

# Jointly Optimal Bit Loading, Channel Pairing and Power Allocation for Multi-channel Relaying

Mahdi Hajiaghayi\*, Min Dong<sup>†</sup>, and Ben Liang\*

\*Dept. of Electrical and Computer Engineering, University of Toronto, Canada

<sup>†</sup>Faculty of Engineering and Applied Science, University of Ontario Institute of Technology, Canada

Email: {mahdih, liang}@comm.utoronto.ca, min.dong@uoit.ca

**Abstract**—We aim to enhance the end-to-end rate of a general dual-hop relay network with multiple channels and finite modulation formats, by jointly optimizing channel pairing, power allocation, and integer bit loading. Such an optimization problem has both a discrete feasible region, due to the combinatoric nature of channel pairing, and a discrete objective, due to the bit loading requirement. For this type of mixed-integer programming problems, the Lagrange dual method generally is inapplicable, due to the non-zero duality gap. However, by exploring the structure of our problem, we are able to bound the gap to within one bit, allowing the extraction of an exact optimal integer solution. We further present complexity reduction techniques, and demonstrate that the proposed solution only requires a computational complexity that is polynomial in the number of channels, realizing efficient implementation in practical systems. Through numerical experiments, we show that the jointly optimal solution can significantly outperform common sub-optimal alternatives.

## I. INTRODUCTION

Multicarrier transmission is an essential component of the physical-layer technology for next-generation wireless systems [1]. Correspondingly, future system designs are evolving toward the adoption of a multi-channel relaying architecture to allow broadband access and coverage improvement [2], [3]. As opposed to a narrow-band single-channel relay, a multi-channel relay has access to multiple channels, e.g., the subcarriers in an Orthogonal Frequency Division Multiplexing (OFDM) relaying system. It may receive a signal from one channel and transmit a processed version of the signal on a different channel. This multi-channel relaying capability can be exploited to adaptively route the incoming signals, taking advantage of the diverse strength of different channels.

The flexibility enabled by multi-channel relaying also brings the new challenge of power allocation among different channels within a given power budget. From an information theoretic point of view, the problem is to optimally allocate power at the source and the relay to maximize the end-to-end data rate. It has been investigated in the literature with various system setups and constraints, and various allocation schemes have been proposed [4]-[7]. However, they all lead to non-integer rate adaptations. Although they provide theoretical understanding for the performance of such systems, the solutions obtained are not suited for practical systems, where finite modulation formats are used for data transmission.

Finite modulation formats incur a finite set of integer bits per transmission, along with a possible requirement on the

error probability. With the practical constraint of integer-valued bit rates, the power allocation problem becomes even more challenging: how can the power and bits be allocated optimally to each channel, in order to maximize the system performance? In the literature of point-to-point communications, this problem is generally referred to as *bit loading* design. Relaying adds another dimension to the problem, since the incoming and outgoing channels at the relay need to be optimally paired for maximal performance. In particular, the bit allocation on an incoming channel of the relay is no longer the same as that of the same channel on the outgoing side.

We aim to derive a practical and computationally efficient resource allocation method that leads to optimal channel selection and power allocation for a multi-channel capable relaying system with integer bit loading. Three issues are addressed jointly: 1) channel pairing: the pairing of incoming and outgoing channels at the relay; 2) power allocation: the determination of power to transmit one of the allowable modulation formats on each channel; 3) bit loading: the selection of discrete bit values from a finite set of integers (corresponding to a modulation scheme) for the transmission rate. There is often strong correlation among the three issues. Intuitively, to maximize the source-destination rate performance, the choice of channel pairing at the relay depends on the specific power allocation scheme used, and vice versa. Together, they affect the selection of bits to be allocated to each channel.

In this work, we design a jointly optimal channel pairing, power allocation and bit loading scheme for multi-channel relaying, to maximize the overall bit rate under a single power constraint. The power constraint can be on the source, the relay, or as an upper limit to their total power. The resulting optimization problem is of non-convex mixed-integer programming in nature, with both a discrete objective and a discrete feasible region. Although the Lagrange duality approach can be applied to such problems [8], it generally does not lead to an optimal solution due to the non-zero duality gap. It is shown in [9] that, for many non-convex problems in broadband multi-channel resource allocation, a time-sharing property – a concavity condition of the objective function with respect to the total power budget – still ensures zero-duality gap, and thus the problem can be solved by the Lagrange dual. This result has been found useful to provide asymptotically optimal solution to some resource allocation problems in OFDM systems [10], [6]. In our problem, however, the discrete

objective function renders such condition invalid, and we show that a non-zero duality gap exists.

Nonetheless, we develop our solution through the Lagrange dual approach. By exploring the structure of our problem, we are able to bound the gap to the original objective to be within one bit. This knowledge allows us to extract the exact optimal integer solution. We further develop numerical techniques to reduce the algorithm runtime. Despite the general NP-hardness of many mixed-integer optimization problems, we demonstrate that the proposed algorithm has a computational complexity that is polynomial in the number of channels. This suggests that it is more amenable to practical implementation than combinatoric search approaches, especially for practical systems, such as IEEE 802.16 OFDMA, with hundreds to thousands of frequency sub-channels. Extensive numerical studies further demonstrate that the proposed optimal solution can substantially outperform common sub-optimal alternatives.

The rest of the paper is organized as follows. We first provide a literature review of the related work in Section II. In Section III, we describe the system model and formulate the joint optimization problem. In Section IV, we describe the Lagrange dual approach that produces an upper bound to our optimal objective, and in Section V, we present a study on the special structure of our problem that enables the extraction of the required integer optimal solution. We further present in Section VI a complexity reduction method and show that the proposed algorithm has polynomial runtime. Numerical results are given in Section VII, followed by conclusion in Section VIII.

## II. RELATED WORK

We explain the relation between our work and the related prior work in the following categories.

### A. Bit Loading for Point-to-Point Communication

For conventional point-to-point OFDM systems, various power allocation and bit loading schemes were proposed incorporating the discrete nature of modulation levels, with the performance metric being either total bit rate [11]-[16], or symbol error rate [17] [18]. In existing literature, [11] was the first to consider this problem and developed a greedy algorithm that is optimal in maximizing throughput. Subsequently, optimal algorithms with reduced complexity were presented in [14] for QAM and in [15] and [16] for general modulation schemes. To further alleviate the computational burden, some less complex but suboptimal solutions were proposed in [12], [13] and [18]. The authors of [12] and [13] took advantage of the water-filling solution obtained from continuous rate maximization and then rounded it off to the integer rates. On the other hand, [18] adopted uniform power allocation to reduce complexity.

### B. Bit Loading for Relay Communication

Studies on bit allocation for multi-channel relaying has so far been scarce. In [19], the authors investigated the bit loading and power allocation problem for both AF and DF relaying OFDM systems with an aim to minimize the total

transmit power subject to a target bit rate and symbol error rate. Under total power constraint, they converted the problem to that of a equivalent point-to-point system, and developed the optimal solution based on the greedy approach in [11]. For a scenario with individual power constraints, a suboptimal approach was proposed assuming fixed source-relay power ratio for each channel. For a multi-hop DF relaying system, [20] proposed a suboptimal approach for power optimization through bit loading, bandwidth and channel assignment with a minimum required rate for each hop. The greedy algorithm [11] is applied over each individual hop to meet the minimum target bit rate with minimal power. Neither of these works have considered channel pairing. The common technique in these works is to convert the dual-hop system to an equivalent point-to-point system, where the conventional optimal greedy algorithm can be applied. However, when channel pairing needs to be jointly optimized, such technique is no longer applicable, because the conversion of the dual-hop system to an equivalent point-to-point system changes whenever a different channel pairing is used.

### C. Channel Pairing with Information Theoretic Rate

For multi-channel relaying, channel pairing was first proposed in [21] and [22]. Its optimization for sum-rate maximization was considered in [23], [24] for a given power allocation. Jointly optimizing channel pairing and power allocation for single-user relaying was considered in several studies for sum-rate maximization, including [25], [26] for dual-hop DF relaying and [27] for general multi-hop AF or DF relaying, all without direct link, and [6], [28] with direct link. For relay-assisted multi-user scenarios, joint optimization of channel pairing, channel-user assignment, and power allocation for maximizing weighted sum-rate was obtained in [7]. All of these studies on channel pairing adopt information theoretic achievable rate as the objective.

### D. Channel Pairing with Discrete Bit Rate

To the best of our knowledge, [29] is the only work considering the joint resource allocation problem consisting of channel pairing and bit and power allocation in a multi-channel system. It focused on DF relaying and total power constraint. Aiming at minimizing the total transmit power under a target bit rate constraint, the authors proposed a suboptimal approach based on separate optimization, where pairing is conducted first and then, similar to [19] and [20], the greedy algorithm [11] is applied over the paired channels that can be viewed as an equivalent point-to-point transmission. In contrast, we present a new approach to jointly optimize channel pairing, power allocation, and bit loading. In Section VII, we demonstrate that our jointly optimal solution can significantly outperform many simpler suboptimal solutions, including one based on the above separate optimization.

## III. SYSTEM MODEL AND PROBLEM FORMULATION

We consider a three-node cooperative scenario where a source communicates with a destination with the assistance of a relay. We assume a broadband multi-channel system where

the available communication spectrum is partitioned into an equivalent set of  $N$  channels, each being used for sending a block of information symbols. A practical example is the OFDM system with  $N$  equal-bandwidth subcarriers.

#### A. Channel Model and Transmission Strategy

We denote by  $h_n$  and  $g_n$  for  $n = 1, \dots, N$ , the channel gain on channel  $n$  over the first hop and the second hop, respectively. The additive noise at the relay and the destination are modeled as zero-mean i.i.d. Gaussian random variables with variance  $\sigma_r^2$  and  $\sigma_d^2$ , respectively. In the system formulation, we always use the normalized channel power gain, which for channel  $n$  of the first and second hops are respectively defined as  $a_n \triangleq \frac{|h_n|^2}{\sigma_r^2}$  and  $b_n \triangleq \frac{|g_n|^2}{\sigma_d^2}$ .

A transmission from the source to the destination takes place in two time slots. In the first time slot, the source transmits a data block to the relay on each channel. In the second time slot, the relay forwards a processed version of the received signals over the outgoing channels to the destination. We are not limited to any specific relaying strategy (DF, AF, etc), as long as the rate-power relation is known and satisfies some conditions as explained below.

#### B. Channel Pairing

In channel pairing, the relay matches each incoming channel with an outgoing channel. As different channels experience different quality and condition, an adaptive channel pairing scheme is essential for maximum spectral efficiency. We say path  $(m, n)$  is selected if incoming channel  $m$  is paired with outgoing channel  $n$ . We further define a binary pairing indicator

$$\phi_{mn} \in \{0, 1\}, \quad \forall m, n, \quad (1)$$

such that  $\phi_{mn}$  is 1 if path  $(m, n)$  is selected and 0 otherwise. The channel pairing matrix is then defined as  $\Phi = [\phi_{mn}]_{N \times N}$ . We consider one-to-one mappings between the incoming channels and outgoing channels. This imposes the following constraints on  $\Phi$ :

$$\sum_{n=1}^N \phi_{mn} = 1, \quad \forall m; \quad \sum_{m=1}^N \phi_{mn} = 1, \quad \forall n. \quad (2)$$

#### C. Power Allocation

We assume there is a single power constraint  $P_t$ . It can be the total power constraint on the source and relay transmissions, which may be due to, *e.g.*, hardware limitations or legal mandates on the overall interference created by the system. It can also be an individual power constraint imposed on the source or the relay.

Let  $p_{mn}$  be the part of  $P_t$  adaptively allocated to path  $(m, n)$ , and define the power allocation matrix  $\mathbf{p} = [p_{mn}]_{N \times N}$ . The power constraint is then expressed by

$$p_{mn} \geq 0, \quad (3)$$

$$\sum_{m=1}^N \sum_{n=1}^N \phi_{mn} p_{mn} \leq P_t. \quad (4)$$

#### D. Bit Loading

For each path, the data transmission may adopt varying rates based on the modulation or coding scheme in use. We define a bit loading matrix  $\mathbf{b} = [b_{mn}]_{N \times N}$ , where  $b_{mn}$  denotes the number of bits per transmission over path  $(m, n)$ . We assume  $b_{mn}$  takes only non-negative integral values, *i.e.*,

$$b_{mn} \in \mathcal{M}, \quad (5)$$

where  $\mathcal{M} = \{0, 1, 2, \dots, M\}$ . For example, in  $x$ -QAM, we have  $x = 2^{b_{mn}}$  for  $b_{mn} = 1, 2, \dots$ . Note that  $b_{mn} = 0$  corresponds to the case where either path  $(m, n)$  is not selected or not allocated sufficient power for useful transmission.

In order to support the transmission of  $b_{mn}$  bits, the allocated power  $p_{mn}$  must satisfy

$$b_{mn} \leq R_{mn}(p_{mn}), \quad (6)$$

where  $R_{mn}$  is in general a real-valued function that depends on the channel power gains  $a_m$  and  $b_n$ . Moreover, other system parameters as well as performance requirements may be involved in constructing  $R_{mn}$ , such as the modulation and coding schemes, the relaying strategy, and the required error rate constraints.

For instance, with a repetition-coding based DF relaying strategy, given a total power constraint  $p_{mn}$  between the source and the relay, we may have  $R_{mn}(p_{mn}) = \frac{1}{2} \log(1 + \Gamma^{-1} a_{mn} p_{mn})$  on path  $(m, n)$ , where  $a_{mn} = (\frac{1}{a_m} + \frac{1}{b_n})^{-1}$ , and  $\Gamma$  is an SNR gap corresponding to the distance between the theoretic capacity and the achievable rate due to modulation and coding. As another example, for AF relaying under relay power constraint  $p_{mn}$ , the achievable rate is given by  $R_{mn}(p_{mn}) = \frac{1}{2} \log(1 + \frac{\Gamma^{-1} p_m^s a_m p_{mn} b_n}{1 + p_m^s a_m + p_{mn} b_n})$ , where  $p_m^s$  denotes the given source power on channel  $m$  of the first hop. In this study, we consider relaying strategies whose  $R_{mn}(p_{mn})$  is a strictly increasing and concave function of  $p_{mn}$ . Both examples above can be shown to satisfy this assumption.

#### E. Optimization Problem

Our objective is to design a scheme that maximizes the overall source-destination bit rate. It is cast in the following form:

$$B^*(P_t) = \max_{\Phi, \mathbf{b}, \mathbf{p}} \sum_{m=1}^N \sum_{n=1}^N \phi_{mn} b_{mn} \quad (7)$$

*s.t.* (1), (2), (3), (4), (5), (6).

For given total power budget  $P_t$ , the optimization problem (7) seeks the jointly optimal solution for channel pairing, power allocation, and bit loading by optimizing  $(\Phi, \mathbf{b}, \mathbf{p})$ .

For convenience of exposition in the rest of this paper, we also consider the minimum required power given an integer sum rate target  $B$ :

$$P_t^*(B) = \min_{\Phi, \mathbf{b}, \mathbf{p}} \sum_{m=1}^N \sum_{n=1}^N \phi_{mn} p_{mn} \quad (8)$$

*s.t.* (1), (2), (3), (5), (6)

$$\sum_{m=1}^N \sum_{n=1}^N \phi_{mn} b_{mn} = B.$$

We note that  $P_t^*(k) < P_t^*(l)$  for all  $k < l$ , so  $B^*(P_t)$  is an integer-valued increasing staircase function in  $P_t$ . Let  $\Delta P_t^*(k) \triangleq P_t^*(k+1) - P_t^*(k)$ . We consider systems that have the following *strictly* diminishing rate-power relation:

$$\Delta P_t^*(k) > \Delta P_t^*(k-1), \quad \forall k \geq 0. \quad (9)$$

This reflects the general reality in practical systems that, as the data rate increases, each one-bit increment requires more and more marginal power. This can also be viewed as a discrete and strict version of the time-sharing property, which is widely applicable in broadband wireless resource management [9].

The optimization problem (7) is non-convex, with integer constraints on both the objective function and the feasibility region. Such problems usually is NP-hard owing to the combinatorial nature of channel pairing and bit loading. The Lagrange dual decomposition method is a commonly used approach to tackle this type of problems. In a recent study, the authors have found an efficient method to jointly optimize channel assignment and power allocation for systems without bit loading [7], showing that a zero Lagrangian duality gap can be achieved even with discrete channel pairing. However, the discrete nature of the objective function in (7) renders that method inapplicable.

Nonetheless, the Lagrange dual decomposition method provides an upper bound to the optimal solution. In Sections IV and V, we show that an optimal solution can be extracted from the minimizer of the Lagrange dual function, even though the problem under consideration has a positive duality gap. Furthermore, in Section VII we observe that the proposed solution can be achieved within polynomial time in  $N$ , providing substantial savings in computing resources over brute-force search methods.

#### IV. UPPER BOUND TO $B^*(P_t)$ VIA DUAL DECOMPOSITION

Let  $\lambda$  be the Lagrange multiplier associated with the power constraint (4). We have the following Lagrange function for (7):

$$\mathcal{L}(\Phi, \mathbf{p}, \mathbf{b}, \lambda) = \sum_{m,n} \phi_{mn} b_{mn} - \lambda \left( \sum_{m,n} \phi_{mn} p_{mn} - P_t \right). \quad (10)$$

The dual function  $g(\lambda)$  is defined as the maximum of (10):

$$g(\lambda) = \max_{\Phi, \mathbf{p}, \mathbf{b}} \mathcal{L}(\Phi, \mathbf{p}, \mathbf{b}, \lambda) \quad (11)$$

*s.t.* (1), (2), (3), (5), (6).

Then, due to the weak duality property of Lagrange decomposition [30], we can obtain an upper bound to  $B^*(P_t)$  by taking the minimum of  $g(\lambda)$ , *i.e.*,

$$B^*(P_t) \leq B_{LD} \triangleq \min_{\lambda} \{g(\lambda) : \lambda \geq 0\}. \quad (12)$$

The maximization and minimization problems of (11) and (12) can be solved in the following three consecutive steps.

#### A. Power and Bit Allocation Given $\lambda$

Let  $(\Phi^*(\lambda), \mathbf{b}^*(\lambda), \mathbf{p}^*(\lambda))$  be an optimal solution to (11). We first consider maximizing  $\mathcal{L}(\Phi, \mathbf{p}, \mathbf{b}, \lambda)$  over  $\mathbf{b}$  and  $\mathbf{p}$  for any given  $\Phi$ .

This problem can be decomposed into  $N \times N$  underlying independent subproblems, concerning the optimal power and bit allocation for each possible path  $(m, n)$ .

$$\begin{aligned} & \max_{p_{mn}, b_{mn}} b_{mn} - \lambda p_{mn} \quad (13) \\ \text{s.t. } & \text{i) } b_{mn} \leq R_{mn}(p_{mn}), \\ & \text{ii) } p_{mn} \geq 0, \quad b_{mn} \in \mathcal{M}. \end{aligned}$$

Since  $R_{mn}(p_{mn})$  is an increasing function of  $p_{mn}$ , it is not difficult to see that, at optimality, the constraint (i) in (13) must be satisfied by equality, *i.e.*,

$$p_{mn}^*(\lambda) = R_{mn}^{-1}(b_{mn}^*(\lambda)). \quad (14)$$

Furthermore, since  $R_{mn}(p_{mn})$  is a strictly concave function, there is a unique maximizer  $b_{mn}^*(\lambda)$  for (13). Substituting (14) into (13), we obtain a discrete concave function and can find  $b_{mn}^*(\lambda)$  by using an efficient algorithm such as integer bisection search over the domain of  $b_{mn}$ .

#### B. Channel Pairing Given $\lambda$

Next, we substitute  $b_{mn}^*(\lambda)$  and  $p_{mn}^*(\lambda)$  into (11) and optimize it over  $\Phi$ , for the given  $\lambda$ . Let  $\mathcal{L}_{mn}^*(\lambda) = b_{mn}^*(\lambda) - \lambda p_{mn}^*(\lambda)$ . Then the maximization in (11) is equivalent to

$$\begin{aligned} g(\lambda) &= \max_{\Phi} \sum_{m,n} \phi_{mn} \mathcal{L}_{mn}^*(\lambda) \quad (15) \\ \text{s.t. } & (1), (2). \end{aligned}$$

This optimization is a two-dimensional assignment problem [8], whose binary solution  $\Phi^*(\lambda)$  can be efficiently obtained by several existing methods, such as the Hungarian Algorithm, which has computational complexity  $O(N^3)$ .

#### C. Dual Minimization through Subgradient Updating

The dual minimization problem (12) can be solved by employing the subgradient algorithm [30]. For completeness, this baseline algorithm is summarized in the following.

- 1) Initialize  $\lambda^{(0)}$ .
- 2) Given  $\lambda^{(l)}$ , obtain  $b_{mn}^*(\lambda^{(l)})$ ,  $p_{mn}^*(\lambda^{(l)})$  and  $\phi_{mn}^*(\lambda^{(l)})$  for all  $(m, n)$  (Sections IV-A, IV-B).
- 3) Update  $\lambda$  through  $\lambda^{(l+1)} = \max\{\lambda^{(l)} - \theta(\lambda^{(l)})\nu^{(l)}, 0\}$  where we use the subgradient

$$\theta(\lambda) = P_t - \sum_{m,n} \phi_{mn}^*(\lambda) p_{mn}^*(\lambda), \quad (16)$$

and  $\nu^{(l)}$  is the step size at the  $l$ th iteration.

- 4) Set  $l = l + 1$ ; repeat from Step (2) until the convergence of  $\min_l g(\lambda^{(l)})$ .

We may adopt the square-summable, non-summable step-size rule [31], under which the algorithm is guaranteed to converge optimally. Let  $\lambda^*$  be the global minimizer for the dual minimization problem (12).

The above Lagrange dual method is widely used to provide an upper bound to the optimal objective function. However,

for general non-convex optimization problems, there is no guarantee that the primal solution corresponding to  $\lambda^*$  is feasible, so generally this method does not produce an optimal solution. As an alternative, in the general context of mixed-integer programming, the bound  $B_{LD}$  can be exploited by other algorithms such as branch-and-bound to find the optimal solution [8]. However, the application of such an approach usually requires exponential computational complexity due to the combinatoric nature of the problem.

Nonetheless, we next show a surprising observation that, in the problem under consideration, a jointly optimal solution for channel pairing, power allocation, and bit loading can be efficiently extracted, despite the non-zero duality gap.

## V. EXTRACTION OF JOINTLY OPTIMAL SOLUTION

There is a fundamental difference between our proposed approach and the conventional Lagrange dual approaches that require a zero duality gap to derive an optimal solution [8]. In this section we show that a small perturbation of  $\lambda^*$  allows the extraction of the optimal solution to the primal problem (7).

In the subsequent derivations, we define

$$B^*(\lambda) \triangleq \sum_{m,n} \phi_{mn}^*(\lambda) b_{mn}^*(\lambda), \quad (17)$$

$$P^*(\lambda) \triangleq \sum_{m,n} \phi_{mn}^*(\lambda) p_{mn}^*(\lambda). \quad (18)$$

We further denote by  $\lambda^{*+}$  and  $\lambda^{*-}$ , respectively,  $\lambda^* + d\lambda$  and  $\lambda^* - d\lambda$ , where  $d\lambda$  is an infinitesimal increment in  $\lambda$ . We will use also the unit step function  $u(z - c) \triangleq \begin{cases} 0, & z < c, \\ 1, & z \geq c \end{cases}$ .

Our main result is stated in the following theorem:

**Theorem 1:** ( $\Phi = [\phi_{m,n}^*(\lambda^{*+})]$ ,  $\mathbf{p} = [p_{m,n}^*(\lambda^{*+})]$ ,  $\mathbf{b} = [b_{m,n}^*(\lambda^{*+})]$ ) is an optimal solution to (7).

This theorem indicates that, after  $\lambda^*$  is obtained as described in Section IV-C, one may simply substitute  $\lambda = \lambda^{*+}$  into Sections IV-A and IV-B to extract the jointly optimal channel pairing, power allocation, and bit loading solution. We provide the proof of Theorem 1 in Section V-C, after presenting in Sections V-A and V-B several interesting observations and required lemmas for the proof.

### A. Properties of $B^*(\lambda)$

As shown in Section III-E, it is clear that  $B^*(P_t)$  is an integer-valued increasing staircase function in the allowed total power budget  $P_t$ . The following lemma states a less obvious, but important fact that each step of the staircase has unit height, *i.e.*, if  $B^*(P_t) = k$  is achievable for some  $P_t$  then  $B^*(P_t) = i$  is achievable for some  $P_t$  for all  $i = 1, 2, \dots, k-1$ . Its proof is provided in Appendix A.

**Lemma 1:**  $B^*(P_t)$  is a unit-step increasing staircase function of  $P_t$ , such that

$$B^*(P_t) = \sum_k u(P_t - P_t^*(k)). \quad (19)$$

Figure 1 illustrates the staircase function  $B^*(P_t)$ . Note that the corner points  $(P_t^*(k), k)$  are also the points where

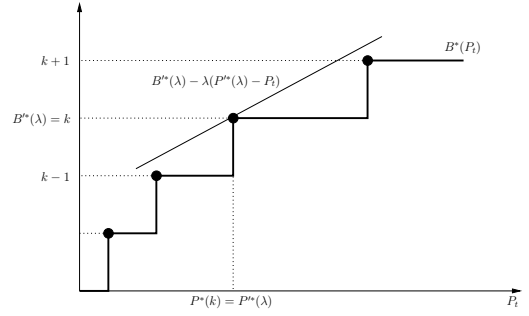


Fig. 1. Function  $B^*(P_t)$ . The point  $(P_t^*(\lambda), B^*(\lambda))$  is on a corner, and  $g(\lambda)$  passes through it.

$B^*(P_t)$  is achieved with equality for the power constraint (4). Further observing the line that represents  $g(\lambda)$ , this leads to the following lemma, whose proof is given in Appendix B.

**Lemma 2:** For all  $\lambda \geq 0$ , the point  $(P_t^*(\lambda), B^*(\lambda))$  always falls on the curve  $B^*(P_t)$  described by (7). Furthermore, for all  $\lambda > 0$ ,  $P_t^*(\lambda) = P_t^*(B^*(\lambda))$ .

Lemma 2 states that, if we substitute  $P_t = P_t^*(\lambda)$  into (7) and solve the optimization problem, the resultant maximum sum rate is  $B^*(\lambda)$ , and if we substitute  $B = B^*(\lambda)$  into (8) and solve the optimization problem, the resultant minimum required power is  $P_t^*(\lambda)$ .

We can then draw the following conclusion:

**Proposition 1:**  $B^*(\lambda)$  is a unit-step decreasing staircase function of  $\lambda$ .

*Proof:* We need to show that  $B^*(\lambda)$  is a decreasing staircase function, and that it does not jump down by more than one for any small increase in  $\lambda$ , *i.e.*,

$$B^*(\lambda) - B^*(\lambda + d\lambda) \leq 1. \quad (20)$$

The first assertion can be seen considering the form of (11) and its geometric representation on Figure 1 as the highest line with slope  $\lambda$  that touches a point  $(P_t^*(k), k)$  on the function  $B^*(P_t)$ . More formally, suppose there exist some  $\lambda_1 < \lambda_2$  such that  $B^*(\lambda_1) < B^*(\lambda_2)$ . Then by Lemma 2 we have

$$P_t^*(\lambda_1) = P_t^*(B^*(\lambda_1)) < P_t^*(B^*(\lambda_2)) = P_t^*(\lambda_2). \quad (21)$$

Furthermore, from (11) we have

$$B^*(\lambda_2) - \lambda_1 P_t^*(\lambda_2) \leq B^*(\lambda_1) - \lambda_1 P_t^*(\lambda_1) \quad (22)$$

$$B^*(\lambda_1) - \lambda_2 P_t^*(\lambda_1) \leq B^*(\lambda_2) - \lambda_2 P_t^*(\lambda_2). \quad (23)$$

Summing the above, and re-arranging, we have

$$(\lambda_1 - \lambda_2)(P_t^*(\lambda_1) - P_t^*(\lambda_2)) \leq 0, \quad (24)$$

which contradicts (21).

We use contradiction again to show the second assertion. Suppose for some  $\lambda, k$ , and  $1 \leq i \leq k$ , we have  $B^*(\lambda) = k+1$  and  $B^*(\lambda^+) = k-i$ , where  $\lambda^+ = \lambda + d\lambda$ . Consider the points  $(P_t^*(\lambda^+), k-i)$  and  $(P_t^*(\lambda), k+1)$  on the  $P_t$ - $B^*$  plane. By Lemma 2, both points are on the curve  $B^*(P_t)$ , and in particular  $P_t^*(k-i) = P_t^*(\lambda^+)$  and  $P_t^*(k+1) = P_t^*(\lambda)$ .

Now consider the point  $(P_t^*(k), k)$ . Lemma 1 ensures that such a point exists. Furthermore, we have

$$\begin{aligned} k - \lambda P_t^*(k) &= k - \lambda^+ P_t^*(k) \\ &> k - \lambda^+ \left( \frac{i}{i+1} P_t^*(k+1) + \frac{1}{i+1} P_t^*(k-i) \right) \end{aligned}$$

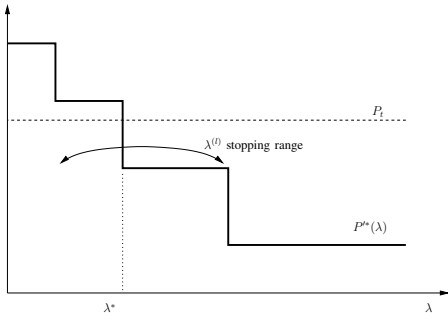


Fig. 2. Function  $P^{l*}(\lambda)$ . It is discontinuous at  $\lambda = \lambda^*$ .

$$= \frac{i}{i+1}((k+1) - \lambda P^{l*}(\lambda)) + \frac{1}{i+1}((k-i) - \lambda^+ P^{l*}(\lambda^+)), \quad (25)$$

where the second line is a consequence of (9) by induction. This implies that either one of the following inequalities must hold,

$$k - \lambda P_t^*(k) > (k+1) - \lambda P^{l*}(\lambda), \quad (26)$$

$$k - \lambda^+ P_t^*(k) > (k-i) - \lambda^+ P^{l*}(\lambda^+), \quad (27)$$

which contradicts with the fact that  $P^{l*}(\lambda)$  and  $P^{l*}(\lambda^+)$  respectively maximize (11) for the given  $\lambda$  and  $\lambda^+$ . ■

### B. Properties of $P^{l*}(\lambda)$

Proposition 1 immediately leads to the following property for  $P^{l*}(\lambda)$ , whose proof is given in Appendix C.

*Lemma 3:*  $P^{l*}(\lambda)$  is a decreasing staircase function of  $\lambda$ . Furthermore,  $P^{l*}(\lambda)$  and  $B^{l*}(\lambda)$  have the same discontinuity points in  $\lambda$ .

Figure 2 illustrates the staircase function  $P^{l*}(\lambda)$ . We can further characterize the relation between  $P_t$  and  $P^{l*}(\lambda)$  near  $\lambda = \lambda^*$  in the following:

*Proposition 2:* If  $\lambda^* > 0$ , then  $\lambda^*$  is at a point where the function  $P^{l*}(\lambda)$  is discontinuous, and furthermore

$$P^{l*}(\lambda^{*+}) \leq P_t < P^{l*}(\lambda^{*-}). \quad (28)$$

*Proof:* If there exists some  $\lambda$  such that  $P^{l*}(\lambda) = P_t$ , then the subgradient at  $\lambda$  as defined in (16) is zero. Since  $\lambda$  is a global minimizer of  $g(\lambda)$  as long as there exists one zero subgradient at  $\lambda$  [30], we have  $\lambda^* = \lambda$ . Then, since  $P^{l*}(\lambda)$  is a decreasing staircase function by Lemma 3, we have (28).

If there does not exist  $\lambda$  such that  $P^{l*}(\lambda) = P_t$ , then the minimum subgradient in (16) is

$$\theta_m \triangleq \min_{\lambda} |P_t - P^{l*}(\lambda)| > 0. \quad (29)$$

Consider the subgradient updating algorithm to minimize  $g(\lambda)$ . Suppose at iteration  $l$ , we have  $P^{l*}(\lambda^{(l)}) < P_t$ . The Lagrange multiplier for the next iteration is updated as  $\lambda^{(l+1)} = \lambda^{(l)} + \nu^{(l)}(P^{l*}(\lambda^{(l)}) - P_t)$ , which results in  $\lambda^{(l+1)} < \lambda^{(l)}$ . In fact, as long as  $P^{l*}(\lambda^{(l)}) < P_t$ ,  $\lambda$  continues to decrease as  $l$  increases. Since  $\{\nu^{(l)}\}$  is a non-summable series, and  $P_t - \lambda^{(l)} \geq \theta_m > 0$  for all  $\lambda^{(l)}$ , this process can continue either until  $\lambda$  decreases to 0 or until an iteration  $k > l$  such that  $P^{l*}(\lambda^{(k)}) > P_t$ . However, since the former case cannot occur by the proposition's assumption that  $\lambda^* > 0$ ,

the latter case holds. Similarly, suppose at iteration  $l'$  we have  $P^{l*}(\lambda^{(l')}) > P_t$ , then there exists an iteration  $k' > l'$ , such that  $P^{l*}(\lambda^{(k')}) < P_t$ .

The convergence of the subgradient updating algorithm, in terms of  $\min_l g(\lambda^{(l)})$ , is guaranteed [30]. This implies that  $\lambda^{(l)}$  converges to  $\lambda^*$ , since there are multiple minimizers only when there exists  $\lambda$  such that  $P^{l*}(\lambda) = P_t$ . By the above analysis, at convergence when  $\lambda^*$  is obtained, we can have neither  $P^{l*}(\lambda^*) > P_t$  nor  $P^{l*}(\lambda^*) < P_t$ . The only possible candidate for such  $\lambda^*$  is at a discontinuity point of the function  $P^{l*}(\lambda)$  where (28) holds. ■

### C. Proof of Main Theorem

We are now ready to prove Theorem 1.

*Proof:* If  $\lambda^* = 0$ , we must have  $P_t^*(0) \leq P_t$ , since otherwise the objective in (12) can be decreased by increasing  $\lambda^*$  from 0. By Lemma 3, we have  $P_t^*(\lambda^{*+}) \leq P_t$ , so the solution  $(\Phi = [\phi_{m,n}^*(\lambda^{*+})], \mathbf{p} = [p_{m,n}^*(\lambda^{*+})], \mathbf{b} = [b_{m,n}^*(\lambda^{*+})])$  is feasible. Furthermore, we have  $B^{l*}(\lambda^{*+}) = B^{l*}(0)$ . Since  $B^{l*}(0)$  maximizes the objective in (11), and that is the same as maximizing the objective in (7) when  $\lambda = 0$ , we conclude that the solution  $(\Phi = [\phi_{m,n}^*(\lambda^{*+})], \mathbf{p} = [p_{m,n}^*(\lambda^{*+})], \mathbf{b} = [b_{m,n}^*(\lambda^{*+})])$  is optimal.

If  $\lambda^* > 0$ , Proposition 2 states that  $\lambda^*$  is at a discontinuity point of the function  $P^{l*}(\lambda)$ . Then, by Lemma 3,  $\lambda^*$  is also at a discontinuity point of the function  $B^{l*}(\lambda)$ . From Proposition 1, we see that  $B^{l*}(\lambda)$  is a unit-step decreasing staircase function. Hence,

$$B^{l*}(\lambda^{*+}) = B^{l*}(\lambda^{*-}) - 1. \quad (30)$$

Then, we have

$$\begin{aligned} B^*(P_t) &\leq g(\lambda^*) \\ &\leq g(\lambda^{*-}) = B^{l*}(\lambda^{*-}) - \lambda^{*-}(P^{l*}(\lambda^{*-}) - P_t) \\ &< B^{l*}(\lambda^{*-}) = B^{l*}(\lambda^{*+}) + 1, \end{aligned} \quad (31)$$

where the first line is due to weak duality, the second line results from the definition of the dual function (11), and the third line is due to  $P^{l*}(\lambda^{*-}) > P_t$  from Proposition 2. Since  $B^*(P_t)$  is an integer, it follows that

$$B^*(P_t) \leq B^{l*}(\lambda^{*+}). \quad (32)$$

However, since  $B^{l*}(\lambda^{*+})$  is feasible, we also have

$$B^*(P_t) \geq B^{l*}(\lambda^{*+}). \quad (33)$$

This implies that  $B^{l*}(\lambda^{*+})$  is the maximum achievable sum rate, i.e.,  $(\Phi = [\phi_{m,n}^*(\lambda^{*+})], \mathbf{p} = [p_{m,n}^*(\lambda^{*+})], \mathbf{b} = [b_{m,n}^*(\lambda^{*+})])$  is an optimal solution. ■

## VI. COMPLEXITY REDUCTION AND ANALYSIS

The baseline subgradient updating algorithm presented in Section IV-C, using the non-summable, square summable step-size rule, requires an infinite number of iterations for convergence. It presents challenges in setting a precise criterion for convergence testing. However, we can drastically reduce the computational complexity of the proposed method by terminating the subgradient updating iterations before convergence to  $\lambda^*$ , which still maintains our ability to recover the optimal primal solution. A salient observation here is that, as long as

$\lambda^{(l)}$  reaches a point where  $B^{*}(\lambda^{(l)})$  is at the same plateau as  $B^{*}(\lambda^{*+})$ , we may stop and use  $\lambda^{(l)}$  to extract the optimal primal solution.

The following theorem formalizes this idea. It gives two conditions upon which the algorithm may be safely terminated. Figure 2 contains a pictorial illustration of the early termination conditions.

*Theorem 2:* At iterations  $l$  and  $l + 1$ , if

- 1)  $|B^{*}(\lambda^{(l)}) - B^{*}(\lambda^{(l+1)})| = 1$ , and
- 2)  $\text{sgn}(P_t - P^{*}(\lambda^{(l)})) \neq \text{sgn}(P_t - P^{*}(\lambda^{(l+1)}))$ ,

then  $B^{*}(P_t) = \min\{B^{*}(\lambda^{(l)}), B^{*}(\lambda^{(l+1)})\}$ .

*Proof:* Suppose  $B^{*}(\lambda^{(l)}) > B^{*}(\lambda^{(l+1)})$ . This gives rise to  $P^{*}(\lambda^{(l)}) > P^{*}(\lambda^{(l+1)})$ , since both  $B^{*}$  and  $P^{*}$  are decreasing in  $\lambda$ . Thus, condition 2) implies  $P^{*}(\lambda^{(l+1)}) < P_t < P^{*}(\lambda^{(l)})$ . Similar to (31), we have

$$B^{*}(P_t) < B^{*}(\lambda^{(l)}) . \quad (34)$$

Since  $B^{*}(P_t)$  is an integer, it follows that

$$B^{*}(P_t) \leq B^{*}(\lambda^{(l)}) - 1 = B^{*}(\lambda^{(l+1)}), \quad (35)$$

where the equality is due to condition 1) above. Since  $B^{*}(\lambda^{(l+1)})$  is also a feasible solution with  $P^{*}(\lambda^{(l+1)}) < P_t$ , (35) implies that  $B^{*}(P_t) = B^{*}(\lambda^{(l+1)})$ .

For the case where  $B^{*}(\lambda^{(l)}) < B^{*}(\lambda^{(l+1)})$ , a similar argument applies by swapping index  $(l + 1)$  for  $(l)$ , to show that  $B^{*}(P_t) = B^{*}(\lambda^{(l)})$ . ■

From this, we see that there is no need for the subgradient updating algorithm to converge exactly to  $\lambda^{*}$ . Instead, when  $\lambda^{(l)}$  reaches within a certain neighborhood of  $\lambda^{*}$ , which is determined by the plateau defined by  $B^{*}(\lambda^{*+}) = B^{*}(P_t)$ , the optimal primal solution can be extracted. This enables a variant of the proposed algorithm that can be shown to require only polynomial complexity in the number of channels  $N$ .

We first make the following observation that  $\lambda^{*}$  is bounded by a constant. The proof is provided in Appendix D.

*Lemma 4:*  $\lambda^{*}$  is upper bounded by a constant  $\lambda_{max}$  independent of  $N$ .

Then, we may modify the original subgradient updating in Section IV-C to the following without affecting the optimality of the overall method:

$$\lambda^{(l+1)} = \min\{\max\{\lambda^{(l)} - \theta(\lambda^{(l)})\nu^{(l)}, 0\}, \lambda_{max}\} . \quad (36)$$

Furthermore, without the need for exact convergence to  $\lambda^{*}$ , we may adopt a constant step-size rule [31], *i.e.*,  $\nu^{(l)} = \nu$ , which improves the convergence speed when the iteration count is high. The polynomial-time complexity of this alternate method is stated in the following theorem:

*Theorem 3:* To achieve within an arbitrary  $\epsilon > 0$  neighborhood of the optimal  $g(\lambda^{*})$ , using a constant step-size in subgradient updating, the alternate method with (36) has polynomial computational complexity in  $N$ .

*Proof:* At each iteration of the subgradient updating algorithm, the procedures described in Sections IV-A and IV-B have overall computational complexity polynomial in  $N$ , since the two-dimensional assignment problem can be solved in  $O(N^3)$ . Therefore, it remains to show that the total number of iterations is polynomial in  $N$ .

At each iteration  $l$ , the subgradient is

$$\theta(\lambda^{(l)}) = P_t - P^{*}(\lambda^{(l)}) \leq P_t . \quad (37)$$

Furthermore,  $P^{*}(\lambda^{(l)})$  is upper bounded by the total power to support transmission at the maximum data rate  $M$  for each of the  $N$  chosen paths, which has order  $O(N)$ . Hence,  $|\theta(\lambda^{(l)})|$  is upper bounded by  $\Theta = O(N)$ .

If we choose  $\lambda^{(0)}$  in the interval  $[0, \lambda_{max}]$ , then the distance between  $\lambda^{(0)}$  and  $\lambda^{*}$  is upper bounded by  $\lambda_{max}$ . Then, it can be shown that, at the  $l$ th iteration, the distance between the current best objective to the optimum objective  $g(\lambda^{*})$  is upper bounded by  $\frac{\lambda_{max}^2 + \nu^2 \Theta^2 l}{2\nu l}$  if a constant step-size  $\nu^{(l)} = \nu$  is used [31]. If we set  $\nu = \epsilon/\Theta^2$ , it is easy to see that the above bound is less than  $\epsilon$  when  $l > \lambda_{max}^2 \Theta^2 / \epsilon^2 = O(N^2)$ . Hence, the number of required iterations until convergence is  $O(N^2)$ . ■

Most mixed integer optimization problems are NP-hard, and in particular, many previously proposed solutions to optimize resource allocation with bit loading are suboptimal. Therefore, it is notable to observe that for the problem under consideration, one can achieve jointly optimal solutions using the proposed method within polynomial time in  $N$ . In the context of practical OFDM systems such as IEEE 802.16, which operate over hundreds or thousands of subcarriers, the proposed method provides substantial savings in computing resources over brute-force search methods.

## VII. NUMERICAL RESULTS

This section presents simulation results to assess and compare the performance of the jointly optimal scheme with that of the following suboptimal schemes:

- Separate Optimization: The two-stage solution proposed in [29], which first pairs the channels based on their sorted channel gains and then allocates bits over the paired channels via the greedy algorithm in [11].
- No Pairing: Carry out bit loading and power allocation without pairing (*i.e.*, same channel is used for relaying a received signal from that channel).
- Uniform PA: Allocate available power uniformly among the paths without pairing.

We consider a multi-channel AF relaying system where a relay lies between a pair of source and destination with  $d_{sr} = d_{rd} = d$ , where  $d_{sr}$  and  $d_{rd}$  denote the source-relay and relay-destination distances, respectively. A 4-tap frequency selective propagation channel is assumed for each hop. To normalize the effect of power over different network settings, we define the average received SNR under a hypothetical uniform allocation of the total transmission power, which for the S-R link is  $\text{SNR}_{\text{avg}}^{S-R} = \frac{P_t^s}{N\sigma_r^2 d_{sr}^{\kappa}}$  and for the R-D link is  $\text{SNR}_{\text{avg}}^{R-D} = \frac{P_t^r}{N\sigma_d^2 d_{rd}^{\kappa}}$ , where  $P_t^s$  and  $P_t^r$  are the total source and relay available powers, respectively, and  $\kappa = 3$  denotes the pathloss exponent.

In this example, we focus on optimal resource allocation at the relay. Hence,  $P_t$  and  $p_{mn}$  in our analysis here refer to  $P_t^r$  and the power allocated by the relay to path  $(m, n)$ , respectively. The SNR gap  $\Gamma$  is set to 1. Then the total achievable

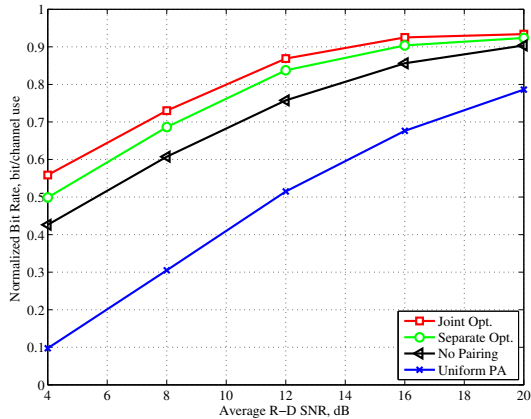


Fig. 3. Per channel rate vs.  $\text{SNR}_{\text{avg}}^{R-D}$  for multi-channel AF relaying with  $\text{SNR}_{\text{avg}}^{S-R} = 10\text{dB}$ .

rate function  $R_{mn}(p_{mn}) = \frac{1}{2} \log(1 + \frac{p_m^s a_m p_{mn} b_n}{1 + p_m^s a_m + p_{mn} b_n})$ , where  $p_m^s$  is a given power allocated to the S-R channel  $m$ . In the following, we assume uniform power allocation at the source across all S-R channels, so  $p_m^s = \frac{P_s}{N}$ , although other arbitrary power allocation is applicable as well. The maximum bit rate per channel  $M$  is set to 5.

Our first experiment compares the performance of various schemes at different  $\text{SNR}_{\text{avg}}^{R-D}$  levels for  $N = 16$ . Fig. 3 depicts the average bit rate per channel vs.  $\text{SNR}_{\text{avg}}^{R-D}$ , while  $\text{SNR}_{\text{avg}}^{S-R}$  is fixed to 10 dB. As we observe, the jointly optimal scheme outperforms the other suboptimal solutions, offering as much as 1dB improvement over the Separate Optimization scheme for a large range of SNR values. Furthermore, the only difference between Separate Optimization and No Pairing is that the former scheme conducts pairing while the latter does not. It is therefore interesting to notice the substantial gain provided by channel pairing alone. At high received  $\text{SNR}_{\text{avg}}^{R-D}$ , it is anticipated that the performance of all schemes converge to a single point as the S-R link becomes the bottleneck of communication.

The second experiment examines the effect of the number of channels on the performance of various resource allocation schemes. This is done under the same level of channel gain variation across channels, and under the fixed channel spacing in terms of bandwidth. Fig. 4 illustrates the average bit rate per channel vs. the number of channels for  $\text{SNR}_{\text{avg}}^{S-R} = \text{SNR}_{\text{avg}}^{R-D} = 10\text{dB}$ . We observe that the gap between different schemes widens as the number of channels increase. This indicates that as more channels become available, they can be exploited more judiciously for pairing and power optimization.

### VIII. CONCLUSION

We have studied the problem of jointly optimal channel pairing, power allocation, and bit loading in a dual-hop multi-channel system. Despite the non-zero duality gap due to its mixed-integer nature, we show that the problem's special structure allows the extraction of an exact optimal solution. The proposed approach is applicable to general relaying strategies,

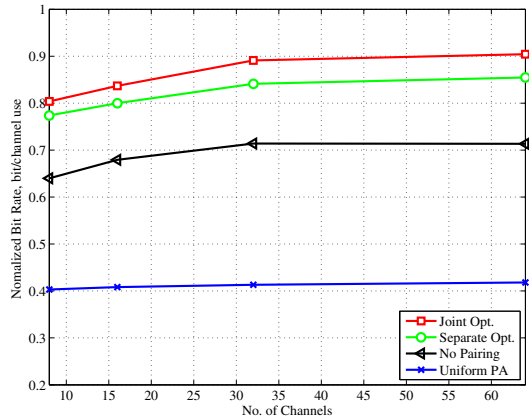


Fig. 4. Per channel rate vs. number of channels for multi-channel AF relaying with  $\text{SNR}_{\text{avg}}^{S-R} = \text{SNR}_{\text{avg}}^{R-D} = 10\text{dB}$ .

modulation schemes, and performance metrics. We have developed complexity reduction techniques and demonstrated that the algorithm runtime is polynomial in the number of channels. The efficiency of the proposed solution and its significant performance gain over suboptimal alternatives suggests that it is favorable in practical system design.

### APPENDIX A

#### PROOF OF LEMMA 1

Consider  $P_t^*(k) > 0$  for any  $k > 0$ . We only need to show that there exists  $P < P_t^*(k)$  such that  $B^*(P) = k - 1$ .

The required  $P$  can be found using the following procedure. Within the solution that renders  $P_t^*(k)$ , pick any path  $(m, n)$  such that  $\phi_{mn}^* = 1$  and  $b_{mn}^* > 0$ , which implies  $p_{mn}^* > 0$ . From (5) and the concavity of  $R(\cdot)$ , we see that there exists  $p_{mn} < p_{mn}^*$  such that the corresponding data rate  $b_{mn} = b_{mn}^* - 1$ . Let  $P = P_t^*(k) - p_{mn}^* + p_{mn}$ . Then  $P < P_t^*(k)$ , and the maximum sum rate given  $P$  is no worse than  $k - 1$ , i.e.,  $B^*(P) \geq k - 1$ . However, since  $P_t^*(k)$  is the minimum required power to achieve sum rate  $k$ , we also have  $B^*(P) < k$  for any  $P < P_t^*(k)$ . Therefore,  $B^*(P) = k - 1$ .

By taking the minimum of all  $P$  such that  $B^*(P) = k - 1$ , we obtain  $P_t^*(k - 1) < P_t^*(k)$ . With further induction, we see that  $B^*$  is a unit-step increasing staircase function of  $P_t$ , with integer-valued corner points at  $(P_t^*(k), k)$ .

### APPENDIX B

#### PROOF OF LEMMA 2

Given any  $\lambda \geq 0$ , consider the maximization problem (7) with  $P_t = P^{I*}(\lambda)$ . The power constraints (3) and (4) in this problem are automatically satisfied. Furthermore, the other constraints are the same as those in the maximization problem (11). Further comparing the objective functions, we see that maximizing the objective function of (11), given  $\lambda$  and  $P^{I*}(\lambda)$ , is exactly the same as maximizing the objective function of (7). Hence, we have

$$B^{I*}(\lambda) = B^*(P^{I*}(\lambda)), \quad (38)$$

i.e., the point  $(P^{I*}(\lambda), B^{I*}(\lambda))$  always falls on the curve  $B^*(P_t)$ .



Furthermore, for all  $\lambda > 0$ , if  $B'^*(\lambda) = B$  for some  $B \in \mathcal{M}$ , then in order to maximize the objective in (11),  $P'^*(\lambda)$  should be the minimum allowed sum power, subject to (1), (2), (3), (5), and  $\sum_{m,n} \phi_{mn}^*(\lambda) b_{mn}^*(\lambda) = B$ . This is of the same form as (8). Hence,  $P'^*(\lambda) = P_t^*(B)$ .

#### APPENDIX C PROOF OF LEMMA 3

From Lemma 2, we have  $P'^*(\lambda) = P_t^*(B'^*(\lambda))$ . Therefore, it is constant as long as  $B'^*(\lambda)$  remains constant. As  $\lambda$  increases, Proposition 1 dictates that  $B'^*(\lambda)$  decreases in unit steps. Each time  $B'^*(\lambda)$  steps down from  $k$  to  $k-1$ ,  $P'^*(\lambda)$  jumps from  $P_t^*(k)$  to  $P_t^*(k-1)$ . Finally, Lemma 1 suggests that  $P_t^*(k) - P_t^*(k-1) > 0$ .

#### APPENDIX D PROOF OF LEMMA 4

Consider  $\lambda^{*+}$  instead of  $\lambda^*$  to avoid ambiguity at the discontinuities of  $B'^*(\lambda)$  and  $P'^*(\lambda)$ . Let  $k = B'^*(\lambda^{*+})$ . Then, by (11) and Lemma 2, we have

$$k - \lambda^{*+} P_t^*(k) \geq (k-1) - \lambda^{*+} P_t^*(k-1). \quad (39)$$

This implies that

$$\begin{aligned} \lambda^* &\leq \frac{1}{P_t^*(k) - P_t^*(k-1)} \\ &\leq \frac{1}{P_t^*(1) - P_t^*(0)} = \frac{1}{P_t^*(1)} = \frac{1}{\sum_{m,n} \phi_{mn}^* p_{mn}^*}, \end{aligned} \quad (40)$$

where the second line is due to (9), and  $\phi_{mn}^*$  and  $p_{mn}^*$  denote the optimal solution to (8) that gives  $P_t^*(1)$ . Since the sum rate is 1, all power in  $P_t^*(1)$  must go to the one path  $(m', n')$  that supports  $b_{m'n'} = 1$ . It follows that

$$\lambda^* \leq \frac{1}{p_{m'n'}^*}. \quad (41)$$

Therefore,  $\lambda^*$  is upper bounded by the minimum power required to transmit at bit rate  $b_{mn} = 1$  for any path  $(m, n)$ , which we denote by  $\lambda_{max}$ . Since such minimum required power is generally determined by the modulation and coding schemes, it is independent of the number of channels  $N$ .

#### REFERENCES

- [1] "Long Term Evolution (LTE): A Technical Overview," Motorola Inc., Schaumburg, IL, Technical White Paper, 2010. [Online]. Available: <http://www.motorola.com/Business/US-EN/Business+Product+and+Services/LTE>
- [2] 3GPP TR 36.814 V1.2.1, "Further advancements for EUTRA: Physical layer aspects," Tech. Specification Group Radio Access Network Rel. 9, June 2009.
- [3] IEEE P802.16j/D9, "Draft amendment to IEEE standard for local and metropolitan area networks part 16: Air interface for fixed and mobile broadband wireless access systems: Multihop relay specification," Feb. 2009.
- [4] L. Vandendorpe, R. Duran, J. Louveaux, and A. Zaidi, "Power allocation for OFDM transmission with DF relaying," in *Proc. IEEE Int. Conf. Communications (ICC)*, May 2008.
- [5] I. Hammerstrom and A. Wittneben, "Power allocation schemes for amplify-and-forward MIMO-OFDM relay links," *IEEE Trans. Wireless Commun.*, vol. 6, no. 8, pp. 2798–2802, Aug. 2007.
- [6] W. Dang, M. Tao, H. Mu, and J. Huang, "Subcarrier-pair based resource allocation for cooperative multi-relay OFDM systems," *IEEE Trans. Wireless Commun.*, vol. 9, no. 5, pp. 1640–1649, May 2010.
- [7] M. Hajiaghayi, M. Dong, and B. Liang, "Jointly optimal channel and power assignment for dual-hop multi-channel multi-user relaying," to appear in the *IEEE J. Select. Areas Commun.*, Special Issue on Cooperative Networking - Challenges and Applications, 2012. [Online]. Available: <http://arxiv.org/abs/1102.5314>
- [8] L. Wolsey, *Integer programming*, ser. Wiley-Interscience series in discrete mathematics and optimization. Wiley, 1998.
- [9] W. Yu and R. Lui, "Dual methods for nonconvex spectrum optimization of multicarrier systems," *IEEE Trans. Commun.*, vol. 54, no. 7, pp. 1310–1322, Jul. 2006.
- [10] T. C.-Y. Ng and W. Yu, "Joint optimization of relay strategies and resource allocations in cooperative cellular networks," *IEEE J. Select. Areas Commun.*, vol. 25, no. 2, pp. 328–339, Feb. 2007.
- [11] D. Hughes-Hartogs, "Ensemble modem structure for imperfect transmission media," U.S. Pats. Nos. 4,679,227 (Jul. 1988) 4,731,816 (Mar. 1988) and 4,833,796 (May 1989).
- [12] P. Chow, J. Cioffi, and J. Bingham, "A practical discrete multitone transceiver loading algorithm for data transmission over spectrally shaped channels," *IEEE Trans. Commun.*, vol. 43, no. 234, pp. 773–775, Feb/Mar/Apr 1995.
- [13] A. Leke and J. Cioffi, "A maximum rate loading algorithm for discrete multitone modulation systems," in *Proc. IEEE Global Telecommun. Conf. (GLOBECOM)*, vol. 3, Nov. 1997.
- [14] J. Campello, "Practical bit loading for DMT," in *Proc. IEEE Int. Conf. Communications (ICC)*, vol. 2, Jun. 1999.
- [15] B. Krongold, K. Ramchandran, and D. Jones, "Computationally efficient optimal power allocation algorithms for multicarrier communication systems," *IEEE Trans. Commun.*, vol. 48, no. 1, pp. 23–27, Jan. 2000.
- [16] L. Piazza, "Fast optimal bit-loading algorithm for adaptive OFDM systems," Technical Report No. 002-04-0, University of Rome, Tech. Rep., Sep. 2003.
- [17] R. Fischer and J. Huber, "A new loading algorithm for discrete multitone transmission," in *Proc. IEEE Global Telecommun. Conf. (GLOBECOM)*, vol. 1, Nov. 1996.
- [18] A. Wyglinski, F. Labeau, and P. Kabal, "Bit loading with BER-constraint for multicarrier systems," *IEEE Trans. Commun.*, vol. 4, no. 4, pp. 1383–1387, Jul. 2005.
- [19] Y. Ma, N. Yi, and R. Tafazolli, "Bit and power loading for OFDM-based three-node relaying communications," *IEEE Trans. Signal Processing*, vol. 56, no. 7, pp. 3236–3247, Jul. 2008.
- [20] L. Chen and B. Krongold, "An efficient resource allocation algorithm for OFDMA-based multihop wireless transmission," in *Proc. IEEE Int. Symp. on Personal, Indoor and Mobile Radio Commun. (PIMRC)*, Sep. 2007.
- [21] A. Hottinen and T. Heikkinen, "Subchannel assignment in OFDM relay nodes," in *Proc. of Annual Conf. on Information Sciences and Systems (CISS)*, Princeton, NJ, Mar. 2006.
- [22] M. Herdin, "A chunk based OFDM amplify-and-forward relaying scheme for 4G mobile radio systems," in *Proc. IEEE Int. Conf. Communications (ICC)*, June 2006.
- [23] A. Hottinen and T. Heikkinen, "Optimal subchannel assignment in a two-hop OFDM relay," in *Proc. IEEE Workshop on Signal Processing Advances in Wireless Commun. (SPAWC)*, June 2007.
- [24] M. Dong, M. Hajiaghayi, and B. Liang, "On linear processing for dual-hop multi-channel relaying," in *Proc. of Annual Conf. on Information Sciences and Systems (CISS)*, Mar. 2011.
- [25] W. Wang, S. Yan, and S. Yang, "Optimally joint subcarrier matching and power allocation in OFDM multihop system," *EURASIP J. Appl. Signal Process. (USA)*, Mar. 2008.
- [26] W. Wang and R. Wu, "Capacity maximization for OFDM two-hop relay system with separate power constraints," *IEEE Trans. Veh. Technol.*, vol. 58, no. 9, pp. 4943–4954, Nov. 2009.
- [27] M. Hajiaghayi, M. Dong, and B. Liang, "Jointly optimal channel pairing and power allocation for multi-channel multi-hop relaying," *IEEE Trans. Signal Processing*, vol. 59, no. 10, pp. 4998–5012, Oct. 2011.
- [28] C.-N. Hsu, H.-J. Su, and P.-H. Lin, "Joint subcarrier pairing and power allocation for OFDM transmission with decode-and-forward relaying," *IEEE Trans. Signal Processing*, vol. 59, no. 99, pp. 399–414, Jan. 2011.
- [29] B. Gui and L. J. Cimini, Jr., "Bit loading algorithms for cooperative OFDM systems," *EURASIP J. Wirel. Commun. Netw.*, Jan. 2008.
- [30] N. Z. Shor, *Minimization Methods for Non-Differentiable Functions*. Springer-Verlag, 1985, translated by K. C. Kiwiel and A. Ruszczyński.
- [31] S. Boyd, L. Xiao, and A. Mutapcic, *Subgradient Methods [Lecture notes online available at <http://www.stanford.edu/class/ee392o/subgrad-method.pdf>]*, Stanford University, Oct. 2003.