (the two lines in Fig. 11(b) converge as N increases).

TABLE 4
Throughput values of different relay power profiles.

Relay power profile: RS1	RS2	RS3	Throughput (Mbps)	Improvement (%)
$10P$	$10P$	$10P$	6.69	—
$15P$	$9P$	$6P$	6.93	3.58
$15P$	$6P$	$9P$	6.91	3.29
$18P$	$7P$	$5P$	7.14	6.73
$18P$	$5P$	$7P$	7.11	6.28

The convergence speed of our power control algorithms is much slower since they involve a nested subgradient update loop. In our simulations, we find that Algorithm 4 with Algorithm 1 for the power control version of the RSS-XOR problem takes on average around 1500 iterations to solve, which may not be feasible for practical use. Given that power allocation yields marginal improvement for both XOR-CD and conventional CD as discussed above, in most cases it is sufficient to adopt the simple uniform power allocation across mobile stations.

We finally study non-uniform power constraints at RS. For the same configuration and network topology with 3 RS, we fix the total power constraint to be $30P$ and vary individual RS's power constraint. Table 4 summarizes the results of different relay power profiles. Observe that allocating more power to RS1 has a positive effect on the average throughput, while adjusting the constraints of RS2 and RS3 does not. The reason is that in our simulation RS are randomly located inside the cell. RS1 is located closest to the BS, providing a much better relay channel for cooperative transmissions. Allocating more power to RS1 boosts its relaying capacity and improves throughput. The result also suggests that power allocation at RS needs to be location-adaptive to best utilize resources.

It is possible to consider the optimization of relay power budget in our framework. We only need to change the relay power budget P_r from a constant to a new variable, and add a new total power constraint across all the P_r: $\sum_r P_r = P_R$. Solving the relay power budget optimization is technically more involved. Subgradient methods can still be used by relaxing the additional constraint, but the convergence of the algorithm will be negatively affected. Since the performance gain of relay power budget allocation is marginal as shown in Table 3, we do not to discuss this issue in detail here.

7 Conclusion

This work represents an early attempt to study network coding assisted cooperative diversity in multi-channel cellular networks. We presented XOR-CD, a simple cooperative diversity scheme with XOR in OFDMA networks. As our main contribution, we proposed a unifying optimization framework based on network utility maximization to exploit multi-user diversity, cooperative diversity and network coding jointly. We established the hardness of the decoupled resource allocation problem in the physical layer, and proposed efficient approximation and optimal algorithms. Simulation results demonstrated that network coding has the potential to significantly improve throughput of OFDMA networks.

References

[1] H. Xu and B. Li, "XOR-assisted cooperative diversity in OFDMA wireless networks: Optimization framework and approximation algorithms," in *Proc. IEEE INFOCOM*, 2009.

[2] S. Chachulski, M. Jennings, S. Katti, and D. Katabi, "Trading structure for randomness in wireless opportunistic routing," in *Proc. of ACM SIGCOMM*, 2007.

[3] Y. Wu, P. A. Chou, and S.-Y. Kung, "Information exchange in wireless networks with network coding and physical-layer broadcast," in *Proc. of CISS*, 2005.
[4] A. Sendonaris, E. Erkip, and B. Aazhang, "User cooperation diversity-Part I: System description," *IEEE Trans. Commun.*, vol. 51, no. 11, pp. 1927–1938, November 2003.
[5] ——, "User cooperation diversity-Part II: Implementation aspects and performance analysis," *IEEE Trans. Commun.*, vol. 51, no. 11, pp. 1939–1948, November 2003.
[6] J. N. Laneman, D. N. C. Tse, and G. W. Wornell, "Cooperative diversity in wireless networks: Efficient protocols and outage behavior," *IEEE Trans. Inf. Theory*, vol. 50, no. 12, pp. 3062–3080, December 2004.
[7] J. N. Laneman, "Network coding gain of cooperative diversity," in *Proc. of IEEE MILCOM*, 2004.
[8] Y. Chen, S. Kishore, and J. T. Li, "Wireless diversity through network coding," in *Proc. of IEEE WCNC*, 2006.
[9] C. Peng, Q. Zhang, M. Zhao, and Y. Yao, "On the performance analysis of network-coded cooperation in wireless networks," in *Proc. of IEEE INFOCOM*, 2007.
[10] IEEE Standard, "802.16TM: Air interface for fixed wireless access systems," 2004.
[11] T. M. Cover and A. A. E. Gamal, "Capacity theorems for the relay channel," *IEEE Trans. Inf. Theory*, vol. 25, no. 5, pp. 572–584, September 1979.
[12] D. Gunduz and E. Erkip, "Opportunistic cooperation by dynamic resource allocation," *IEEE Trans. Wireless Commun.*, vol. 6, no. 4, pp. 1446–1454, April 2007.
[13] S. Savazzi and U. Spagnolini, "Energy aware power allocation strategies for multihop cooperative transmission schemes," *IEEE J. Sel. Areas Commun.*, vol. 25, no. 2, pp. 318–327, February 2007.
[14] A. Khandani, J. Abounadi, E. Modiano, and L. Zheng, "Cooperative routing in static wireless networks," *IEEE Trans. Commun.*, vol. 55, no. 11, pp. 2185–2192, November 2007.
[15] R. Ahlswede, N. Cai, S.-Y. Li, and R. Yeung, "Network information flow," *IEEE Trans. Inf. Theory*, vol. 46, no. 4, pp. 1204–1216, July 2000.
[16] S. Katti, H. Rahul, W. Hu, D. Katabi, M. Medard, and J. Crowcroft, "XORs in the air: Practical wireless network coding," in *Proc. of ACM SIGCOMM*, July 2006.
[17] P. Chaporkar and A. Proutiere, "Adaptive network coding and scheduling for maximizing throughput in wireless networks," in *Proc. ACM MobiCom*, 2007.
[18] A. Khreishah, C. chun Wang, and N. Shroff, "Cross-layer optimization for wireless multihop networks with pairwise intersession network coding," *IEEE J. Sel. Areas Commun.*, vol. 27, no. 5, pp. 606–621, June 2009.
[19] N. Jones, B. Shrader, and E. Modiano, "Optimal routing and scheduling for a simple network coding scheme," in *Proc. IEEE INFOCOM*, 2012.
[20] X. Bao and J. Li, "Adaptive network coded cooperation (ANCC) for wireless relay networks: Matching code-on-graph with network-on-graph." *IEEE Trans. Wireless Commun.*, vol. 7, no. 2, pp. 574–583, February 2008.
[21] C. Hausl, F. Schreckenbach, I. Oikonomidis, and G. Bauch, "Iterative channel and network decoding on a tanner graph," in *Proc. 43rd Allerton Conference on Communications, Control, and Computing*, 2005.
[22] S. Yang and R. Koetter, "Network coding over a noisy relay: A belief propagation approach," in *Proc. IEEE Intl. Symposium on Information Theory (ISIT)*, 2007.
[23] J. Tang and X. Zhang, "Cross-layer resource allocation over wireless relay networks for quality of service provisioning," *IEEE J. Sel. Areas Commun.*, vol. 25, no. 4, pp. 645–656, May 2007.
[24] L. Le and E. Hossain, "Cross-layer optimization frameworks for multihop wireless networks using cooperative diversity," *IEEE Trans. Wireless Commun.*, vol. 7, no. 7, pp. 2592–2602, July 2008.
[25] S.-J. Kim, X. Wang, and M. Madihian, "Optimal resource allocation in multi-hop ofdma wireless networks with cooperative relay," *IEEE Trans. Wireless Commun.*, vol. 7, no. 5, pp. 1833–1838, May 2008.
[26] T. C.-Y. Ng and W. Yu, "Joint optimization of relay strategies and resource allocations in cooperative cellular networks," *IEEE J. Sel. Areas Commun.*, vol. 25, no. 2, pp. 328–339, February 2007.
[27] F. P. Kelly, A. Maulloo, and D. Tan, "Rate control for communication networks: Shadow prices, proportional fairness and stability," *Journal of Operations Research Society*, vol. 49, no. 3, pp. 237–252, 1998.
[28] G. Song and Y. G. Li, "Cross-layer optimization for ofdm wireless network - Part I and part II," *IEEE Trans. Wireless Commun.*, vol. 4, no. 2, March 2005.
[29] W. Yu and R. Lui, "Dual methods for nonconvex spectrum optimization of multicarrier systems," *IEEE Trans. Commun.*, vol. 54, no. 7, pp. 1310–1322, July 2006.
[30] S. Boyd and A. Mutapcic, "Subgradient methods," Lecture notes of EE364b, Stanford University, Winter Quarter 2006-2007. http://www.stanford.edu/class/ee364b/notes/subgrad_method_notes.pdf.
[31] G. L. Nemhauser and L. A. Wolsey, *Integer and Combinatorial Optimization*. Wiley-Interscience, 1988.
[32] B. Chandra and M. M. Halldórsson, "Greedy local improvement and weighted set packing approximation," *J. Algorithms*, vol. 39, no. 2, pp. 223–240, May 2001.
[33] P. Berman, "A d/2 approximation for maximum weight independent set in d-claw free graphs," *Nordic J. Comput.*, vol. 7, no. 3, pp. 178–184, 2000.
[34] C. Papadimitriou and K. Steiglitz, *Combinatorial Optimization: Algorithms and Complexity*. Prentice Hall, 1982.
[35] S. Valentin, "Chsim – a wireless channel simulator for omnet++," http://www.cs.uni-paderborn.de/en/fachgebiete/research-group-computer-networks/projects/chsim.html, 2006.
[36] "CVX: Matlab software for disciplined convex programming," http://cvxr.com/cvx/.

Hong Xu received his B.Engr. degree from the Department of Information Engineering, The Chinese University of Hong Kong, in 2007 and the M.A.Sc. and Ph.D. degrees from the Department of Electrical and Computer Engineering, University of Toronto in 2010 and 2013. He will join the Department of Computer Science, City University of Hong Kong as an Assistant Professor starting from August 2013. His research interests include cloud computing, datacenters, and computer networking and systems. He is a student member of ACM and IEEE.

Baochun Li received the B.Engr. degree from the Department of Computer Science and Technology, Tsinghua University, China, in 1995 and the M.S. and Ph.D. degrees from the Department of Computer Science, University of Illinois at Urbana-Champaign, Urbana, in 1997 and 2000.

Since 2000, he has been with the Department of Electrical and Computer Engineering at the University of Toronto, where he is currently a Professor. He holds the Nortel Networks Junior Chair in Network Architecture and Services from October 2003 to June 2005, and the Bell Canada Endowed Chair in Computer Engineering since August 2005. His research interests include large-scale distributed systems, cloud computing, peer-to-peer networks, applications of network coding, and wireless networks.

Dr. Li was the recipient of the IEEE Communications Society Leonard G. Abraham Award in the Field of Communications Systems in 2000. In 2009, he was a recipient of the Multimedia Communications Best Paper Award from the IEEE Communications Society, and a recipient of the University of Toronto McLean Award. He is a member of ACM and a senior member of IEEE.