Power Allocation for Cooperative Bit-Interleaved Coded Modulation Systems with Decode-Remap-and-Forward Relaying

Tsang-Wei Yu, Wern-Ho Sheen, *Member, IEEE*, and Chung-Hsuan Wang, *Member, IEEE*

***Abstract*—Cooperative bit-interleaved coded modulation (BICM) is a key technology for the next-generation wireless communication systems. This paper investigates power allocation for the cooperative BICM systems with decode-remap-and-forward (DRF) relaying. Unlike the conventional decode-and-forward relays, the DRF relays may choose different constellation mappings from that of source so as to obtain a remapping gain. In spite of its importance, the power allocation in such a system has not yet been explored in the literature. Two new power allocation methods are proposed, aiming to minimize bit error rate at the destination. The first uses a cost function based on the minimum weighted squared-Euclidean distance (called PA-MWSED) and is optimized with the sub-gradient method. The second is based on a generalized MWSED (called PA-GMWSED) and is optimized with the Simplex method after the optimization is re-cast as a linear programming problem. Generally speaking, PA-MWSED has a better performance than PA-GMWSED but requires a higher complexity. Numerical results show that both of the proposed methods outperform the equal gain power allocation with large margins.**

***Index Terms*—Bit-interleaved coded modulation, cooperative relaying, decode-remap-and-forward, power allocation.**

I. Introduction

RECENTLY, the cooperative relaying technique has drawn an increasing interest in wireless communications for gaining space diversity yet without using multiple physical antennas [1]-[5]. The basic idea is to allow other nodes, called relays, to forward the information of source to destination in a cooperative way [1][2]. Basically, the relay can simply amplify the received signal and forward it to the destination, or the relay can decode the received signal and re-encode before forwarding to the destination. The former is called the amplify-and-forward (AF) relaying mode and the latter the decode-and-forward (DF) relaying mode. Each mode has its merit over the other under different system setups and/or channel conditions [1]-[5]. Furthermore, different types of hybrid modes that are able to adapt to channel conditions have been proposed in the literature to improve system performance [5]-[7].

Manuscript received February 18, 2011; revised July 7 and November 21, 2011; accepted January 17, 2012. The associate editor coordinating the review of this paper and approving it for publication was S. Affes.

T. W. Yu and C. H. Wang are with the Department of Electrical Engineering, National Chiao Tung University, Hsinchu 300, Taiwan, R.O.C. (e-mail: toshiba.cm93g@nctu.edu.tw, chwang@mail.nctu.edu.tw).

W. H. Sheen is with the Department of Information and Communication Engineering, Chaoyang University of Technology, Wufong, Taichung 41349, Taiwan (e-mail: whsheen@cyut.edu.tw).

Digital Object Identifier 10.1109/TWC.2012.031212.110309

Bit-interleaved coded modulation (BICM) [8][9] is a bandwidth-efficient technique for fading channels and has been used widely in real systems [10][11]. Very recently, studies have begun to look at the BICM systems with cooperative techniques [12]-[16][26]. In particular, bit error rate (BER) performance of a cooperative BICM system was analyzed in [12] with the DF relay being modeled by a post-BSC (binary symmetric channel). Reference [13] derived the achievable rates for different combinations of modulation and number of antennas used at the source and relay nodes. Pre-coding for a non-orthogonal AF cooperative BICM system was studied in [14] aiming to achieve maximum diversity order and high coding gains. For the cooperative BICM-OFDM (orthogonal frequency-division multiplexing) systems, [15] considered the relay selection and sub-carrier allocation with AF relaying in order to minimize the asymptotic worst-case, pair-wise error probability, whereas [16] considered the issue of relay placement with DF relaying.

In this paper, we investigate power allocation for a cooperative BICM system with decode-remap-and-forward (DRF) relaying, where relays do the forwarding only if the received packet is decoded correctly and are allowed to use different constellation mappings from that of source so as to obtain a remapping gain [17]-[19]. Two power allocation methods are proposed aiming to minimize BER at the destination based on perfect knowledge of channel state information (CSI). To the best of our knowledge, power allocation in such a system has not yet been discussed in the literature[1].

Theoretically, the optimal power allocation for the considered cooperative BICM systems can be achieved by first obtaining the exact BER at the destination, followed by an exhaustive search of the optimal power allocation. Unfortunately, such an approach is not practical because the exact BER is very difficult to obtain if not impossible for a coded system, and all possible allocations have to be examined in the exhaustive search. In this paper, two low-complexity power allocation methods are proposed. The first employs a cost function based on the minimum weighted squared-Euclidean distance (MWSED), which is shown to be convex and can be optimized with the sub-gradient method [20]. The second

[1]Joint optimization of power and mappers is a very complicated problem in the considered cooperative BICM systems. In fact, the mapping design itself is very cumbersome and lacks of analytical solutions. This paper mainly focuses on the power allocation methods that are applicable to systems with different mappers at source and relays, although the joint optimization is a very important topic and deserves a full investigation.

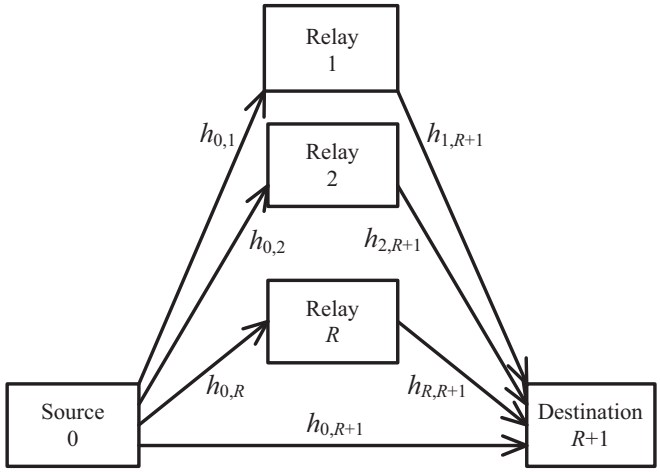

Fig. 1. A cooperative relaying network with one source, one destination and R relays.

employs a cost function based on a generalized MWSED (GMWSED) to further reduce complexity; the optimization is done with the Simplex method [21] after the problem is re-cast as a linear programming problem. Simulation results show that the proposed methods significantly outperform the equal gain power allocation. Furthermore, it was shown in [26] that the cooperative BICM system with DRF relaying achieves full diversity order.

The remainder of this paper is organized as follows. Section II describes the system model, and Section III analyzes BER at the destination. In Section IV, the proposed power allocation methods are presented. Simulation results and complexity comparisons are given in Section V, followed by conclusions in Section VI.

II. System Models

As is shown in Fig. 1, we consider a cooperative relaying network with one source, R relays and one destination, which are indexed by 0, 1, ..., R and $R+1$, respectively. Each node is equipped with one antenna and operates in the half-duplexing manner, implying that it cannot transmit and receive simultaneously. The channel gain between nodes i and j is denoted by $h_{i,j}$ which, in a block fading environment, remains unchanged over the transmission of a packet. Perfect CSI of $h_{0,j}$ is assumed to be known at the relay j, and perfect $h_{0,R+1}$, $h_{0,j}$ and $h_{j,R+1}$, $j = 1, 2, \cdots, R$ are known at the destination and the power allocation unit. In addition, since all links remain unchanged over the transmission of a packet, they can be treated as AWGN (additive white Gaussian noise) channels from the power allocation perspective. Note that power allocation can be done either at source or destination depending on the required signaling overhead and where the complexity of power allocation is to be placed. From the signaling overhead aspect, allocation at destination seems favorable because only the CSIs of the source-to-relay links have to be reported to the destination.

In the considered DRF relaying, transmission of a packet is divided into two phases. At phase-I, the source broadcasts the packet to all relays and the destination. If a relay decodes correctly, aided by a cyclic redundancy check (CRC), it re-encodes, remaps and forwards the packet to the destination at phase-II; otherwise, the relay keeps silent. The transmissions of relays are through orthogonal channels with the transmit power determined by the power allocation unit. At the destination, the received signals from phase-I and phase-II are combined and decoded jointly.

BICM is employed at all nodes. At the source, an information bit sequence $\mathbf{b}$ of length K is encoded by a channel encoder to a coded sequence $\mathbf{c}$ of length N and then interleaved by an interleaver π. The resulting binary sequence $\mathbf{v} = \pi(\mathbf{c})$ is successively partitioned into groups of m bits, called the labels. The k-th label of $\mathbf{v}$, denoted by v_k, is then mapped to a complex symbol $x_k^{(0)}$ in the constellation χ, according to the constellation mapper $\mu^{(0)}$, i.e., $x_k^{(0)} = \mu^{(0)}(v_k)$, where $|\chi| = 2^m$, and $|\chi|$ is the cardinality of χ. For notation simplicity, the time index k will be omitted in the following development.

The received signals at relays and the destination at phase-I are given by

$$y_{0,j} = h_{0,j}\sqrt{P_0}x^{(0)} + \omega_{0,j},\ j = 1, 2, \cdots, R+1, \tag{1}$$

where P_0 is the source transmit power, and $\omega_{0,j}$ is AWGN at node j. The noises at relays and destination are modeled as i.i.d. (independent and identically distributed) zero-mean, circularly-symmetric complex Gaussian random variables with variance $N_0/2$ per dimension. Upon receiving $y_{0,j}$, the demapper of relay j evaluates the simplified log-likelihood ratio (LLR) of the i-th bit of the considered label v, according to

$$\log \frac{\max\limits_{v\in\Gamma_1^i} p(y_{0,j}|v)}{\max\limits_{v\in\Gamma_0^i} p(y_{0,j}|v)} = \min_{v\in\Gamma_0^i} \frac{|y_{0,j} - h_{0,j}\sqrt{P_0}\mu^{(0)}(v)|^2}{N_0} - \min_{v\in\Gamma_1^i} \frac{|y_{0,j} - h_{0,j}\sqrt{P_0}\mu^{(0)}(v)|^2}{N_0}, \tag{2}$$

where Γ_b^i is the set of labels with the binary value b at the i-th position. The LLRs are then de-interleaved and passed to the decoder. If $\mathbf{b}$ is decoded correctly at the relay, $\mathbf{b}$ is re-encoded and re-interleaved by the same encoder and interleaver as those of source, resulting the label sequence $\mathbf{v}$. A particular v of $\mathbf{v}$ is then mapped to $x^{(j)} \in \chi$ by the mapper $\mu^{(j)}$, which may or may not be the same as $\mu^{(0)}$.

Denote Ω the set of relays which have succeeded in decoding (called active relays). The received signals at destination at phase-II can then be expressed as

$$y_{j,R+1} = h_{j,R+1}\sqrt{P_j}x^{(j)} + \omega_{j,R+1}, j \in \Omega, \tag{3}$$

where P_j is the transmit power of relay j, and $\omega_{j,R+1}$ is AWGN at destination. After combining the signals received at phase-I and II, the LLR for the i-th bit of the considered

label v is evaluated at destination by

$$\log \frac{\max\limits_{v\in\Gamma_1^i}\prod\limits_{j\in\widetilde{\Omega}} p(y_{j,R+1}|v)}{\max\limits_{v\in\Gamma_0^i}\prod\limits_{j\in\widetilde{\Omega}} p(y_{j,R+1}|v)}$$
$$= \min_{v\in\Gamma_0^i}\sum_{j\in\widetilde{\Omega}}\frac{\left|y_{j,R+1}-h_{j,R+1}\sqrt{P_j}\mu^{(j)}(v)\right|^2}{N_0} - \min_{v\in\Gamma_1^i}\sum_{j\in\widetilde{\Omega}}\frac{\left|y_{j,R+1}-h_{j,R+1}\sqrt{P_j}\mu^{(j)}(v)\right|^2}{N_0}, \quad (4)$$

where $\widetilde{\Omega}=\Omega\cup\{0\}$ is the set of all active nodes including the source. The LLRs of the coded sequence are then de-interleaved and passed to the decoder.

III. Bit Error Rate at Destination

The objective of this work is to determine the power allocation $\mathbf{P}=[P_0\ P_1\cdots P_R]^T$ that minimizes BER at the destination under the power constraint of $\sum_{j=0}^{R}P_j\le P_T$, where $[\cdot]^T$ denotes taking transpose of a vector. In this section, the BER performance is analyzed, and the new power allocation methods will be discussed in the next section.

Let $p_{b,j}$ and $p_{f,j}$ denote BER and PER (packet error rate) at node j, respectively. Then, BER at the destination $p_{b,R+1}$ is evaluated by

$$P_{b,R+1} = \sum_{\Omega\subseteq\{1,2,\cdots,R\}} p_{b,R+1}(\Omega)\left[\prod_{j\in\Omega}(1-p_{f,j})\right]\left[\prod_{j\notin\Omega}p_{f,j}\right] \approx \sum_{\Omega\subseteq\{1,2,\cdots,R\}} p_{b,R+1}(\Omega)\prod_{j\notin\Omega}K\cdot p_{b,j}, \quad (5)$$

where $p_{b,R+1}(\Omega)$ is BER at destination given the active set Ω, and $j=1,2,\cdots,R$. In (5), the approximation is obtained with the assumptions that $p_{f,j}\ll 1$ in practical systems and that $p_{f,j}\approx K\cdot p_{b,j}$ for a small $p_{b,j}$.

With the assumptions of ideal interleaving (an interleaver with infinite depth) and symmetrization as in [9], a BICM can be regarded as a linear code, and $p_{b,R+1}(\Omega)$ is upper-bounded by [9],

$$p_{b,R+1}(\Omega)\le\sum_{d_h=d_f}^{N} W_I(d_h)f_{ub}(d_h,\mathbf{P},\Omega), \quad (6)$$

where $W_I(d_h)$ denotes the total input weight of error events with Hamming weight d_h divided by K, d_f is the free distance of the code, and $f_{ub}(d_h,\mathbf{P},\Omega)$ is an upper bound of the pair-wise error probability (PEP) between two coded sequences with Hamming distance d_h, given by

$$f_{ub}(d_h,\mathbf{P},\Omega)=\frac{1}{2\pi\mathbf{j}}\int_{s_0-\mathbf{j}\infty}^{s_0+\mathbf{j}\infty}\left[\Phi_{\mathbf{P},\Omega}(s)\right]^{d_h}\frac{ds}{s}. \quad (7)$$

In (7),

$$\Phi_{\mathbf{P},\Omega}(s)=\frac{1}{m2^m}\sum_{i=1}^{m}\sum_{b=0}^{1}\sum_{v\in\Gamma_b^i}\sum_{w\in\Gamma_{\bar{b}}^i}\prod_{j\in\widetilde{\Omega}}\Phi_{\Delta(x^{(j)},z^{(j)})}(s), \quad (8)$$

$\mathbf{j}=\sqrt{-1}$, and $\Phi_{\Delta(x^{(j)},z^{(j)})}(s)$ is the moment generating function of

$$\Delta(x^{(j)},z^{(j)})=\log p(y_{j,R+1}|x^{(j)})-\log p(y_{j,R+1}|z^{(j)}) \quad (9)$$

with $x^{(j)}=\mu^{(j)}(v)$ and $z^{(j)}=\mu^{(j)}(w)$. Similar upper-bound on $p_{b,j}$ for $j=1,2,\cdots,R$ can be obtained by applying the same procedure.

IV. New Power Allocation Methods

In this section, two low-complexity power allocation methods are proposed: one is based on the minimum weighted squared-Euclidean distance (PA-MWSED), and the other on the generalized MWSED (PA-GMWSED). Generally speaking, PA-MWSED has a better performance than PA-GMWSED but requires a higher complexity. Both of the proposed methods outperform the equal gain power allocation (PA-EG) with large margins, as is to be shown in Section V.

A. PA-MWSED

At high signal-to-noise ratios (SNRs), the summation in (6) is dominated by the error events that have the smallest Hamming weight, i.e., the free distance d_f. Thus, (6) can be approximated by

$$p_{b,R+1}(\Omega)\approx W_I(d_f)f_{ub}(d_f,\boldsymbol{\alpha},\Omega), \quad (10)$$

where $\mathbf{P}$ is replaced by $\boldsymbol{\alpha}=[\alpha_0\ \alpha_1\cdots\alpha_R]^T$ with $\alpha_j=P_j/P_T$.

The inverse Laplace transform in (7) can be evaluated efficiently along the vertical line of the saddle point [22]. In addition, given the channels, $\Delta(x^{(j)},z^{(j)})$ is a Gaussian random variable with the moment generating function

$$\Phi_{\Delta(x^{(j)},z^{(j)})}(s)=\exp\left[\frac{-s+s^2}{N_0}P_T\alpha_j|h_{j,R+1}|^2D^{(j)}\right], \quad (11)$$

where $D^{(j)}=|x^{(j)}-z^{(j)}|^2$ is the squared-Euclidean distance between $x^{(j)}$ and $z^{(j)}$. Note that $-s+s^2=(s-0.5)^2-0.25$, and the saddle point of (11) always occurs at $s=0.5$, regardless of the values of P_T, α_j, $|h_{j,R+1}|^2$ and $D^{(j)}$. Substituting $s=0.5+\mathbf{j}t$ into (11) and (8), we have

$$\Phi_{\boldsymbol{\alpha},\Omega}(t) = \frac{1}{m2^m}\sum_{i=1}^{m}\sum_{b=0}^{1}\sum_{v\in\Gamma_b^i}\sum_{w\in\Gamma_{\bar{b}}^i}\exp\left[-\frac{t^2+1/4}{N_0}P_T\sum_{j\in\widetilde{\Omega}}\alpha_j|h_{j,R+1}|^2D^{(j)}\right]. \quad (12)$$

Furthermore, at high SNRs, (12) can be simplified by just considering the (v,w) pairs which achieve MWSED, i.e.,

$$\Phi_{\boldsymbol{\alpha},\Omega}(t)\approx\frac{N_\Omega(\boldsymbol{\alpha})}{m2^m}\exp\left[-\frac{t^2+1/4}{N_0}P_TM_\Omega(\boldsymbol{\alpha})\right], \quad (13)$$

where $M_\Omega(\boldsymbol{\alpha})$ is MWSED, given by

$$M_\Omega(\boldsymbol{\alpha}) = \min_{v\in\Gamma_b^i,w\in\Gamma_{\bar{b}}^i,i=1,\cdots,m,\ b=0,1}\sum_{j\in\widetilde{\Omega}}\alpha_j|h_{j,R+1}|^2D^{(j)} = \min_{\mathbf{D}\in\Psi_{PA-MWSED}}\sum_{j\in\widetilde{\Omega}}\alpha_j|h_{j,R+1}|^2D^{(j)}, \quad (14)$$

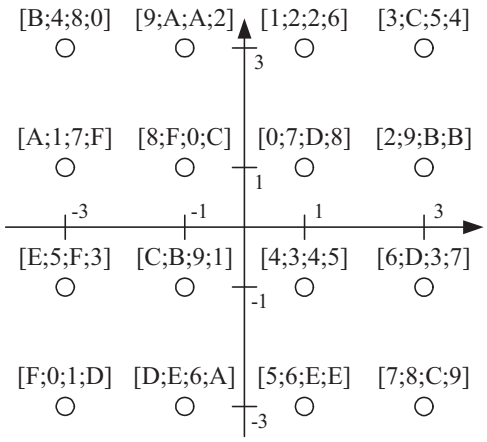

Fig. 2. Four example mappers μ_G, μ_A, μ_B and μ_C for the 16-QAM constellation. The signal point labels $[v_G; v_A; v_B; v_C]$ are in hexadecimal format, where v_G, v_A, v_B and v_C are to denote the label of μ_G, μ_A, μ_B and μ_C, respectively. (The μ_A, μ_B and μ_C mappers are the MBER mappings which maximize the minimum Euclidean distance between transmit symbols for the second, third and forth transmissions of the hybrid automatic repeat-request system in [23], respectively.)

$\Psi_{PA-MWSED}$ is the set of all distinct $\mathbf{D} = [D^{(0)}\ D^{(1)} \cdots D^{(R)}]^T$ obtained by exhausting all possible (v, w) pairs, and $N_\Omega(\boldsymbol{\alpha})$ is the number of (v, w) pairs which achieve MWSED. Note that the numbers of (v, w) pairs for different $\mathbf{D}$'s in $\Psi_{PA-MWSED}$ are usually not the same so that $N_\Omega(\boldsymbol{\alpha})$ may not be a continuous function of $\boldsymbol{\alpha}$. Table I enumerates $\Psi_{PA-MWSED}$ for the following setups.

- Setup-1: $R = 1$, 16-QAM, and $\mu^{(0)} = \mu^{(1)} = \mu_G$ (the Gray mapping) in Fig. 2. Note that all elements in $\mathbf{D}$ are same if $\mu^{(j)} = \mu^{(0)}$ for all j.
- Setup-2: $R = 1$, 16-QAM, $\mu^{(0)} = \mu_G$ and $\mu^{(1)} = \mu_A$ in Fig. 2.
- Setup-3: $R = 2$, 16-QAM, $\mu^{(0)} = \mu_G$, $\mu^{(1)} = \mu_A$ and $\mu^{(2)} = \mu_B$ in Fig. 2.

Using (13), we have

$$\begin{aligned} & f_{ub}(d_f, \boldsymbol{\alpha}, \Omega) \\ &\approx \frac{1}{2\pi}\int_{-\infty}^{\infty}\left[\frac{N_\Omega(\boldsymbol{\alpha})}{m2^m}\exp\left[-\frac{t^2+1/4}{N_0}P_T M_\Omega(\boldsymbol{\alpha})\right]\right]^{d_f}\frac{dt}{1/2+\mathbf{j}t} \\ &= \frac{1}{4\pi}\left(\frac{N_\Omega(\boldsymbol{\alpha})}{m2^m}\right)^{d_f}\exp\left[-\frac{d_f P_T}{4N_0}M_\Omega(\boldsymbol{\alpha})\right] \\ &\quad \cdot\int_{-\infty}^{\infty}\exp\left[-\frac{d_f P_T}{N_0}M_\Omega(\boldsymbol{\alpha})t^2\right]\frac{dt}{1/4+t^2}, \end{aligned} \tag{15}$$

where the equality is obtained by using the fact that the imaginary part in the integral is an odd function of t. Furthermore, at high SNRs, i.e., $P_T/N_0 \to \infty$, the integral in (15) is dominated by the value integrated over a small interval $[-\tau, \tau]$, where $0 < \tau \ll 1$. Thus, (15) can be approximated by

$$\begin{aligned} f_{ub}(d_f, \boldsymbol{\alpha}, \Omega) \approx\ & \frac{1}{\pi}\left(\frac{N_\Omega(\boldsymbol{\alpha})}{m2^m}\right)^{d_f}\exp\left[-\frac{d_f P_T}{4N_0}M_\Omega(\boldsymbol{\alpha})\right] \\ & \cdot\int_{-\infty}^{\infty}\exp\left[-\frac{d_f P_T}{N_0}M_\Omega(\boldsymbol{\alpha})t^2\right]dt \\ =\ & \left(\frac{N_\Omega(\boldsymbol{\alpha})}{m2^m}\right)^{d_f}\sqrt{\frac{N_0}{\pi d_f P_T M_\Omega(\boldsymbol{\alpha})}} \\ & \cdot\exp\left[-\frac{d_f P_T}{4N_0}M_\Omega(\boldsymbol{\alpha})\right], \end{aligned} \tag{16}$$

where the equality is obtained by re-arranging the term in the integral into the form of a Gaussian probability density function. Applying similar steps in (10)-(16), we can obtain similar results for $p_{b,j}$, $j = 1, 2, \cdots, R$ as follows. Since relay j receives the signal from the source only, the corresponding MWSED is degenerated to $\alpha_0|h_{0,j}|^2 D_\chi$, where

$$D_\chi = \min_{v\in\Gamma_b^i, w\in\Gamma_b^i, i=1,\cdots,m,\ b=0,1}|\mu^{(0)}(v) - \mu^{(0)}(w)|^2, \tag{17}$$

and consequently

$$\begin{aligned} p_{b,j} \approx\ & W_I(d_f)\left(\frac{N_\emptyset(\boldsymbol{\alpha})}{m2^m}\right)^{d_f}\sqrt{\frac{N_0}{\pi d_f P_T \alpha_0|h_{0,j}|^2 D_\chi}} \\ & \cdot\exp\left[-\frac{d_f P_T}{4N_0}\alpha_0|h_{0,j}|^2 D_\chi\right]. \end{aligned} \tag{18}$$

Now, we can approximate (5) by using (10), (16) and (18).

As to power allocation, we note that (i) $N_\Omega(\boldsymbol{\alpha})$'s may not be continuous functions of $\boldsymbol{\alpha}$, and (ii) at high SNRs $N_\Omega(\boldsymbol{\alpha})$ and $N_\emptyset(\boldsymbol{\alpha})$ (the numbers of pairs) are less dominant than $M_\Omega(\boldsymbol{\alpha})$ and $\alpha_0|h_{0,j}|^2 D_\chi$ (the distances) in (16) and (18), respectively. Therefore, for simplicity, the effects of $N_\Omega(\boldsymbol{\alpha})$ and $N_\emptyset(\boldsymbol{\alpha})$ are neglected by replacing them with $m2^m$. Through above steps, (5) can be approximated as

$$p_{b,R+1} \approx W_I(d_f)G(\boldsymbol{\alpha}), \tag{19}$$

where

$$G(\boldsymbol{\alpha}) = \sum_{\Omega\subseteq\{1,2,\cdots,R\}} g_\Omega \prod_{j\notin\tilde{\Omega}} K W_I(d_f)\tilde{g}_j, \tag{20}$$

$$g_\Omega = \sqrt{\frac{N_0}{\pi d_f P_T M_\Omega(\boldsymbol{\alpha})}}\exp\left[-\frac{d_f P_T}{4N_0}M_\Omega(\boldsymbol{\alpha})\right], \tag{21}$$

$$\tilde{g}_j = \sqrt{\frac{N_0}{\pi d_f P_T \alpha_0|h_{0,j}|^2 D_\chi}}\exp\left[-\frac{d_f P_T}{4N_0}\alpha_0|h_{0,j}|^2 D_\chi\right]. \tag{22}$$

Now, we introduce our first power allocation method called PA-MWSED as follow: given the channel realizations and the mappers $\mu^{(0)}, \mu^{(1)}, \cdots, \mu^{(R)}$,

$$\begin{aligned} & \boldsymbol{\alpha}_{PA-MWSED} = \arg\min_{\boldsymbol{\alpha}} G(\boldsymbol{\alpha}), \\ & \text{s.t.} \qquad \mathbf{1}^T\boldsymbol{\alpha} \le 1,\ 0 \le \alpha_j \le 1,\ j = 0, 1, \cdots, R, \end{aligned} \tag{23}$$

where $\mathbf{1}$ is the all-one vector of dimension $R + 1$. Note that the feasible set of $\boldsymbol{\alpha}$ is convex. In addition, as is proved in Appendix A, $G(\boldsymbol{\alpha})$ is a convex function of $\boldsymbol{\alpha}$, but may not be differentiable because $M_\Omega(\boldsymbol{\alpha})$ is the minimum of linear functions of $\boldsymbol{\alpha}$ and may not be differentiable. Such a constrained convex optimization problem with a non-differentiable cost

TABLE I
$\Psi_{PA-MWSED}$ FOR SETUP-1, SETUP-2 AND SETUP-3

Setup	$\Psi_{PA-MWSED}$
Setup-1	$[4\ 4]^T$, $[8\ 8]^T$, $[16\ 16]^T$, $[20\ 20]^T$, $[32\ 32]^T$, $[36\ 36]^T$, $[40\ 40]^T$, $[52\ 52]^T$, $[72\ 72]^T$
Setup-2	$[4\ 16]^T$, $[4\ 20]^T$, $[4\ 36]^T$, $[4\ 52]^T$, $[8\ 16]^T$, $[8\ 20]^T$, $[8\ 52]^T$, $[8\ 72]^T$, $[16\ 4]^T$, $[16\ 8]^T$, $[16\ 40]^T$, $[20\ 4]^T$, $[20\ 8]^T$, $[20\ 32]^T$, $[20\ 40]^T$, $[32\ 20]^T$, $[36\ 4]^T$, $[40\ 16]^T$, $[40\ 20]^T$, $[40\ 40]^T$, $[52\ 4]^T$, $[52\ 8]^T$, $[72\ 8]^T$
Setup-3	$[4\ 16\ 20]^T$, $[4\ 20\ 20]^T$, $[4\ 20\ 40]^T$, $[4\ 36\ 8]^T$, $[4\ 36\ 40]^T$, $[4\ 36\ 72]^T$, $[4\ 52\ 4]^T$, $[4\ 52\ 8]^T$, $[8\ 16\ 16]^T$, $[8\ 16\ 52]^T$, $[8\ 20\ 16]^T$, $[8\ 20\ 20]^T$, $[8\ 20\ 40]^T$, $[8\ 20\ 52]^T$, $[8\ 52\ 4]^T$, $[8\ 52\ 36]^T$, $[8\ 72\ 32]^T$, $[16\ 4\ 20]^T$, $[16\ 8\ 16]^T$, $[16\ 8\ 52]^T$, $[16\ 40\ 4]^T$, $[16\ 40\ 16]^T$, $[20\ 4\ 20]^T$, $[20\ 4\ 40]^T$, $[20\ 8\ 16]^T$, $[20\ 8\ 20]^T$, $[20\ 8\ 40]^T$, $[20\ 8\ 52]^T$, $[20\ 32\ 4]^T$, $[20\ 32\ 8]^T$, $[20\ 32\ 16]^T$, $[20\ 40\ 4]^T$, $[20\ 40\ 36]^T$, $[32\ 20\ 4]^T$, $[32\ 20\ 8]^T$, $[32\ 20\ 16]^T$, $[36\ 4\ 4]^T$, $[36\ 4\ 8]^T$, $[36\ 4\ 40]^T$, $[36\ 4\ 72]^T$, $[40\ 16\ 4]^T$, $[40\ 16\ 16]^T$, $[40\ 20\ 8]^T$, $[40\ 20\ 32]^T$, $[40\ 40\ 32]^T$, $[40\ 40\ 36]^T$, $[52\ 4\ 4]^T$, $[52\ 4\ 8]^T$, $[52\ 8\ 8]^T$, $[52\ 8\ 32]^T$, $[72\ 8\ 36]^T$

function can be solved by applying the projected sub-gradient method [20].

Define the sub-gradient of $G(\boldsymbol{\alpha})$ by

$$\mathbf{G}'(\boldsymbol{\alpha}) = \left[\frac{\partial G(\boldsymbol{\alpha})}{\partial \alpha_0} \ \frac{\partial G(\boldsymbol{\alpha})}{\partial \alpha_1} \ \cdots \ \frac{\partial G(\boldsymbol{\alpha})}{\partial \alpha_R} \ \right]^T, \tag{24}$$

where $\partial G(\boldsymbol{\alpha})/\partial \alpha_j$ can be evaluated numerically by using the following difference quotient with a sufficiently small δ, i.e.,

$$\frac{\partial G(\boldsymbol{\alpha})}{\partial \alpha_j} \approx \frac{G(\boldsymbol{\alpha}) - G(\boldsymbol{\alpha} + \boldsymbol{\delta}_j)}{\delta}, \tag{25}$$

with $\boldsymbol{\delta}_j = [\delta_0\ \delta_1\ \cdots\ \delta_R\ ]^T$ and the i-th entry of $\boldsymbol{\delta}_j$ defined by

$$\delta_i = \begin{cases} \delta, & \text{if } i = j \\ 0, & \text{otherwise} \end{cases}.$$

The projected sub-gradient method iterates

$$\boldsymbol{\alpha}^{(t)} = \mathbb{P}\left(\boldsymbol{\alpha}^{(t-1)} - \varepsilon_t \mathbf{G}'(\boldsymbol{\alpha}^{(t-1)}) \right), \tag{26}$$

where ε_t is the step size at the t-th iteration and $\mathbb{P}$ is the Euclidean projection on the feasible set of $\boldsymbol{\alpha}$. Some example performances of PA-MWSED will be provided in Section V.

B. PA-GMWSED

In PA-MWSED, $G(\boldsymbol{\alpha})$ needs to be evaluated repeatedly (according to (24) and (25)) in each iteration when the sub-gradient method is carried out, and there are total of 2^R Ω's (each has different $g_\Omega \prod_{j \notin \widetilde{\Omega}} KW_I(d_f)\tilde{g}_j$) needed to be included in the evaluation of $G(\boldsymbol{\alpha})$ in (20). As a consequence, PA-MWSED can be quite complex in some cases. In this sub-section, the method of PA-GMWSED is proposed to further reduce the complexity of power allocation. The performance and complexity of PA-MWSED and PA-GMWSED will be compared in Section V for some examples.

The first idea in the development of PA-GMWSED is that instead of minimizing $G(\boldsymbol{\alpha})$ in (20), an $\boldsymbol{\alpha}$ is searched to minimize $\max_{\Omega \subseteq \{1,2,\cdots,R\}} g_\Omega \prod_{j \notin \widetilde{\Omega}} KW_I(d_f)\tilde{g}_j$, the largest term in the summation of (20). Equivalently, an $\boldsymbol{\alpha}$ is searched to minimize

$$\max_{\Omega \subseteq \{1,2,\cdots,R\}} \left\{ \ln A_\Omega(\boldsymbol{\alpha}) + \ln B_\Omega(\boldsymbol{\alpha}) \right\}, \tag{27}$$

where

$$\ln A_\Omega(\boldsymbol{\alpha}) = -\frac{1}{2} \ln \left[M_\Omega(\boldsymbol{\alpha}) \prod_{j \notin \widetilde{\Omega}} \alpha_0 |h_{0,j}|^2 D_\chi \right], \tag{28}$$

and

$$\ln B_\Omega(\boldsymbol{\alpha}) = \ln \left(KW_I(d_f) \sqrt{\frac{N_0}{\pi d_f P_T}} \right)^{R-|\Omega|} - \frac{d_f P_T}{4N_0} \left[M_\Omega(\boldsymbol{\alpha}) + \alpha_0 \sum_{j \notin \widetilde{\Omega}} |h_{0,j}|^2 D_\chi \right]. \tag{29}$$

Note that (i) $\ln A_\Omega(\boldsymbol{\alpha})$ does not change with P_T/N_0, (ii) $|\ln B_\Omega(\boldsymbol{\alpha})| \to \infty$ as $P_T/N_0 \to \infty$, and (iii) $|\ln A_\Omega(\boldsymbol{\alpha})| \to \infty$ as $\alpha_0 \to 0$ for $|\Omega| < R$. Using a very small α_0, however, is impractical because the source power will be too small to activate any relay in this case. Therefore, in practical systems, $\ln A_\Omega(\boldsymbol{\alpha}) + \ln B_\Omega(\boldsymbol{\alpha})$ is dominated by $\ln B_\Omega(\boldsymbol{\alpha})$ at high SNRs, and it can be employed in search for good $\boldsymbol{\alpha}$ with a reduced complexity.

In the proposed PA-GMWSED, the optimization problem is cast as follows.

$$\begin{aligned} &\boldsymbol{\alpha}_{PA-GMWSED} \\ &= \operatorname*{argmin}_{\boldsymbol{\alpha}} \max_{\Omega \subseteq \{1,2,\cdots,R\}} \ln B_\Omega(\boldsymbol{\alpha}) \\ &= \operatorname*{argmax}_{\boldsymbol{\alpha}} \min_{\Omega \subseteq \{1,2,\cdots,R\}} \left\{ M_\Omega(\boldsymbol{\alpha}) + \alpha_0 \sum_{j \notin \widetilde{\Omega}} |h_{0,j}|^2 D_\chi - \eta_{R-|\Omega|} \right\}, \end{aligned} \tag{30}$$

where

$$\eta_{R-|\Omega|} = \frac{4N_0}{d_f P_T} \ln \left(KW_I(d_f) \sqrt{\frac{N_0}{\pi d_f P_T}} \right)^{R-|\Omega|}. \tag{31}$$

In (30), $M_\Omega(\boldsymbol{\alpha})$ is the MWSED contributed by active nodes, $\alpha_0 \sum_{j \notin \widetilde{\Omega}} |h_{0,j}|^2 D_\chi$ accounts for the effect of inactive relays, and $\eta_{R-|\Omega|}$, which is independent to $\boldsymbol{\alpha}$, is a term relating to the channel coding $(K, d_f, W_I(d_f))$, the number of inactive relays $(R - |\Omega|)$ and the SNR (P_T/N_0). Note that $\eta_{R-|\Omega|} = 0$ when $P_T/N_0 \to \infty$ or $|\Omega| = R$.

Using $M_\Omega(\boldsymbol{\alpha})$ in (14), we have

$$\begin{aligned} & M_\Omega(\boldsymbol{\alpha}) + \alpha_0 \sum_{j\notin\widetilde{\Omega}} |h_{0,j}|^2 D_\chi - \eta_{R-|\Omega|} \\ = & \min_{\mathbf{D}\in\Psi_{PA-MWSED}} \sum_{j\in\widetilde{\Omega}} \alpha_j |h_{j,R+1}|^2 D^{(j)} \\ & + \alpha_0 \sum_{j\notin\widetilde{\Omega}} |h_{0,j}|^2 D_\chi - \eta_{R-|\Omega|} \sum_{j=0}^{R} \alpha_j \\ = & \min_{\bar{\mathbf{D}}\in\bar{\Psi}_\Omega} \bar{\mathbf{D}}^T \boldsymbol{\alpha}, \end{aligned} \tag{32}$$

where

$$\bar{\Psi}_\Omega = \left\{ \bar{\mathbf{D}} = [\bar{D}^{(0)}\ \bar{D}^{(1)}\ \cdots\ \bar{D}^{(R)}]^T | \mathbf{D}\in\Psi_{PA-MWSED} \right\}, \tag{33}$$

and

$$\bar{D}^{(j)} = \begin{cases} |h_{j,R+1}|^2 D^{(j)} + D_\chi \sum\limits_{j_1\notin\widetilde{\Omega}} |h_{0,j_1}|^2 - \eta_{R-|\Omega|}, & \text{if } j=0 \\ |h_{j,R+1}|^2 D^{(j)} - \eta_{R-|\Omega|}, & \text{if } j\in\Omega \\ -\eta_{R-|\Omega|}, & \text{otherwise} \end{cases}.$$

Substituting (32) into (30), we have

$$\arg\max_{\boldsymbol{\alpha}} \min_{\Omega\subseteq\{1,2,\cdots,R\}} \left\{ \min_{\bar{\mathbf{D}}\in\bar{\Psi}_\Omega} \bar{\mathbf{D}}^T\boldsymbol{\alpha} \right\} = \arg\max_{\boldsymbol{\alpha}} \min_{\bar{\mathbf{D}}\in\bar{\Psi}} \bar{\mathbf{D}}^T\boldsymbol{\alpha}, \tag{34}$$

where

$$\bar{\Psi} = \bigcup_{\Omega\subseteq\{1,2,\cdots,R\}} \bar{\Psi}_\Omega. \tag{35}$$

Define $\bar{M}(\boldsymbol{\alpha}) = \min_{\bar{\mathbf{D}}\in\bar{\Psi}} \bar{\mathbf{D}}^T\boldsymbol{\alpha}$, which can be regarded as a generalized MWSED (GMWSED) that jointly considers $M_\Omega(\boldsymbol{\alpha})$, $\alpha_0 \sum_{j_1\notin\widetilde{\Omega}} |h_{0,j_1}|^2 D_\chi$ and $\eta_{R-|\Omega|}$ among all Ω's. With this formulation, the optimization in (30) becomes

$$\begin{aligned} \arg\max_{\boldsymbol{\alpha}} \quad & \bar{M}(\boldsymbol{\alpha}) \\ \text{s.t.} \quad & \mathbf{1}^T\boldsymbol{\alpha} \le 1,\ 0\le\alpha_j\le 1,\ j=0,1,\cdots,R. \end{aligned} \tag{36}$$

To solve (36), firstly some $\bar{\mathbf{D}}$'s, which are irrelevant in finding $\boldsymbol{\alpha}_{PA-GMWSED}$, are removed from the set $\bar{\Psi}$ according to the following Lemma whose proof is given in Appendix B.

Lemma-1: For a $\bar{\mathbf{D}}_2\in\bar{\Psi}$, if there exist a $\bar{\mathbf{D}}_1\in\bar{\Psi}$ with $\bar{D}_1^{(j)} \le \bar{D}_2^{(j)}$ for all j, then $\bar{\mathbf{D}}_2$ can be removed from $\bar{\Psi}$ without changing $\bar{M}(\boldsymbol{\alpha})$ in (36).

After removal of irrelevant $\bar{\mathbf{D}}$'s, the optimization in (36) becomes

$$\begin{aligned} \arg\max_{\boldsymbol{\alpha}} \quad & \min_{\bar{\mathbf{D}}\in\Psi_{PA-GMWSED}} \bar{\mathbf{D}}^T\boldsymbol{\alpha} \\ \text{s.t.} \quad & \mathbf{1}^T\boldsymbol{\alpha} \le 1,\ 0\le\alpha_j\le 1,\ j=0,1,\cdots,R, \end{aligned} \tag{37}$$

where $\Psi_{PA-GMWSED}$ is the set after removing irrelevant $\bar{\mathbf{D}}$'s from $\bar{\Psi}$.

Example-1: Consider Setup-1 with $|h_{0,1}|^2=5$, $|h_{0,2}|^2=1$ and $|h_{1,2}|^2=10$. In this case, (32) becomes

$$\begin{aligned} & M_\Omega(\boldsymbol{\alpha}) + \alpha_0 \sum_{j\notin\widetilde{\Omega}} |h_{0,j}|^2 D_\chi - \eta_{R-|\Omega|} \\ = & \begin{cases} \min\limits_{\mathbf{D}\in\Psi_{PA-MWSED}} \left\{ \alpha_0 D^{(0)} + 10\alpha_1 D^{(1)} \right\}, & \Omega=\{1\} \\ 24\alpha_0 - \eta_1, & \Omega=\emptyset \end{cases} \end{aligned} \tag{38}$$

with $\Psi_{PA-MWSED}$ given as Setup-1 in Table I and $D_\chi=4$ for 16-QAM. Substituting $24\alpha_0-\eta_1$ by $(24-\eta_1)\alpha_0-\eta_1\alpha_1$, $\bar{\Psi}_\Omega$ becomes

$$\bar{\Psi}_\Omega = \begin{cases} \left\{ \begin{matrix} [4\ 40]^T, [8\ 80]^T, [16\ 160]^T, [20\ 200]^T, \\ [32\ 320]^T, [36\ 360]^T, [40\ 400]^T, \\ [52\ 520]^T, [72\ 720]^T, \end{matrix} \right\}, & \Omega=\{1\} \\ \left\{ [24-\eta_1\ \ -\eta_1]^T \right\}, & \Omega=\emptyset \end{cases}.$$

After defining $\bar{\Psi}=\bar{\Psi}_\emptyset\cup\bar{\Psi}_{\{1\}}$ and removing all irrelevant $\bar{\mathbf{D}}$'s from $\bar{\Psi}$ according to Lemma-1, one has

$$\begin{aligned} & \Psi_{PA-GMWSED} \\ = & \begin{cases} \left\{ [4\ 40]^T, [24-\eta_1\ -\eta_1]^T \right\}, & \text{if } \eta_1<20 \\ \left\{ [24-\eta_1\ -\eta_1]^T \right\}, & \text{if } \eta_1\ge 20 \end{cases}, \end{aligned} \tag{39}$$

where the number of vectors is reduced from 10 in $\bar{\Psi}$ to 2 or 1 in (39) depending on the value of η_1.

Example-2: Consider Setup-2 with $|h_{0,1}|^2 = 5$ and $|h_{0,2}|^2=|h_{1,2}|^2=1$. In this case, (32) becomes

$$\begin{aligned} & M_\Omega(\boldsymbol{\alpha}) + \alpha_0 \sum_{j\notin\widetilde{\Omega}} |h_{0,j}|^2 D_\chi - \eta_{R-|\Omega|} \\ = & \begin{cases} \min\limits_{\mathbf{D}\in\Psi_{PA-MWSED}} \left\{ \alpha_0 D^{(0)} + \alpha_1 D^{(1)} \right\}, & \Omega=\{1\} \\ 24\alpha_0 - \eta_1, & \Omega=\emptyset \end{cases} \end{aligned} \tag{40}$$

with $\Psi_{PA-MWSED}$ given as Setup-2 in Table I. Also, it can be shown that

$$\begin{aligned} & \Psi_{PA-GMWSED} \\ = & \begin{cases} \left\{ [16\ 4]^T, [4\ 16]^T, [24-\eta_1\ -\eta_1]^T \right\}, & \text{if } \eta_1<8 \\ \left\{ [4\ 16]^T, [24-\eta_1\ -\eta_1]^T \right\}, & \text{if } 8\le\eta_1<20 \\ \left\{ [24-\eta_1\ -\eta_1]^T \right\}, & \text{if } \eta_1\ge 20 \end{cases}. \end{aligned} \tag{41}$$

In this case, the number of vectors has been reduced from 24 in $\bar{\Psi}$ to 3 or less in $\Psi_{PA-GMWSED}$.

From (37), it can be easily seen that the optimization problem can also be recast as

$$\begin{aligned} \max_{\boldsymbol{\alpha},\ \beta} \quad & \beta \\ \text{s.t.} \quad & \beta \le \bar{\mathbf{D}}^T\boldsymbol{\alpha},\ \forall\bar{\mathbf{D}}\in\Psi_{PA-GMWSED} \\ & \mathbf{1}^T\boldsymbol{\alpha}\le 1, \text{ and } 0\le\alpha_j\le 1,\ j=0,1,\cdots,R. \end{aligned} \tag{42}$$

Since the objective function and all the constraints are linear, (42) is a linear programming problem (though not in a standard form) and can be solved efficiently with the Simplex method [21]. To have a better understanding how the power is allocated, we provide the following example.

Example-3: Consider Example-2. According to (41), when the SNR is low such that $\eta_1\ge 20$, we have only one vector in $\Psi_{PA-GMWSED}$ and consequently $\bar{M}(\boldsymbol{\alpha})=24\alpha_0-\eta_1$. The optimum of $\bar{M}(\boldsymbol{\alpha})$ occurs at $\alpha_0=1$. As the SNR increases to $8\le\eta_1<20$, there are two vectors in $\Psi_{PA-GMWSED}$. The corresponding curves of $\bar{\mathbf{D}}^T\boldsymbol{\alpha}$ for $\bar{\mathbf{D}}_2=[4\ 16]^T$ and $\bar{\mathbf{D}}_3=[24-\eta_1\ -\eta_1]^T$ with different values of η_1 are plotted in Fig. 3 as functions of α_0 (note that $\alpha_1=1-\alpha_0$). As

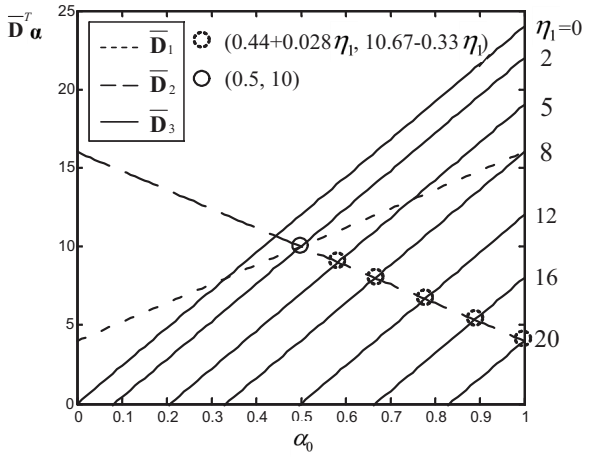

Fig. 3. Curves of $\bar{\mathbf{D}}^T\boldsymbol{\alpha}$, $\bar{\mathbf{D}} \in \Psi_{PA-GMWSED}$ for Example-3, where $\bar{\mathbf{D}}_1 = [16\ 4]^T$, $\bar{\mathbf{D}}_2 = [4\ 16]^T$ and $\bar{\mathbf{D}}_3 = [24-\eta_1\ \ -\eta_1]^T$.

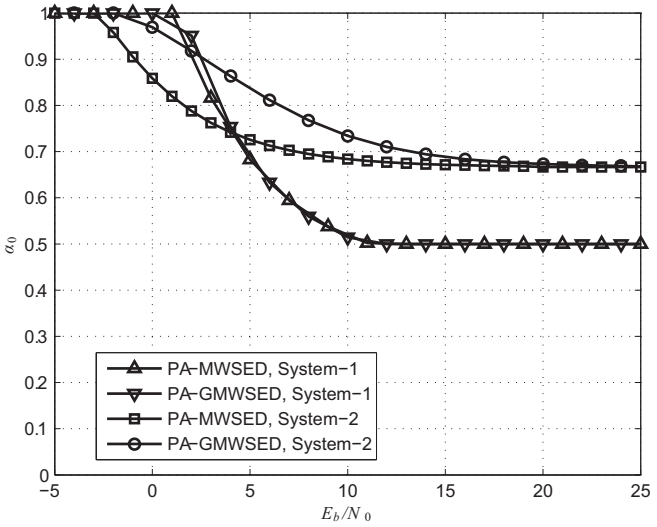

Fig. 4. Power allocation results of different methods for System-1 and System-2.

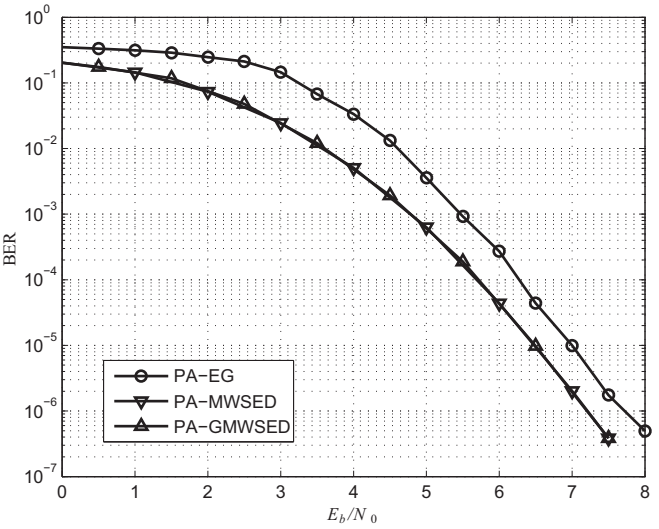

Fig. 5. BER performance of System-1 with power allocations in Fig. 4.

shown in Fig. 3, the optimum of $\bar{M}(\boldsymbol{\alpha}) = \min\{\bar{\mathbf{D}}_2^T\boldsymbol{\alpha}, \bar{\mathbf{D}}_3^T\boldsymbol{\alpha}\}$ occurs at the intersection of the two lines. By solving $\bar{\mathbf{D}}_2^T\boldsymbol{\alpha} = \bar{\mathbf{D}}_3^T\boldsymbol{\alpha}$, the intersection point locates at $\boldsymbol{\alpha}_{PA-GMWSED} = [0.44+0.028\eta_1\ \ 0.56-0.028\eta_1]^T$ and $\bar{M}(\boldsymbol{\alpha}_{PA-GMWSED}) = 10.67 - 0.33\eta_1$. For convenience, we denote this intersection point by $(0.44 + 0.028\eta_1, 10.67 - 0.33\eta_1)$. In the case of $\eta_1 < 8$, we have $\Psi_{PA-GMWSED} = \{\bar{\mathbf{D}}_1, \bar{\mathbf{D}}_2, \bar{\mathbf{D}}_3\}$ with $\bar{\mathbf{D}}_1 = [16\ 4]^T$. When $2 < \eta_1 < 8$, the optimum is still at $(0.44 + 0.028\eta_1, 10.67 - 0.33\eta_1)$. When $\eta_1 \leq 2$, $\bar{M}(\boldsymbol{\alpha})$ is maximized at $\alpha_0 = 0.5$, which is the intersection between $\bar{\mathbf{D}}_1^T\boldsymbol{\alpha}$ and $\bar{\mathbf{D}}_2^T\boldsymbol{\alpha}$. As is observed, the full power is allocated to source at low SNRs. As the SNR increases (η_1 decreases), the allocated source power decreases and finally arrives at 0.5 when the SNR is high enough such that $\eta_1 \leq 2$. In summary, the optimum power allocation occurs at $\alpha_0 = 1$ when the SNR is very low. As the SNR increases, the decreasing η_1 makes the optimal α_0 become smaller. This is quite obvious; when SNR is very low such that relay can never decode correctly, all power should be allocated to the source. And, when the SNR increases, the relay may be given some power to improve the BER performance.

V. Numerical Results

A. Power Allocations and BER Performances

In this subsection, numerical results are given to confirm the effectiveness of the proposed methods. In the following example systems, two 1/2-rate covolutional codes are considered: the CC(7,5) code with the generator matrix $(1 + D + D^2, 1 + D^2)$ is employed in System-1, 3, 4 and 5, and the CC(171,133) code with the generator matrix $(1+D+D^2+D^3+D^6, 1+D^2+D^3+D^5+D^6)$ is employed in System-2. The interleaver is S-random with length of 6144 and depth of 20 [2]. The modulation constellation is 16-QAM. Equations (2) and (4) are used to evaluate the LLRs at relays and destination, respectively. The Max-Log MAP decoder [24] is employed in all receiving nodes.

Figure 4 shows the power allocation results with PA-MWSED and PA-GMWSED for System-1, where $R = 1$, $\mu^{(0)} = \mu_G$, $\mu^{(1)} = \mu_A$, $|h_{0,1}|^2 = 5$ and $|h_{0,2}|^2 = |h_{1,2}|^2 = 1$. As is shown, for $E_b/N_0 \leq 0$ dB, all power is allocated to the source in both methods; as the SNR increases, the allocated source power of PA-MWSED and PA-GMWSED are close with PA-MWSED slightly lower than PA-GMWSED. When $E_b/N_0 \geq 12$ dB, PA-MWSED and PA-GMWSED have the same power allocation at $\alpha_0 = 0.5$ (as predicted in Fig. 3 with $\eta_1 \leq 2$). The BER performance of the system is given in Fig. 5. The proposed methods achieve almost the same BER performance and outperform PA-EG by at all ranges of SNRs. In particular, 0.5 dB gain is observed at BER of 10^{-5}.

The power allocation results for System-2 are also plotted in Fig. 4, where $R = 1$, $\mu^{(0)} = \mu^{(1)} = \mu_G$, $|h_{0,1}|^2 = 5$, $|h_{0,2}|^2 = 1$ and $|h_{1,2}|^2 = 10$. Similarly, both the proposed

[2]For the block-fading channel considered in this paper, the channels remain constant over the transmission of a packet. Therefore, the use of the interleaver is mainly to break the correlation introduced by high-order modulations. In [9], it was shown that an interleaver depth of 10 is sufficient for 8-PSK modulation to approach the ultimate performance. In our case, where 16-QAM is used, an interleaver depth of 20 is generally enough for not incurring performance loss.

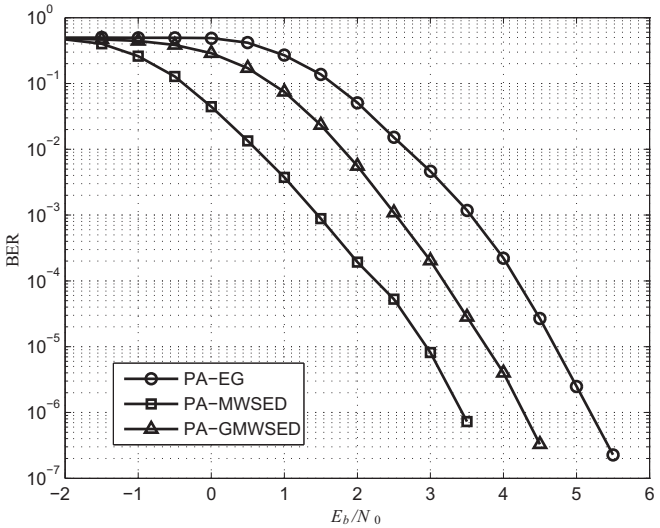

Fig. 6. BER performance of System-2 with power allocations in Fig. 4.

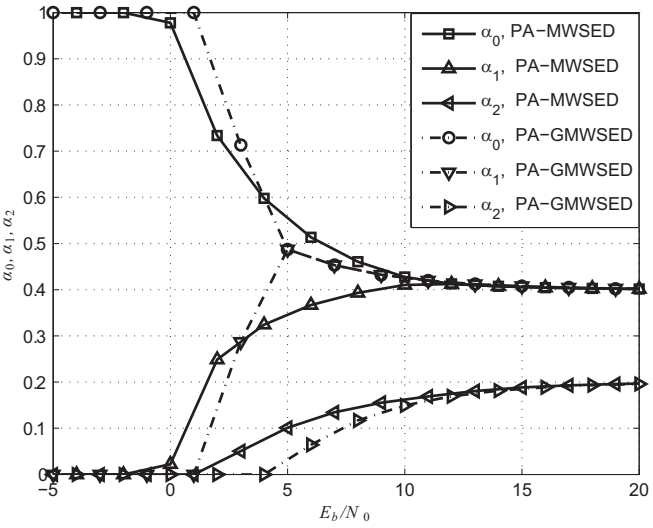

Fig. 7. Power allocation results for System-3.

methods allocate all power to the source when $E_b/N_0 \leq -3$ dB. As SNR increases, the allocated source power with PA-MWSED decreases much faster than that with PA-GMWSED. For $E_b/N_0 \geq 23$ dB, both methods allocate the same power to the source at about $\alpha_0 = 0.67$. Figure 6 shows the BER performance. As is seen, PA-MWSED achieves the best performance at all ranges of SNRs, followed by PA-GMWSED and PA-EG. At BER of 10^{-5}, PA-MWSED has a gain of about 1.8 dB, and PA-GMWSED has about 1 dB over PA-EG.

Figure 7 shows the power allocation results for System-3, where $R = 2$, $\mu^{(0)} = \mu_G$, $\mu^{(1)} = \mu_A$, $\mu^{(2)} = \mu_B$, $|h_{0,3}|^2 = 1$, $|h_{0,1}|^2 = 10$, $|h_{0,2}|^2 = 5$, $|h_{1,3}|^2 = 1$, and $|h_{2,3}|^2 = 2$. The two relays have different qualities of the source-relay and relay-destination links and different mappers. Since $|h_{0,1}|^2 > |h_{0,2}|^2$, relay 1 is more likely to have a correct decoding at low SNRs than relay 2. Therefore, each the power allocation method starts to allocate non-zero α_1 at the SNR earlier than that of α_2. The BER performance is given in Fig. 8, where PA-MWSED still achieves the best BER performance, closely followed by PA-GMWSED. PA-EG is about 1.5 dB and 1.1 dB worse than PA-MWSED and PA-GMWSED at BER of 10^{-5}, respectively.

Figure 9 shows the BER performance for System-4 and 5 with $R = 3$. The channels $|h_{0,4}|^2 = 1$, $|h_{0,1}|^2 = 15$, $|h_{0,2}|^2 = 10$, $|h_{0,3}|^2 = 5$, $|h_{1,4}|^2 = 1$, $|h_{2,4}|^2 = 2$ and $|h_{3,4}|^2 = 3$ are used for both systems but with different mappers at relays; in particular, $\mu^{(0)} = \mu^{(1)} = \mu^{(2)} = \mu^{(3)} = \mu_G$ are employed in System-4 and $\mu^{(0)} = \mu_G$, $\mu^{(1)} = \mu_A$, $\mu^{(2)} = \mu_B$ and $\mu^{(3)} = \mu_C$ in System-5 (see Fig. 2). As is seen in Fig. 9, PA-MWSED and PA-GMWSED perform similarly and outperform PA-EG by 1.1 dB and 1.8 dB at BER of 10^{-5} for System-4 and 5, respectively. In addition, System-5 outperforms System-4 by using different mappers at relays; the performance improvement is quite significant; for example, at BER of 10^{-5}, 1.3 dB is observed for PA-EG and 2 dB for PA-MWSED and PA-GMWSED, respectively.

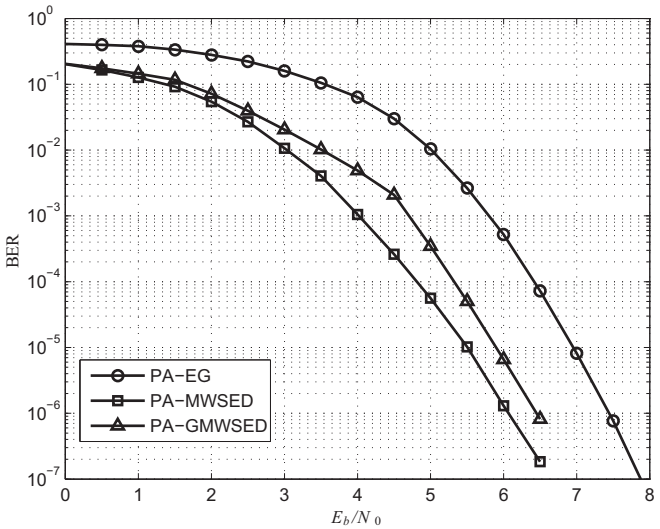

Fig. 8. BER performance of System-3 with power allocations in Fig. 7.

B. Complexity Comparison

The computation complexities of PA-MWSED and PA-GMWSED are compared here for the 5 example systems investigated in the previous subsection (at $E_b/N_0 = 10$ dB). The number of iterations for the sub-gradient method in PA-MWSED is fixed at 10. In Table II, we list the total numbers of required multiplications $(\times)$, additions $(+)$, square-roots $(\sqrt{x})$, exponentials (e^x) and nature-logs $(\ln x)$. As is shown, PA-GMWSED has a much lower complexity than PA-MWSED. Note that the major complexity of PA-MWSED comes from the evaluation of 2^R different terms $(g_\Omega \prod_{j \notin \widetilde{\Omega}} KW_I(d_f)\tilde{g}_j)$ in the summation of $G(\boldsymbol{\alpha})$. Furthermore, in each iteration, $G(\boldsymbol{\alpha})$ is calculated twice as in (25) and is repeated $(R+1)$ times because of taking the sub-gradients with respect to all entries in $\boldsymbol{\alpha}$. On the other hand, the omission of $\ln A_\Omega(\boldsymbol{\alpha})$ in (27) and the removal of the irrelevant $\bar{\mathbf{D}}$'s in $\bar{\Psi}$ effectively reduce the complexity of PA-GMWSED. It is also observed that in each of the proposed methods the complexity increases substantially with R due to that the number of Ω's increases exponentially with R.

TABLE II
COMPLEXITY COMPARISONS BETWEEN PA-MWSED (M) AND PA-GMWSED (G)

	System-1		System-2		System-3		System-4		System-5	
R	1		1		2		3		3	
$\{\mu^{(j)}\}$	$\mu_G,\ \mu_A$		$\mu_G,\ \mu_G$		$\mu_G,\ \mu_A,\ \mu_B$		$\mu_G,\ \mu_G,\ \mu_G,\ \mu_G$		$\mu_G,\ \mu_A,\ \mu_B,\ \mu_C$	
Coding	CC(7,5)		CC(171,133)		CC(7,5)		CC(7,5)		CC(7,5)	
$\{\lvert h_{i,j}\rvert^2\}$	$\lvert h_{0,2}\rvert^2=1,$ $\lvert h_{0,1}\rvert^2=5,$ $\lvert h_{1,2}\rvert^2=1$		$\lvert h_{0,2}\rvert^2=1,$ $\lvert h_{0,1}\rvert^2=5,$ $\lvert h_{1,2}\rvert^2=10$		$\lvert h_{0,3}\rvert^2=1,$ $\lvert h_{0,1}\rvert^2=10,\ \lvert h_{0,2}\rvert^2=5,$ $\lvert h_{1,3}\rvert^2=1,\ \lvert h_{2,3}\rvert^2=2$		$\lvert h_{0,4}\rvert^2=1,$ $\lvert h_{0,1}\rvert^2=15,\ \lvert h_{0,2}\rvert^2=10,\ \lvert h_{0,3}\rvert^2=5,$ $\lvert h_{1,4}\rvert^2=1,\ \lvert h_{2,4}\rvert^2=2,\ \lvert h_{3,4}\rvert^2=3$			
PA	M	G	M	G	M	G	M	G	M	G
$\times$	6820	330	3460	210	52470	2016	42440	1299	196040	17595
$+$	2780	629	1100	232	24090	8503	14840	3207	91640	93281
$\sqrt{x}$	160	2	160	2	480	3	1280	4	1280	4
e^x	80	0	80	0	240	0	640	0	640	0
$\ln x$	0	2	0	2	0	3	0	4	0	4

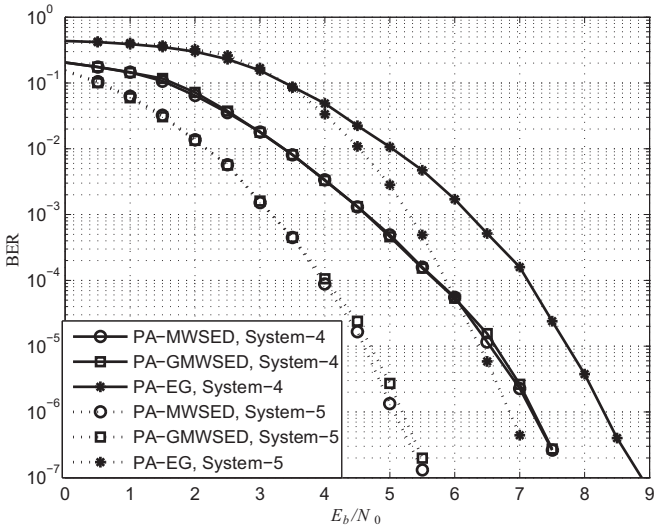

Fig. 9. BER performance of System-4 and System-5 with different power allocation methods.

VI. CONCLUSIONS

This paper investigates power allocation for the cooperative BICM systems with DRF relaying. Two methods of power allocation, PA-MWSED and PA-GMWSED, are proposed under the assumption of perfect knowledge of CSI. In PA-MWSED, a cost function is derived based on the saddle-point integration and MWSED and optimized with the sub-gradient method. In PA-GMWSED, with the introduction of a generalized MWSED, the optimization is re-cast as a linear programming problem which can be solved efficiently via the Simplex algorithm. Several examples with different relay number, mappers and channels are given to confirm the effectiveness of the proposed methods. Simulation results show that both the proposed methods outperform PA-EG with large margins under different system configurations.

APPENDIX A
THE CONVEXITY OF $G(\boldsymbol{\alpha})$

To show that $G(\boldsymbol{\alpha})$ is convex, we first re-write $G(\boldsymbol{\alpha}) = \sum_{\Omega\subseteq\{1,2,\cdots,R\}} F_\Omega(\boldsymbol{\alpha})$ and $F_\Omega(\boldsymbol{\alpha}) = h(q_1,q_2) = q_1\cdot q_2$, where $q_1 = g_\Omega$ and $q_2 = \prod_{j\notin\tilde{\Omega}} KW_I(d_f)\tilde{g}_j$. We aim to show that $F_\Omega(\boldsymbol{\alpha})$ is convex for any Ω, and so is $G(\boldsymbol{\alpha})$ which is a sum of convex functions. According to [25], $F_\Omega(\boldsymbol{\alpha})$ is convex if (a) h is convex in each argument, (b) h is non-decreasing in each argument and (c) both q_1 and q_2 are convex. Firstly, (a) and (b) can be proved to be true by evaluating the first and second derivatives of h w.r.t. q_1 and q_2. Now consider the convexity of q_1 and q_2. According to [25], a function $S(T(\boldsymbol{\alpha}))$ is convex if S is convex non-increasing and T is concave. Define $S(x) = a_1 x^{-a_2/2} e^{-a_3 x}$. By evaluating the first and second derivatives of S, it can be shown that S is convex non-increasing in the region $x>0$ for any $\alpha_1>0$, $\alpha_3>0$ and any positive integer α_2. Let $T(\boldsymbol{\alpha}) = M_\Omega(\boldsymbol{\alpha})$, which is the minimum of linear functions of $\boldsymbol{\alpha}$, is positive and concave. Since we can write $q_1 = S(T(\boldsymbol{\alpha}))$ by properly choosing a_1, a_2 and a_3, q_1 is convex. Similarly, q_2 can also be proved convex by letting $T(\boldsymbol{\alpha}) = \alpha_0$. Q.E.D.

APPENDIX B
PROOF OF LEMMA-1

Since $\bar{D}_1^{(j)} \le \bar{D}_{(2)}^{(j)}$ for all j, one has

$$\bar{\mathbf{D}}_1^T\boldsymbol{\alpha} = \sum_{j=0}^{R} \bar{D}_1^{(j)}\alpha_j \le \sum_{j=0}^{R} \bar{D}_2^{(j)}\alpha_j = \bar{\mathbf{D}}_2^T\boldsymbol{\alpha}, \tag{43}$$

for any $\boldsymbol{\alpha}$ with non-negative entries, and thus,

$$\bar{M}(\boldsymbol{\alpha}) = \min_{\bar{\mathbf{D}}\in\bar{\Psi}} \bar{\mathbf{D}}^T\boldsymbol{\alpha} = \min_{\bar{\mathbf{D}}\in\{\bar{\Psi}\setminus\bar{\mathbf{D}}_2\}} \bar{\mathbf{D}}^T\boldsymbol{\alpha}. \tag{44}$$

One can remove $\bar{\mathbf{D}}_2$ from $\bar{\Psi}$ without changing $\bar{M}(\boldsymbol{\alpha})$. Q.E.D.

REFERENCES

[1] J. N. Laneman and G. W. Wornell, "Distributed space-time coded protocols for exploiting cooperative diversity in wireless networks," *IEEE Trans. Inf. Theory*, vol. 49, pp. 2415–2425, Oct. 2003.

[2] 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, pp. 3062–3080, Dec. 2004.

[3] X. J. Zhang and Y. Gong, "Joint power allocation and relay positioning in multi-relay cooperative systems," *IET Commun.*, vol. 3, pp. 1683–1692, Sep. 2009.

[4] N. Ahmed, M. A. Khojastepour, and B. Aazhang, "Outage minimization and optimal power control for the fading relay channel," in *Proc. 2004 IEEE Inform. Theory Workshop*, pp. 458–462.

[5] W. Su and X. Liu, "On optimum selection relaying protocols in cooperative wireless networks," *IEEE Trans. Commun.*, vol. 58, pp. 52–57, Jan. 2010.

[6] M. R. Souryal and B. R. Vojcic, "Performance of amplify-and-forward and decode-and-forward relaying in Rayleigh fading with turbo codes," in *Proc. 2006 IEEE ICASSP*, pp. 681–684.

[7] Q. Li, S. H. Ting, A. Pandharipande, and Y. Han, "Adaptive two-way relaying and outage analysis," *IEEE Trans. Wireless. Commun.*, vol. 8, pp. 3288–3299, June 2009.

[8] E. Zehavi, "Eight-PSK trellis codes for a Rayleigh channel," *IEEE Trans. Commun.*, vol. 40, pp. 873–884, May 1992.

[9] G. Caire, G. Taricco, and E. Biglieri, "Bit-interleaved coded modulation," *IEEE Trans. Inf. Theory*, vol. 44, pp. 927–946, May 1998.

[10] IEEE Standard for Local and Metropolitan Area Networks, Part 11: Wireless LAN medium access control (MAC) and physical layer (PHY) specifications, Amendment 5: Enhancements for higher throughputs, IEEE Std 802.11n-2009, 2009.

[11] IEEE Standard for Local and Metropolitan Area Networks, Part 16: Air interface for broadband wireless access systems, IEEE Std 802.16e-2009, 2009.

[12] R. Hoshyar and R. Tafazolli, "BER performance analysis of a cooperative BICM system based on post-BSC model," in *Proc. IEEE 2008 PIMRC*, pp. 1–5.

[13] R. Hoshyar and R. Tafazolli, "Achievable full decode and forward rates for cooperative MIMO BICM system," in *Proc. 2008 IEEE PIMRC*, pp. 1–5.

[14] G. M. Kraidy, N. Gresset, and J. J. Boutros, "Coding for the nonorthogonal amplify-and-forward cooperative channel," *IEEE Trans. Inf. Theory*, vol. 56, pp. 2601–2610, June 2010.

[15] T. Islam, R. Schober, R. K. Mallik, and V. K. Bhargava, "Analysis and design of cooperative BICM-OFDM systems," *IEEE Trans. Commun.*, vol. 59, pp. 1742–1751, June 2011.

[16] T. Islam, A. Nasri, R. Schober, and R. K. Mallik, "Analysis and relay placement for DF cooperative BICM-OFDM systems," in *Proc. 2011 IEEE WCNC*, pp. 1670–1675.

[17] M. N. Khormuji and E. G. Larsson, "Rate-optimized constellation rearrangement for the relay channel," *IEEE Commun. Lett.*, vol. 12, pp. 618–620, Sep. 2008.

[18] M. N. Khormuji and E. G. Larsson, "Improving collaborative transmit diversity by using constellation rearrangement," in *Proc. 2007 IEEE WCNC*, pp. 804–808.

[19] Z. Si, R. Thobaben, and M. Skoglund, "Instantaneous relaying with bit-interleaved coded modulation: design and optimization," in *Proc. 2010 IEEE ISTC*, pp. 221–225.

[20] N. Z. Shor, *Minimization Methods for Non-Differentiable Functions*. Springer Series in Computational Mathematics, 1985.

[21] G. B. Dantzig, *Linear Programming and Extensions*. Princeton University Press, 1963.

[22] C. W. Helstrom, *Elements of Signal Detection and Estimation*. PTR Prentice Hall, 1995.

[23] L. Szczecinski, F.-K. Diop, M. Benjillali, A. Ceron, and R. Feick, "BICM in HARQ with mapping rearrangement: capacity and performance of practical schemes," in *Proc. 2007 IEEE GLOBECOM*, pp. 1410–1415.

[24] P. Robertson, E. Villebrun, and P. Hoeher, "A comparison of optimal and sub-optimal decoding algorithms in the log domain," in *Proc. 1995 IEEE ICC*, pp. 1009–1013.

[25] S. P. Boyd and L. Vandenberghe, *Convex Optimization*. Cambridge University Press, 2003.

[26] T. W. Yu and W. H. Sheen, "On the performance of cooperative BICM systems with selection decode-and-forward relaying over block fading channels," unpublished.

Tsang-Wei Yu was born in Kaohsiung, Taiwan in 1982. He received the B.S. degree in communication engineering from the National Chiao Tung University (NCTU), Taiwan in 2004. He is currently a Ph.D. student with the Graduate Institute of Communication Engineering, NCTU. His research interest is in wireless communications, particularly in coding theory and cooperative communications.

Wern-Ho Sheen (M'91) received his Ph.D. degree from the Georgia Institute of Technology, Atlanta, USA in 1991. From 1991 to 1993, he was with Chunghwa Telecom. Labs. as an Associate Researcher. From 1993 to 2001, he was with the National Chung Cheng University, where he held positions as a Professor in the Department of Electrical Engineering and the Managing Director of the Center for Telecommunication Research. From 2001 to 2009, he was a Professor in the Department of Communications Engineering, National Chiao Tung University. Currently he is a Professor in the Department of Information and Communication Engineering, Chaoyang University of Technology. His research interests include communication theory, wireless communication systems, and signal processing for communications.

Chung-Hsuan Wang (S'94 - M'01) received the B.S. and Ph.D. degrees from the National Tsing Hua University, Hsinchu, Taiwan, in 1994 and 2001, respectively, both in electrical engineering.

From August 2001 to July 2004, he held a faculty position with the Department of Electronic Engineering, Chung Yuan Christian University, Chung-Li, Taiwan. From August 2004 to July 2009, he was with the Department of Communication Engineering, National Chiao Tung University (NCTU), Hsinchu. Since August 2009, he has been with the Institute of Communication Engineering and the Department of Electrical Engineering, NCTU. His current research interests include digital communications, error-correcting codes, information theory, and signal processing.