Dynamic Resource Allocation in MIMO-OFDMA Systems with Full-Duplex and Hybrid Relaying

Derrick Wing Kwan Ng, *Student Member, IEEE,* Ernest S. Lo, *Member, IEEE,*
and Robert Schober, *Fellow, IEEE*

***Abstract*—In this paper, we formulate a joint optimization problem for resource allocation and scheduling in full-duplex multiple-input multiple-output orthogonal frequency division multiple access (MIMO-OFDMA) relaying systems with amplify-and-forward (AF) and decode-and-forward (DF) relaying protocols. Our problem formulation takes into account heterogeneous data rate requirements for delay sensitive and non-delay sensitive users. We also consider a theoretically optimal hybrid relaying scheme as a performance benchmark, which allows a dynamic selection between AF relaying and DF relaying protocols with full-duplex and half-duplex relays. We show that under some mild conditions the optimal transmitter precoding and receiver post-processing matrices jointly diagonalize the MIMO-OFDMA relay channels for all considered relaying protocols transforming the resource allocation and scheduling problem into a scalar optimization problem. Dual decomposition is employed to solve this optimization problem and a distributed iterative resource allocation and scheduling algorithm with closed-form power and subcarrier allocation is derived. Simulation results not only illustrate that the proposed distributed algorithm converges to the optimal solution in a small number of iterations, but also demonstrate the potential performance gains achievable with full-duplex relaying protocols.**

***Index Terms*—MIMO-OFDMA, amplify-and-forward relays, decode-and-forward relays, full-duplex relaying, loop interference, dual decomposition, distributed resource allocation, multiuser diversity.**

I. INTRODUCTION

MULTIPLE-input multiple-output (MIMO) and orthogonal frequency division multiple access (OFDMA) [1]-[3] are important techniques for high data rate wireless multiuser communication systems, such as 3GPP Long Term Evolution (LTE) and IEEE 802.16 Worldwide Interoperability for Microwave Access (WiMAX), not only because of its flexibility in resource allocation, but also because of its ability to exploit multiuser diversity. On the other hand, cooperative relaying for wireless networks has received considerable interest, as it provides coverage extension and reduced power consumption without incurring the high costs of additional base station (BS) deployment. There are two main types of relaying schemes, namely half-duplex (HD) relaying and full-duplex (FD) relaying. In the literature, a large amount of work has been devoted to HD relaying [4]-[8] as it enables a low-complexity relay design. Nevertheless, HD relaying systems require additional resources (time slots or frequencies) to transmit data in a multi-hop manner which results in a loss of spectral efficiency. Even though there exist several approaches for minimizing/recovering the spectral efficiency loss associated with HD relaying, such as non-orthogonal relaying [9], [10] and two-way relaying [11]-[13], these schemes do not solve the problem fundamentally since the associated protocols are still using HD relaying [14]. On the contrary, although FD relaying suffers from inherent loop interference which was considered impractical in the past, it has regained the attention of both industry [15], [16] and academia [14], [17]-[22]. Recent research shows that FD relaying is feasible by using interference cancellation techniques and transmit/receive antenna isolation [19], [22]. Several prototypes of FD transceivers using different self-interference cancellation techniques have been built to demonstrate the feasibility of FD relaying and the expected performance gains compared to HD relays [23]-[26]. However, efficient resource allocation and scheduling algorithms for MIMO-OFDMA FD relaying systems with interference cancellation error have not been studied, yet.

Next generation wireless communication systems are required to support heterogeneous data rate services and guarantee certain quality of service (QoS) requirements. The combination of MIMO, OFDMA, and relaying provides a viable solution for addressing these issues. In [4]-[8] and [27]-[29], best effort resource allocation for homogeneous users in OFDMA HD relaying systems and MIMO HD relaying systems were studied for different system configurations, respectively. However, QoS requirements are driven by heterogeneous applications and different users may demand different data rates, which best effort resource allocation cannot guarantee. On the other hand, FD relaying could provide a substantial performance gain and should not be overlooked in the system design. Furthermore, existing works focus on centralized resource allocation at the BS. As the numbers of users/relays and subcarriers in the system increase, brute force optimization of resource allocation may overload the BS which limits the system's scalability. Therefore, a distributed resource allocation algorithm, which enables the exploitation of the

Paper approved by J. Wang, the Editor for Wireless Spread Spectrum of the IEEE Communications Society. Manuscript received April 18, 2011; revised November 23, 2011.

D. W. K. Ng and R. Schober are with the Department of Electrical and Computer Engineering, University of British Columbia, Vancouver, BC, V6T 1Z4, Canada (e-mail: wingn@ece.ubc.ca; rschober@ece.ubc.ca).

E. S. Lo is with the Centre Tecnològic de Telecomunicacions de Catalunya-Hong Kong (CTTC-HK) (e-mail: ernest.lo@cttc.hk).

This paper has been presented in part at the IEEE International Conference on Communications (ICC 2011), Kyoto, Japan, June 2011.

This work was supported by the Natural Science and Engineering Research Council of Canada (NSERC) through Grant STPGP 396545-2010.

Digital Object Identifier 10.1109/TCOMM.2012.031712.110233

benefits of different relaying protocols and duplexing schemes, fulfills heterogeneous QoS requirements, and converges fast to the optimal solution is needed for practical implementation.

In this paper, we address the above issues. For this purpose, we formulate the scheduling and resource allocation problem for MIMO-OFDMA FD-relaying systems as an optimization problem. A theoretically optimal hybrid relaying scheme, which dynamically selects between amplify-and-forward (AF) relaying and decode-and-forward (DF) relaying with FD and HD relays, is also considered in the problem formulation and serves as a performance benchmark. By exploiting the structure of the optimal transmitter precoding and receiver post-processing matrices, the considered problem can be transformed into a scalar optimization problem. This optimization problem is further decomposed into a master problem and several subproblems by dual decomposition, which leads to a distributed iterative resource allocation and scheduling algorithm with closed-form power and subcarrier allocation. The BS solves the master problem with a gradient method and updates the dual variables through the concept of pricing, while each relay solves its own subproblem by utilizing the dual variables and its local channel state information (CSI) without any help from other relays.

II. MIMO-OFDMA Relay Network Model

A. Notation

For a square-matrix $\mathbf{S}$: $\det[\mathbf{S}]$, diag($\mathbf{S}$), eig($\mathbf{S}$), tr($\mathbf{S}$), $(\mathbf{S})^{-1}$, and $(\mathbf{S})^{1/2}$ denote the determinant, diagonal element vector, eigenvalue vector, trace, inverse, and square root of matrix $\mathbf{S}$, respectively; $\mathbf{S} \succeq 0$ means that $\mathbf{S}$ is a positive semi-definite matrix. $(\mathbf{S})^H$ and rank($\mathbf{S}$) denote the conjugate transpose and the rank of matrix $\mathbf{S}$, respectively. Matrix $\mathbf{I}_N$ denotes the $N \times N$ identity matrix. $\mathcal{E}\{\cdot\}$ denotes statistical expectation. $\mathbb{C}^{N\times M}$ and $(\mathbb{R}_0^+)^{N\times M}$ represent the space of $N \times M$ matrices with complex and non-negative real entries, respectively. The distribution of a circularly symmetric complex Gaussian (CSCG) vector with mean vector $\mathbf{x}$ and covariance matrix $\mathbf{\Sigma}$ is denoted by $\mathcal{CN}(\mathbf{x}, \mathbf{\Sigma})$, and $\sim$ means "distributed as". $(x)^+ = \max\{0, x\}$. Operator $\mu_i(\cdot)$ returns the i-th largest eigenvalue of an input matrix. $\mathcal{O}(g(x))$ represents an asymptotic upper bound, i.e., $f(x) = \mathcal{O}(g(x))$ if $\lim\limits_{x\to\infty} |\frac{f(x)}{g(x)}| \leq N$ for $0 < N < \infty$.

B. System Model

We consider a relay assisted MIMO-OFDMA downlink packet transmission network with n_F subcarriers, one BS, M relays, and K mobile users which belong to one of two categories, namely, *delay sensitive* users and *non-delay sensitive* users. *Delay sensitive* users require a minimum constant data rate while *non-delay sensitive* users require only best-effort service. The BS, relays, and users are assumed to be equipped with N antennas, respectively. A cell is modeled by two concentric ring-shaped discs as shown in Fig. 1. The cell coverage is divided into M areas corresponding to the M relays and each user is assigned to one relay. In this paper, we focus on resource allocation and scheduling for heterogeneous users who need the help of relays, i.e., cell edge users in the shaded

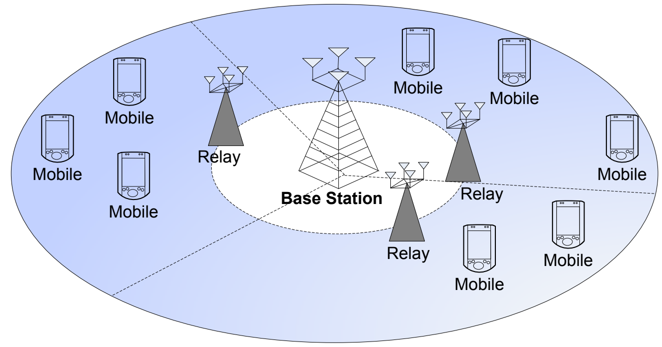


Fig. 1. Relay assisted MIMO OFDMA packet transmission system model with $K = 8$ users and $M = 3$ relays. Each transceiver is equipped with $N = 4$ antennas. The shaded area contains the users who require the help of relays.

region of Fig. 1. We assume that there is no direct transmission between the BS and the mobile users due to heavy blockage and long distance transmission. A distributed algorithm is derived for resource allocation and scheduling purposes. Based on the CSI of the users, the algorithm selects between four transmission strategies t on a per subcarrier basis for hybrid relaying, namely: $t = 1$, decode-and-forward full-duplex (DF-FD) relaying; $t = 2$, decode-and-forward half-duplex (DF-HD) relaying; $t = 3$, amplify-and-forward full-duplex (AF-FD) relaying; and $t = 4$, amplify-and-forward half-duplex (AF-HD) relaying.[1] A time-division channel allocation with two time slots is used to facilitate transmission. In the first time slot, the BS broadcasts a data packet to the relays. Then, in the second time slot, if subcarrier i is using full-duplex (FD) relaying, the corresponding relay decodes/amplifies the previously received signal on subcarrier i and forwards it to the corresponding user while the BS transmits the next packet, cf. Fig. 2. If half-duplex (HD) relaying is used on subcarrier i, the relays perform the same signal processing on subcarrier i as for FD transmission, however, the BS remains silent during the second time slot. In this paper, we assume a slowly time-varying time-division-duplex (TDD) system where channel reciprocity holds within the coherence time[2]. CSI can be obtained by exploiting the pilots of previously received packets or via handshaking signals. Hence, local CSI is available at both the relays and the base station.

C. Amplify-and-Forward Full-Duplex Relaying Channel Model

The channel impulse response is assumed to be time-invariant within a scheduling slot. The data symbol vector $\mathbf{s}_k^{[i]} \in \mathbb{C}^{N\times 1}$ on subcarrier $i \in \{1, \dots, n_F\}$ using transmission strategy $t \in \{1, \dots, 4\}$ for user $k \in \{1, \dots, K\}$ is linearly precoded at the BS as

$$\mathbf{x}_{m,k}^{[t,i]} = \mathbf{B}_{m,k}^{[t,i]} \mathbf{s}_k^{[i]}, \tag{1}$$

[1]The hybrid relaying serves as a theoretical upper bound of the system performance and provides useful system design insights such as the operating region of each transmission strategy. As can be seen in the following sections, we can always restrict the proposed algorithm to select only one of the considered relay protocols for a practical implementation.

[2]In practice, the coherence time depends on the mobility of the users. For example, the coherence time is roughly 200 ms with a central carrier frequency of 2.5 GHz and a user mobility of 2 km/h [30].

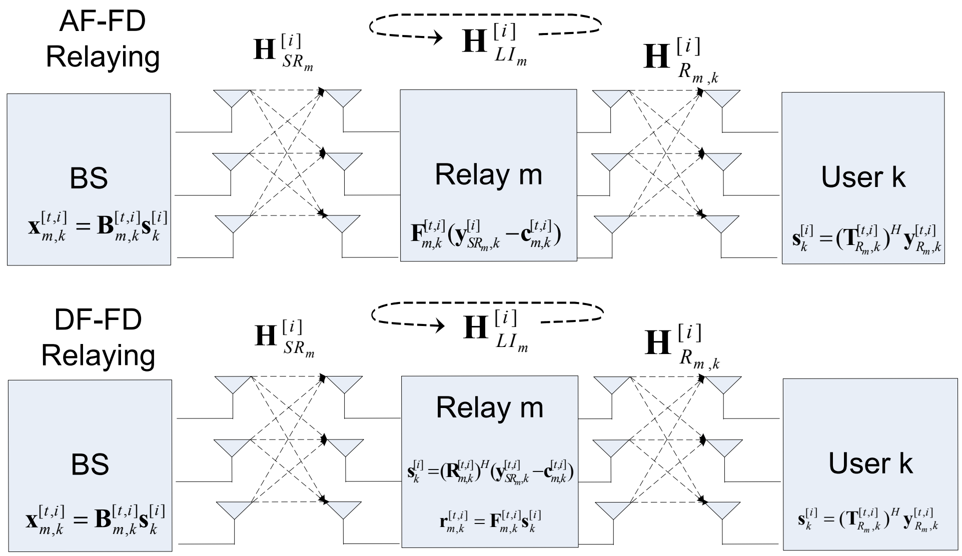


Fig. 2. Block diagrams for AF and DF full-duplex relaying with loop interference cancellation.

where $\mathbf{B}^{[t,i]}_{m,k} \in \mathbb{C}^{N\times N}$ is the transmit precoding matrix on subcarrier i. In the first time slot, if the AF-FD relaying protocol is used for transmission, i.e., $t = 3$, the (frequency domain) received vector symbol on subcarrier i at relay $m \in \{1, \ldots, M\}$ for user k is given by

$$\mathbf{y}^{[i]}_{SR_m,k} = \mathbf{H}^{[i]}_{SR_m}\mathbf{x}^{[t,i]}_{m,k} + \mathbf{H}^{[i]}_{LI_m}\mathbf{F}^{[t,i]}_{m,k}\mathbf{q}^{[t,i]}_{m,k} + \mathbf{z}^{[i]}_m, \quad (2)$$

where $\mathbf{H}^{[i]}_{SR_m}$ is the $N \times N$ MIMO channel matrix between the BS and relay m on subcarrier i and captures the effect of both multi-path fading and path loss, $\mathbf{H}^{[i]}_{LI_m}$ is the $N \times N$ loop interference channel matrix, and $\mathbf{F}^{[t,i]}_{m,k} \in \mathbb{C}^{N\times N}$ is a post-processing matrix used at relay m. $\mathbf{q}^{[t,i]}_{m,k}$ is the $N \times 1$ accumulated loop interference signal vector at relay m on subcarrier i caused by AF-FD relaying. $\mathbf{z}^{[i]}_m \in \mathbb{C}^{N\times 1}$ is the additive white Gaussian noise (AWGN) vector with distribution $\mathcal{CN}(0, \mathbf{\Sigma_m})$ on subcarrier i at relay m, where $\mathbf{\Sigma_m} \in (\mathbb{R}^+_0)^{N\times N}$ is a diagonal covariance matrix with each main diagonal element equal to N_0. The relay subtracts vector $\mathbf{c}^{[t,i]}_{m,k} = \widehat{\mathbf{H}}^{[i]}_{LI_{m,k}}\mathbf{F}^{[t,i]}_{m,k}\mathbf{q}^{[t,i]}_{m,k}$ from $\mathbf{y}^{[i]}_{SR_m,k}$ for loop interference cancellation which yields

$$\begin{aligned}\tilde{\mathbf{y}}^{[i]}_{SR_m,k} &= \mathbf{y}^{[i]}_{SR_m,k} - \mathbf{c}^{[t,i]}_{m,k}\\ &= \mathbf{H}^{[i]}_{SR_m}\mathbf{x}^{[t,i]}_{m,k} + \Delta\mathbf{H}^{[i]}_{LI_m}\mathbf{F}^{[t,i]}_{m,k}\mathbf{q}^{[t,i]}_{m,k} + \mathbf{z}^{[i]}_m, \quad (3)\end{aligned}$$

where $\widehat{\mathbf{H}}^{[i]}_{LI_{m,k}}$ and $\Delta\mathbf{H}^{[i]}_{LI_m} \sim \mathcal{CN}(0, \mathbf{\Xi}^{[i]}_m)$ denote, respectively, the estimated loop interference channel and the corresponding estimation error, which are mutually uncorrelated. $\mathbf{\Xi}^{[i]}_m$ is a diagonal covariance matrix with each main diagonal element equal to σ^2_e.

In the second time slot, relay m multiplies the received signal vector on subcarrier i by $\mathbf{F}^{[t,i]}_{m,k}$ and forwards the processed signal vector to user k on subcarrier[3] i. Then, the signal vector received at user k on subcarrier i from relay m using AF-FD relaying is given by

$$\begin{aligned}&\mathbf{y}^{[t,i]}_{R_m,k}\\ &= \mathbf{H}^{[i]}_{R_m,k}\mathbf{F}^{[t,i]}_{m,k}\Big(\mathbf{H}^{[i]}_{SR_m}\mathbf{x}^{[t,i]} + \Delta\mathbf{H}^{[i]}_{LI_m}\mathbf{F}^{[t,i]}_{m,k}\mathbf{q}^{[t,i]}_{m,k} + \mathbf{z}^{[i]}_m\Big) + \mathbf{n}^{[i]}_k\\ &= \underbrace{\mathbf{H}^{[i]}_{R_m,k}\mathbf{F}^{[t,i]}_{m,k}\mathbf{H}^{[i]}_{SR_m}\mathbf{B}^{[t,i]}_{m,k}\mathbf{s}^{[i]}_k}_{\text{Desired signal}}\\ &+ \underbrace{\mathbf{H}^{[i]}_{R_m,k}\mathbf{F}^{[t,i]}_{m,k}\Delta\mathbf{H}^{[i]}_{LI_m}\mathbf{F}^{[t,i]}_{m,k}\mathbf{q}^{[t,i]}_{m,k}}_{\text{Amplified loop interference}} + \underbrace{\mathbf{H}^{[i]}_{R_m,k}\mathbf{F}^{[t,i]}_{m,k}\mathbf{z}^{[i]}_m}_{\text{Amplified noise}} + \mathbf{n}^{[i]}_k, \quad (4)\end{aligned}$$

where $\mathbf{n}^{[i]}_k \in \mathbb{C}^{N\times 1}$ is the AWGN vector at user k on subcarrier i with distribution $\mathcal{CN}(0, \mathbf{\Sigma_k})$. $\mathbf{\Sigma_k} \in (\mathbb{R}^+_0)^{N\times N}$ is a diagonal matrix with each main diagonal element equal to N_0. In order to simplify the subsequent mathematical expressions and without loss of generality, we assume in the following a normalized noise variance of $N_0 = 1$ at all receive antennas of the relays and the users. Assuming a linear receiver is employed at user k, the estimated data vector symbol $\hat{\mathbf{s}}^{[i]}_k \in \mathbb{C}^{N\times 1}$ on subcarrier i is given by

$$\hat{\mathbf{s}}^{[i]}_k = (\mathbf{T}^{[t,i]}_{m,k})^H\mathbf{y}^{[t,i]}_{R_m,k}, \quad (5)$$

where $\mathbf{T}^{[t,i]}_{m,k} \in \mathbb{C}^{N\times N}$ is a post-processing matrix used for subcarrier i at user k for transmission strategy t. Assuming $\mathcal{E}\{\mathbf{s}^{[i]}_k(\mathbf{s}^{[i]}_k)^H\} = \mathbf{I}_N$, the mean square error (MSE) matrix for the transmission on subcarrier i via relay m at user k is given by

$$\begin{aligned}\mathbf{E}^{[t,i]}_{m,k} &= \mathcal{E}\{(\hat{\mathbf{s}}^{[i]}_k - \mathbf{s}^{[i]}_k)(\hat{\mathbf{s}}^{[i]}_k - \mathbf{s}^{[i]}_k)^H\}\\ &= (\mathbf{T}^{[t,i]}_{m,k}\mathbf{\Gamma}^{[t,i]}_{m,k} - \mathbf{I}_N)(\mathbf{T}^{[t,i]}_{m,k}\mathbf{\Gamma}^{[t,i]}_{m,k} - \mathbf{I}_N)^H\\ &+ (\mathbf{T}^{[t,i]}_{m,k})^H\mathbf{\Theta}^{[t,i]}_{m,k}(\mathbf{T}^{[t,i]}_{m,k}), \quad (6)\end{aligned}$$

where $\mathbf{\Gamma}^{[t,i]}_{m,k}$ is the effective MIMO channel matrix between the BS and user k via relay m on subcarrier i using transmission strategy t, and $\mathbf{\Theta}^{[t,i]}_{m,k}$ is the corresponding equivalent noise

[3]Note that subcarrier pairing is not considered in this paper since it would increase the computational complexity of the resource allocation algorithm. The interested reader is referred to [5] and [31] for different approaches to optimal subcarrier pairing.

$$C_{m,k}^{[t,i]}=\begin{cases}\min\Big\{-\log_2\Big(\det[\mathbf{\Delta}_{m,k}^{*[1,i]}]\Big),-\log_2\Big(\det[\mathbf{E}_{m,k}^{*[1,i]}]\Big)\Big\}, & \text{for } t=1;\ \text{DF-FD}\\ \frac{1}{2}\min\Big\{-\log_2\Big(\det[\mathbf{\Delta}_{m,k}^{*[2,i]}]\Big),-\log_2\Big(\det[\mathbf{E}_{m,k}^{*[2,i]}]\Big)\Big\}, & \text{for } t=2;\ \text{DF-HD}\\ -\log_2\Big(\det[\mathbf{E}_{m,k}^{*[3,i]}]\Big), & \text{for } t=3;\ \text{AF-FD}\\ -\frac{1}{2}\log_2\Big(\det[\mathbf{E}_{m,k}^{*[4,i]}]\Big), & \text{for } t=4;\ \text{AF-HD}\end{cases} \tag{18}$$

covariance matrix. These matrices are given by

$$\begin{aligned}\mathbf{\Gamma}_{m,k}^{[t,i]} &= \mathbf{H}_{R_m,k}^{[i]}\mathbf{F}_{m,k}^{[t,i]}\mathbf{H}_{SR_m}^{[i]}\mathbf{B}_{m,k}^{[t,i]} \quad \text{and}\\ \mathbf{\Theta}_{m,k}^{[t,i]} &= \Big(\mathbf{H}_{R_m,k}^{[i]}\mathbf{F}_{m,k}^{[t,i]}\Big)\mathbf{K}_{m,k}^{[t,i]}\Big(\mathbf{H}_{R_m,k}^{[i]}\mathbf{F}_{m,k}^{[t,i]}\Big)^H+\mathbf{I}_N \text{ with}\\ \mathbf{K}_{m,k}^{[t,i]} &= \mathcal{E}\{(\Delta\mathbf{H}_{LI_m}^{[i]}\mathbf{F}_{m,k}^{[t,i]}\mathbf{q}_{m,k}^{[t,i]})(\Delta\mathbf{H}_{LI_m}^{[i]}\mathbf{F}_{m,k}^{[t,i]}\mathbf{q}_{m,k}^{[t,i]})^H\}+\mathbf{I}_N\\ &= \mathbf{L}_{m,k}^{[t,i]}+\mathbf{I}_N,\end{aligned} \tag{7}$$

where $\mathbf{L}_{m,k}^{[t,i]}=\mathcal{E}\{(\Delta\mathbf{H}_{LI_m}^{[i]}\mathbf{F}_{m,k}^{[t,i]}\mathbf{q}_{m,k}^{[t,i]})(\Delta\mathbf{H}_{LI_m}^{[i]}\mathbf{F}_{m,k}^{[t,i]}\mathbf{q}_{m,k}^{[t,i]})^H\}$ denotes the interference cancellation error matrix. Thus, the optimal post-processing matrix which minimizes the MSE for transmission strategy $t=3$ and the corresponding minimum MSE (MMSE) matrix are given by

$$\begin{aligned}\mathbf{T}_{R_m,k}^{*[t,i]} &= (\mathbf{\Gamma}_{m,k}^{[t,i]}(\mathbf{\Gamma}_{m,k}^{[t,i]})^H+\mathbf{\Theta}_{m,k}^{[t,i]})^{-1}\mathbf{\Gamma}_{m,k}^{[t,i]} \quad \text{and}\\ \mathbf{E}_{m,k}^{*[t,i]} &= \Big[\mathbf{I}_N+(\mathbf{\Gamma}_{m,k}^{[t,i]})^H(\mathbf{\Theta}_{m,k}^{[t,i]})^{-1}(\mathbf{\Gamma}_{m,k}^{[t,i]})\Big]^{-1},\end{aligned} \tag{8}$$

respectively. If AF-HD relaying is selected on subcarrier i, equations (1)-(8) are still valid if we set $t=4$ and $\sigma_e^2=0$.

D. Decode-and-Forward Full-Duplex Relaying Channel Model

If the DF-FD relaying protocol is selected on subcarrier i, i.e., $t=1$, the received signal vector after interference cancellation at relay m for users k on subcarrier i in the first time slot is given by

$$\begin{aligned}\tilde{\mathbf{y}}_{SR_m,k}^{[t,i]} &= \mathbf{y}_{SR_m,k}^{[t,i]}-\mathbf{c}_{m,k}^{[t,i]}\\ &= \mathbf{H}_{SR_m}^{[i]}\mathbf{x}_{m,k}^{[t,i]}+\Delta\mathbf{H}_{LI_m}^{[i]}\mathbf{r}_{m,k}^{[t,i]}+\mathbf{z}_m^{[i]},\end{aligned} \tag{9}$$

where $\mathbf{r}_{m,k}^{[t,i]}$ is the concurrent transmitted signal vector from relay m to user k when DF-FD relaying is selected and $\mathbf{c}_{m,k}^{[t,i]}=\hat{\mathbf{H}}_{LI_{m,k}}^{[i]}\mathbf{F}_{m,k}^{[t,i]}\mathbf{r}_{m,k}^{[t,i]}$. The received signal vector in the first hop on subcarrier i at relay m is multiplied by an intermediate post-processing matrix $\mathbf{R}_{m,k}^{[t,i]}$ for extracting the original signal vector. The estimated vector symbol at the relay, $\tilde{\mathbf{s}}_k^{[i]}\in\mathbb{C}^{N\times 1}$, is given by

$$\tilde{\mathbf{s}}_k^{[i]}=(\mathbf{R}_{m,k}^{[t,i]})^H\tilde{\mathbf{y}}_{SR_m,k}^{[t,i]}. \tag{10}$$

Thus, the optimal MMSE post-processing matrix and the corresponding MMSE matrix at relay m can be written as

$$\begin{aligned}\mathbf{R}_{m,k}^{*[t,i]} &= (\mathbf{\Phi}_{m,k}^{[t,i]}(\mathbf{\Phi}_{m,k}^{[t,i]})^H+\mathbf{\Upsilon}_{m,k}^{[t,i]})^{-1}\mathbf{\Phi}_{m,k}^{[t,i]} \text{ and}\\ \mathbf{\Delta}_{m,k}^{*[t,i]} &= \Big[\mathbf{I}_N+(\mathbf{\Phi}_{m,k}^{[t,i]})^H(\mathbf{\Upsilon}_{m,k}^{[t,i]})^{-1}(\mathbf{\Phi}_{m,k}^{[t,i]})\Big]^{-1},\end{aligned} \tag{11}$$

respectively. $\mathbf{\Phi}_{m,k}^{[t,i]}$ and $\mathbf{\Upsilon}_{m,k}^{[t,i]}$ are the equivalent MIMO channel matrix of the BS-relay links for DF relaying on subcarrier i and the noise covariance matrix at relay m on subcarrier i given by

$$\begin{aligned}\mathbf{\Phi}_{m,k}^{[t,i]} &= \mathbf{H}_{SR_m}^{[i]}\mathbf{B}_{m,k}^{[t,i]} \quad \text{and}\\ \mathbf{\Upsilon}_{m,k}^{[t,i]} &= \mathcal{E}\Big\{\Big(\Delta\mathbf{H}_{LI_m}^{[i]}\mathbf{r}_{m,k}^{[t,i]}\Big)\Big(\Delta\mathbf{H}_{LI_m}^{[i]}\mathbf{r}_{m,k}^{[t,i]}\Big)^H\Big\}+\mathbf{I}_N.\end{aligned} \tag{12}$$

Assuming error-free decoding[4] at relay m, in the second time slot, the decoded signal vector at relay m on subcarrier i is encoded again with a new precoding matrix $\mathbf{F}_{m,k}^{[t,i]}$ and forwarded to user k. The re-encoded signal vector is given by

$$\mathbf{r}_{m,k}^{[t,i]}=\mathbf{F}_{m,k}^{[t,i]}\mathbf{s}_k^{[i]} \tag{13}$$

and the received signal vector at user k from the BS via relay m on subcarrier i is given by

$$\mathbf{y}_{R_m,k}^{[t,i]} = \mathbf{H}_{R_m,k}^{[i]}\mathbf{r}_{m,k}^{[t,i]}+\mathbf{z}_k^{[i]}. \tag{14}$$

Similar to AF-FD relaying, user k multiplies the received signal vector with a post-processing matrix to extract the original signal vector. The estimated symbol vector $\hat{\mathbf{s}}_k^{[i]}$ at user k on subcarrier i can be expressed as

$$\hat{\mathbf{s}}_k^{[i]}=(\mathbf{T}_{m,k}^{[t,i]})^H\mathbf{y}_{R_m,k}^{[t,i]}. \tag{15}$$

Then, the optimal linear receiver post-processing matrix is given by

$$\mathbf{T}_{R_m,k}^{*[t,i]}=(\mathbf{H}_{R_m,k}^{[i]}\mathbf{F}_{m,k}^{[t,i]}(\mathbf{H}_{R_m,k}^{[i]}\mathbf{F}_{m,k}^{[t,i]})^H+\mathbf{I}_N)^{-1}\mathbf{H}_{R_m,k}^{[i]}\mathbf{F}_{m,k}^{[t,i]} \tag{16}$$

with the corresponding MMSE matrix

$$\mathbf{E}_{m,k}^{*[t,i]}=\Big[\mathbf{I}_N+(\mathbf{H}_{R_m,k}^{[i]}\mathbf{F}_{m,k}^{[t,i]})^H(\mathbf{H}_{R_m,k}^{[i]}\mathbf{F}_{m,k}^{[t,i]})\Big]^{-1}. \tag{17}$$

The signal model for DF-HD relaying is obtained by setting in (9)-(17) $t=2$ and $\sigma_e^2=0$.

III. Resource Allocation and Scheduling Design

A. Instantaneous Channel Capacity and System Throughput

Given perfect CSI at the receiver (CSIR), the channel capacity $C_{m,k}^{[t,i]}$ between the BS and user k via relay m on subcarrier i are function of MSEs of the adopted transmission strategies [33], [34] and are given by (18) at the top of this page where the pre-log factor $\frac{1}{2}$ in (18) for $t=2$ and $t=4$ is due to the two channel uses necessary for transmitting one packet in HD relaying.

[4]This assumption is justified since the data rate of the scheduled users is always less than the instantaneous channel capacity. Thus, an arbitrarily small error probability can be achieved with a powerful error correcting code [32].

Now, we define the instantaneous throughput (bit/s/Hz successfully delivered) for user k who is served by relay m as

$$\rho_m^{(k)} = \frac{1}{n_F} \sum_{i=1}^{n_F} \sum_{t=1}^{4} s_{m,k}^{[t,i]} C_{m,k}^{[t,i]}, \quad (19)$$

where $s_{m,k}^{[t,i]} \in \{0,1\}$ is the subcarrier allocation indicator. The *average weighted system throughput* is defined as the total average number of bit/s/Hz successfully decoded at the K users via the M relays and given by

$$U_{TP}(\mathcal{P},\mathcal{S}) = \sum_{m=1}^{M} \sum_{k\in\mathcal{U}_m} w_k \rho_m^{(k)}, \quad (20)$$

where $\mathcal{P} = \{\mathbf{B}_{m,k}^{[t,i]}, \mathbf{F}_{m,k}^{[t,i]}\}$ and $\mathcal{S}$ are the precoding matrices and subcarrier allocation policies, respectively, $\mathcal{U}_m$ is the set of users served by relay m, and w_k is a positive constant, which is specified in the media access control (MAC) layer and allows the resource allocator to prioritize the users so as to achieve certain fairness objectives.

B. Problem Formulation

The optimal precoding matrices, $\mathcal{P}^* = \{\mathbf{B}_{m,k}^{[t,i]*}, \mathbf{F}_{m,k}^{[t,i]*}\}$, and subcarrier allocation policy, $\mathcal{S}^*$, are given by

$$\begin{aligned}
&(\mathcal{P}^*,\mathcal{S}^*) = \arg\max_{\mathcal{P},\mathcal{S}} \quad U_{TP}(\mathcal{P},\mathcal{S}) \\
\text{s.t.}\quad &\text{C1:} \quad \sum_{m=1}^{M}\sum_{k\in\mathcal{U}_m}\sum_{t=1}^{4}\sum_{i=1}^{n_F} s_{m,k}^{[t,i]} \Big(\text{tr}\Big(\mathbf{B}_{m,k}^{[t,i]}(\mathbf{B}_{m,k}^{[t,i]})^H\Big) \\
&\quad + \quad \text{tr}\Big(\mathbf{G}_{m,k}^{[t,i]}\Big)\Big) \le P_T, \\
&\text{C2:} \quad \rho_m^{(k)} \ge R^{(k)} \quad \forall k \in \mathcal{D}_m, \\
&\text{C3:} \quad \sum_{m=1}^{M}\sum_{k\in\mathcal{U}_m}\sum_{t=1}^{4} s_{m,k}^{[t,i]} \le 1 \quad \forall i, \\
&\text{C4:} \quad s_{m,k}^{[t,i]} \in \{0,1\} \quad \forall m,i,t,k, \\
&\text{C5:} \quad \mathbf{B}_{m,k}^{[t,i]}, \mathbf{F}_{m,k}^{[t,i]} \succeq 0 \quad \forall m,i,k,t, \qquad (21)
\end{aligned}$$

where $\text{tr}\Big(\mathbf{G}_{m,k}^{[t,i]}\Big)$ is the total power transmitted by relay m on subcarrier i for user k using transmission strategy t and is given by

$$\begin{aligned}
\text{tr}\Big(\mathbf{G}_{m,k}^{[t,i]}\Big) &= \text{tr}\Big(\mathbf{F}_{m,k}^{[t,i]}\Big(\mathbf{H}_{SR_m}^{[i]}\mathbf{B}_{m,k}^{[t,i]}(\mathbf{B}_{m,k}^{[t,i]})^H(\mathbf{H}_{SR_m}^{[i]})^H \\
&\quad +\mathbf{K}_{m,k}^{[t,i]}\Big)(\mathbf{F}_{m,k}^{[t,i]})^H\Big) \quad \text{for AF} \quad \text{and} \\
\text{tr}\Big(\mathbf{G}_{m,k}^{[t,i]}\Big) &= \text{tr}\Big(\mathbf{F}_{m,k}^{[t,i]}(\mathbf{F}_{m,k}^{[t,i]})^H\Big) \quad \text{for DF} \qquad (22)
\end{aligned}$$

where $\mathbf{K}_{m,k}^{[t,i]} = \mathbf{I}_N$ for AF-HD relaying, i.e., $t=4$. In (21), $\mathcal{D}_m$ is the set of *delay sensitive* users associated with relay m. C1 is a joint power constraint for the BS and the relays with total maximum power P_T. Although in a practical system the BS and the relays have separate power supplies, a joint power optimization provides useful insight into the power usage of the whole system rather than the per hop required power. Moreover, for separate power constraints for the BS and the relays, obtaining a globally optimal solution in polynomial time does not seem possible due to the non-convexity of the problem. C2 specifies the minimum required data rate, $R^{(k)}$, for *delay sensitive* users which are chosen by the application layer. Constraints C3 and C4 are imposed to guarantee that each subcarrier is only used by one user with one transmission strategy. In other words, intra-cell interference due to multiple access does not exist in the system. C5 is due to the fact that the precoding matrices $\mathbf{B}_{m,k}^{[t,i]}$ and $\mathbf{F}_{m,k}^{[t,i]}$ must be positive semidefinite.

C. Transformation of the Optimization Problem

In general, the considered problem is a mixed combinatorial and non-convex optimization problem. The combinatorial nature comes from the integer constraint for subcarrier allocation while the non-convexity is caused by the precoding matrices in the objective function. Thus, a brute force approach is needed to obtain the global optimal solution. However, such method is computationally infeasible for a large system and does not provide useful system design insight.

In order to obtain an insightful solution for scheduling and resource allocation purposes, we first define the following matrices before stating an important theorem and proposition. By singular value decomposition (SVD), matrices $\mathbf{H}_{SR_m}^{[i]}$ and $\mathbf{H}_{R_m,k}^{[i]}$ can be written as

$$\begin{aligned}
\mathbf{H}_{SR_m}^{[i]} &= \mathbf{U}_{SR_m}^{[i]}\mathbf{\Lambda}_{SR_m}^{[i]}(\mathbf{V}_{SR_m}^{[i]})^H \text{ and} \\
\mathbf{H}_{R_m,k}^{[i]} &= \mathbf{U}_{R_m,k}^{[i]}\mathbf{\Lambda}_{R_m,k}^{[i]}(\mathbf{V}_{R_m,k}^{[i]})^H, \qquad (23)
\end{aligned}$$

respectively, where $\mathbf{U}_{SR_m}^{[i]} \in \mathbb{C}^{N\times N}, \mathbf{V}_{SR_m}^{[i]} \in \mathbb{C}^{N\times N}, \mathbf{U}_{R_m,k}^{[i]} \in \mathbb{C}^{N\times N}$, and $\mathbf{V}_{R_m,k}^{[i]} \in \mathbb{C}^{N\times N}$ are unitary matrices. $\mathbf{\Lambda}_{R_m,k}^{[i]} \in (\mathbb{R}_0^+)^{N\times N}$ and $\mathbf{\Lambda}_{SR_m}^{[i]} \in (\mathbb{R}_0^+)^{N\times N}$ are diagonal matrices with diagonal element vectors $\text{diag}(\mathbf{\Lambda}_{SR_m}^{[i]}) = \Big[\sqrt{\gamma_{SR_m,1}^{[i]}}\sqrt{\gamma_{SR_m,2}^{[i]}}\cdots\sqrt{\gamma_{SR_m,N}^{[i]}}\Big]$ and $\text{diag}(\mathbf{\Lambda}_{R_m,k}^{[i]}) = \Big[\sqrt{\gamma_{RD_m,k,1}^{[i]}}\sqrt{\gamma_{RD_m,k,2}^{[i]}}\ \cdots\sqrt{\gamma_{RD_m,k,N}^{[i]}}\Big]$, respectively. The elements in $\text{diag}(\mathbf{\Lambda}_{SR_m}^{[i]})$ and $\text{diag}(\mathbf{\Lambda}_{R_m,k}^{[i]})$ are assumed to be arranged in ascending order. Variables $\gamma_{SR_m,n}^{[i]}$ and $\gamma_{RD_m,k,n}^{[i]}$ represent the equivalent signal-to-noise ratio (SNR) on spatial channel n in subcarrier i of the BS-to-relay m link and the relay m-to-user k link, respectively. We are now ready to introduce the following theorem.

Theorem 1: Assume that[5] $\text{rank}(\mathbf{B}_{m,k}^{[t,i]}) = \text{rank}(\mathbf{F}_{m,k}^{[t,i]}) = \text{rank}(\mathbf{H}_{SR_m}^{[i]}) = \text{rank}(\mathbf{H}_{R_m,k}^{[i]}) = N$. For both MIMO AF-FD relaying and MIMO DF-FD relaying with linear MMSE receivers implemented at the users, if the loop interference cancellation errors are spatially uncorrelated[6], the optimal linear precoding matrices used at the BS and relay jointly diagonalize the BS-relay-user channels on each subcarrier. The optimal precoding matrices $\mathbf{B}_{m,k}^{[t,i]}$ and $\mathbf{F}_{m,k}^{[t,i]}$ are given by

$$\begin{aligned}
\mathbf{B}_{m,k}^{[t,i]} &= \mathbf{V}_{SR_m}^{[i]}\mathbf{\Lambda}_{B_m,k}^{[t,i]} \text{ and} \\
\mathbf{F}_{m,k}^{[t,i]} &= \begin{cases} \mathbf{V}_{R_m,k}^{[i]}\mathbf{\Lambda}_{F_m,k}^{[t,i]}(\mathbf{U}_{SR_m}^{[i]})^H, & \text{for } t=3,4 \\ \mathbf{V}_{R_m,k}^{[i]}\mathbf{\Lambda}_{F_m,k}^{[t,i]}, & \text{otherwise,} \end{cases} \qquad (24)
\end{aligned}$$

[5]In practice, the full rank assumption can be achieved by placing the antennas sufficiently far apart such that all antennas experience uncorrelated fading.

[6]If the cancellation errors are spatially correlated, i.e., $\mathbf{L}_{m,k}^{[t,i]}$ in (7) is not a diagonal matrix, we can always perform pre-whitening of the received signals at relay m to decorrelate the loop interference cancellation errors.

respectively. Matrices $\mathbf{\Lambda}_{B_m,k}^{[t,i]} \in \mathbb{C}^{N\times N}$ and $\mathbf{\Lambda}_{F_m,k}^{[t,i]} \in \mathbb{C}^{N\times N}$ are diagonal matrices with diagonal element vectors $\text{diag}(\mathbf{\Lambda}_{B_m,k}^{[t,i]}) = \Big[\sqrt{P_{SR_m,k,1}^{[t,i]}}\sqrt{P_{SR_m,k,2}^{[t,i]}} \cdots \sqrt{P_{SR_m,k,N}^{[t,i]}}\Big]$, and $\text{diag}(\mathbf{\Lambda}_{F_m,k}^{[t,i]}) = \Big[\sqrt{P_{RD_m,k,1}^{[t,i]}}\sqrt{P_{RD_m,k,2}^{[t,i]}} \cdots \sqrt{P_{RD_m,k,N}^{[t,i]}}\Big]$, respectively, which are assumed to be arranged in ascending order. Variables $P_{SR_m,k,n}^{[t,i]}$ and $P_{RD_m,k,n}^{[t,i]}$ are, respectively, the equivalent transmit powers of the BS-to-relay m link and the relay m-to-user k link on spatial channel n in subcarrier i for transmission strategy t.

Proof: Please refer to Appendix A. ■

By Theorem 1, the MIMO FD-relaying channel on subcarrier i is converted into N parallel spatial channels if the optimal precoding and post-processing matrices are used. Therefore, the capacity between the BS and user k via relay m on subcarrier i of the considered transmission strategies can be restated as [5], [17]:

For $t=1$, DF-FD relaying is selected and $C_{m,k}^{[1,i]} =$
$$\sum_{n=1}^{N}\log_2\Big(1+\min\Big\{\frac{P_{SR_m,k,n}^{[1,i]}\gamma_{SR_m,n}^{[i]}}{\gamma_{LI_m,n}^{[i]}P_{RD_m,k,n}^{[1,i]}+1}, P_{RD_m,k,n}^{[1,i]}\gamma_{RD_m,k,n}^{[i]}\Big\}\Big);$$
for $t=2$, DF-HD relaying is selected and $C_{m,k}^{[2,i]} =$
$$\sum_{n=1}^{N}\frac{1}{2}\log_2\Big(1+\min\Big\{\gamma_{SR_m,n}^{[i]}P_{SR_m,k,n}^{[2,i]}, P_{RD_m,k,n}^{[2,i]}\gamma_{RD_m,k,n}^{[i]}\Big\}\Big);$$
for $t=3$, AF-FD relaying is selected and $C_{m,k}^{[3,i]} =$
$$\sum_{n=1}^{N}\log_2\Bigg(1+\frac{\frac{P_{SR_m,k,n}^{[3,i]}\gamma_{SR_m,n}^{[i]}}{P_{RD_m,k,n}^{[3,i]}\gamma_{LI_m,n}^{[i]}+1}P_{RD_m,k,n}^{[3,i]}\gamma_{RD_m,k,n}^{[i]}}{1+P_{RD_m,k,n}^{[3,i]}\gamma_{RD_m,k,n}^{[i]}+\frac{P_{SR_m,k,n}^{[3,i]}\gamma_{SR_m,n}^{[i]}}{1+P_{RD_m,k,n}^{[3,i]}\gamma_{LI_m,n}^{[i]}}}\Bigg);$$
for $t=4$, AF-HD relaying is selected and $C_{m,k}^{[4,i]} =$
$$\sum_{n=1}^{N}\frac{1}{2}\log_2\Bigg(1+\frac{\gamma_{SR_m,n}^{[i]}P_{SR_m,k,n}^{[4,i]}P_{RD_m,k,n}^{[4,i]}\gamma_{RD_m,k,n}^{[i]}}{1+\gamma_{SR_m,n}^{[i]}P_{SR_m,k,n}^{[4,i]}+P_{RD_m,k,n}^{[4,i]}\gamma_{RD_m,k,n}^{[i]}}\Bigg);$$
where variable $\gamma_{LI_m,n}^{[i]}$ denotes the instantaneous residual loop interference power of spatial channel n of relay m on subcarrier i.

Although (25) shows that the considered problem in (21) is now transformed into an equivalent scalar optimization problem, this scalar problem is still non-convex which is an obstacle in deriving an efficient resource allocation algorithm. Therefore, we introduce the following proposition for a further transformation of the scalar problem.

Proposition 1 (Equivalent Capacity): Without loss of generality, we define a new variable $P_{m,k,n}^{[t,i]}$ such that $P_{m,k,n}^{[t,i]} = P_{SR_m,k,n}^{[t,i]} + P_{RD_m,k,n}^{[t,i]}$, which represents the power consumption in subcarrier i on spatial channel n for user k via relay m using relaying protocol t. Let $\Gamma_{m,k,n}^{[t,i]}$ be the corresponding equivalent received SNR or signal-to-interference-plus-noise ratio (SINR) of spatial channel n at user k on subcarrier i via relay m using transmission strategy t. Then, the maximized sum channel capacity on subcarrier i for user k via relay m of each transmission strategy is summarized in Table I.

Proof: Please refer to the Appendix B. ■

Hence, we substitute the capacity equations of the different relaying protocols from Table I into the original objective function in (21), to obtain a new objective function. Note that the above theorem and proposition are the two key steps to simplify the considered problem. Theorem 1 transforms the original matrix optimization problem into a scalar optimization problem, while Proposition 1 transforms the objective function into a concave function with respect to (w.r.t.) the new optimization variables $P_{m,k,n}^{[t,i]}$ for all transmission strategies. The next step is to handle the combinatorial subcarrier assignment constraint. We follow the approach in [35] and relax $s_{m,k}^{[t,i]}$ in constraint C4 to be a real value between zero and one instead of a Boolean. Then, $s_{m,k}^{[t,i]}$ can be interpreted as a *time sharing* factor for the K users to utilize subcarrier i through relay m with different transmission strategies. It can be shown that the relaxation is optimal [36] for large numbers of subcarriers. Thus, the optimization problem can be rewritten as

$$(\mathcal{P}^*,\mathcal{S}^*) = \arg\max_{\mathcal{P},\mathcal{S}} \sum_{m=1}^{M}\sum_{k\in\mathcal{U}_m}\sum_{t=1}^{4}\sum_{i=1}^{n_F} w_k s_{m,k}^{[t,i]}\widetilde{C}_{m,k}^{[t,i]}$$
$$\text{s.t. C1: } \sum_{m=1}^{M}\sum_{k\in\mathcal{U}_m}\sum_{t=1}^{4}\sum_{i=1}^{n_F}\sum_{n=1}^{N}\widetilde{P}_{m,k,n}^{[t,i]} \le P_T,$$
$$\text{C2, C3, C5,}$$
$$\text{C4: } 0\le s_{m,k}^{[t,i]}\le 1 \quad \forall m,i,t,k, \tag{25}$$

where $\widetilde{P}_{m,k,n}^{[t,i]} = P_{m,k,n}^{[t,i]}s_{m,k}^{[t,i]} = \widetilde{P}_{SR_m,k,n}^{[t,i]} + \widetilde{P}_{RD_m,k,n}^{[t,i]}$, $\widetilde{P}_{SR_m,k,n}^{[t,i]} = P_{SR_m,k,n}^{[t,i]}s_{m,k}^{[t,i]}$, and $\widetilde{P}_{RD_m,k,n}^{[t,i]} = P_{RD_m,k,n}^{[t,i]}s_{m,k}^{[t,i]}$ are auxiliary power allocation variables which represent the corresponding transmitted powers for time-sharing. $\widetilde{C}_{m,k}^{[t,i]} = C_{m,k}^{[t,i]}|_{P_{m,k,n}^{[t,i]}=\widetilde{P}_{m,k,n}^{[t,i]}/s_{m,k}^{[t,i]}}$ is the time-shared channel capacity for each transmission strategy. It can be shown that (25) is now jointly concave w.r.t. all optimization variables and the duality gap is equal to zero under some mild conditions [37]. Hence, solving the dual problem is equivalent to solving the original primal problem.

TABLE I
CAPACITY AND EQUIVALENT SNR/SINR OF DIFFERENT RELAYING PROTOCOLS

VARIABLES: $\Psi_{m,k,n}^{[i]} = \gamma_{LI_m,n}^{[i]}\gamma_{RD_m,k,n}^{[i]}\gamma_{SR_m,n}^{[i]}$ AND $\Upsilon_{m,k,n}^{[i]} = \gamma_{SR_m,n}^{[i]} + \gamma_{RD_m,k,n}^{[i]}$.

Transmission Strategy	Capacity Equation $C_{m,k}^{[t,i]}$
$t=1$, DF-FD relaying	$\sum_{n=1}^{N}\log_2\left(1+\Gamma_{m,k,n}^{[1,i]}\right)$
$t=2$, DF-HD relaying	$\sum_{n=1}^{N}\frac{1}{2}\log_2\left(1+\Gamma_{m,k,n}^{[2,i]}\right)$
$t=3$, AF-FD relaying	$\sum_{n=1}^{N}\log_2\left(1+\Gamma_{m,k,n}^{[3,i]}\right)$
$t=4$, AF-HD relaying	$\sum_{n=1}^{N}\frac{1}{2}\log_2\left(1+\Gamma_{m,k,n}^{[4,i]}\right)$
Transmission Strategy	**Equivalent SNR/SINR $\Gamma_{m,k,n}^{[t,i]}$**
$t=1$, DF-FD relaying	$\frac{\sqrt{(\Upsilon_{m,k,n}^{[i]})^2+4\Psi_{m,k,n}^{[i]}P_{m,k,n}^{[1,i]}}-\Upsilon_{m,k,n}^{[i]}}{2\gamma_{LI_m,n}^{[i]}}$
$t=2$, DF-HD relaying	$P_{m,k,n}^{[2,i]}\frac{\gamma_{SR_m,n}^{[i]}\gamma_{RD_m,k,n}^{[i]}}{\Upsilon_{m,k,n}^{[i]}}$
$t=3$, AF-FD relaying	$\frac{\sqrt{\gamma_{RD_m,k,n}^{[i]}\gamma_{SR_m,n}^{[i]}P_{m,k,n}^{[3,i]}}}{2\sqrt{\gamma_{LI_m,n}^{[i]}}}$
$t=4$, AF-HD relaying	$\frac{P_{m,k,n}^{[4,i]}\gamma_{SR_m,n}^{[i]}\gamma_{RD_m,k,n}^{[i]}}{(\sqrt{\gamma_{SR_m,n}^{[i]}}+\sqrt{\gamma_{RD_m,k,n}^{[i]}})^2}$

TABLE II
OPTIMAL WATER-FILLING POWER ALLOCATION

VARIABLES: $\Theta^{[i]}_{m,k,n} = \sqrt{\frac{\lambda \ln(2)}{(w_k+\delta_k)}(8\Psi^{[i]}_{m,k,n} + \frac{\lambda \ln(2)}{(w_k+\delta_k)} Y^{[i]}_{m,k,n})}$, $Y^{[i]}_{m,k,n} = (\Upsilon^{[i]}_{m,k,n} - 2\gamma^{[i]}_{LI_m,n})^2$, $\Omega^{[i]}_{m,k,n} = \frac{\lambda}{(w_k+\delta_k)} \ln(2)(2\gamma^{[i]}_{LI_m,n} - \Upsilon^{[i]}_{m,k,n})$, $\Upsilon^{[i]}_{m,k,n} = \gamma^{[i]}_{SR_m,n} + \gamma^{[i]}_{RD_m,k,n}$, AND $\Phi^{[i]}_{m,k,n} = \sqrt{4(\frac{\lambda}{(w_k+\delta_k)})^2 \ln^2(2)(\gamma^{[i]}_{LI_m,n})^2 + \frac{2\ln(2)\lambda}{(w_k+\delta_k)} \Psi^{[i]}_{m,k,n}}$.

Transmission Strategy	Optimal $P^{[t,i]*}_{m,k,n}$	Optimal $P^{[t,i]*}_{R_m,k,n}$
$t=1$	$\Big[\frac{4\Psi^{[i]}_{m,k,n}}{(\Theta^{[i]}_{m,k,n}+\Omega^{[i]}_{m,k,n})^2} - \frac{(\Upsilon^{[i]}_{m,k,n})^2}{4\Psi^{[i]}_{m,k,n}}\Big]^+$	$\Big[\frac{\Upsilon^{[i]}_{m,k,n} - \sqrt{4\Psi^{[i]}_{m,k,n} P^{[1,i]*}_{m,k,n} + (\Upsilon^{[i]}_{m,k,n})^2}}{2\gamma^{[i]}_{LI_m,n}\gamma^{[i]}_{RD_m,k,n}}\Big]^+$
$t=2$	$\Big[\frac{w_k+\delta_k}{2\lambda \ln(2)} - \frac{\gamma^{[i]}_{SR_m,n}+\gamma^{[i]}_{RD_m,k,n}}{\gamma^{[i]}_{SR_m,n}\gamma^{[i]}_{RD_m,k,n}}\Big]^+$	$\Big[\frac{(w_k+\delta_k)\gamma^{[i]}_{SR_m,n}}{2\lambda\ln(2)\big(\gamma^{[i]}_{RD_m,k,n}+\gamma^{[i]}_{SR_m,n}\big)} - \frac{1}{\gamma^{[i]}_{RD_m,k,n}}\Big]^+$
$t=3$	$\frac{\Psi^{[i]}_{m,k,n}}{\big(2\lambda/(w_k+\delta_k)\ln(2)\gamma^{[i]}_{LI_m,n}+\Phi^{[i]}_{m,k,n}\big)^2}$	$\sqrt{\frac{P^{[3,i]*}_{m,k,n}\gamma^{[i]}_{SR_m,n}}{\gamma^{[i]}_{LI_m,n}\gamma^{[i]}_{RD_m,k,n}}}$
$t=4$	$\Big[\frac{w_k+\delta_k}{2\lambda \ln(2)} - \frac{\Big(\sqrt{\gamma^{[i]}_{SR_m,n}}+\sqrt{\gamma^{[i]}_{RD_m,k,n}}\Big)^2}{\gamma^{[i]}_{SR_m,n}\gamma^{[i]}_{RD_m,k,n}}\Big]^+$	$\Big[\frac{\gamma^{[i]}_{SR_m,n} - \sqrt{\gamma^{[i]}_{SR_m,n}\gamma^{[i]}_{RD_m,k,n}}}{\gamma^{[i]}_{SR_m,n}-\gamma^{[i]}_{RD_m,k,n}}\Big]^+$

D. Dual Problem Formulation

In this subsection, we formulate the dual problem of the transformed optimization problem. For this purpose, we first need the Lagrangian function of the primal problem. Upon rearranging terms, the Lagrangian can be written as

$$\begin{aligned}
&\mathcal{L}(\lambda, \boldsymbol{\delta}, \boldsymbol{\beta}, \mathcal{P}, \mathcal{S}) \\
&= \sum_{m=1}^{M} \sum_{k\in\mathcal{U}_m} \sum_{i=1}^{n_F} \sum_{t=1}^{4} (w_k+\delta_k) s^{[t,i]}_{m,k} \widetilde{C}^{[t,i]}_{m,k} \\
&-\lambda \sum_{m=1}^{M} \sum_{k\in\mathcal{U}_m} \sum_{i=1}^{n_F} \sum_{t=1}^{4} \widetilde{P}^{[t,i]}_{m,k,n} - \sum_{k=1}^{K} R^{(k)}\delta_k \\
&- \sum_{m=1}^{M} \sum_{k\in\mathcal{U}_m} \sum_{i=1}^{n_F} \sum_{t=1}^{4} \beta_i s^{[t,i]}_{m,k} + \lambda P_T + \sum_{i=1}^{n_F} \beta_i, \qquad (26)
\end{aligned}$$

where λ is the Lagrange multiplier corresponding to the joint power constraint and $\boldsymbol{\beta}$ with elements β_i, $i \in \{1, \ldots, n_F\}$, is the Lagrange multiplier vector associated with the subcarrier usage constraints. $\boldsymbol{\delta}$ with elements δ_k, $k \in \{1, \ldots, K\}$, is the Lagrange multiplier vector corresponding to the data rate constraint. Note that $\delta_k = 0$ for *non-delay sensitive* users, i.e., $k \notin \mathcal{D}$. The boundary constraints C4 and C5 in (25) will be absorbed into the Karush-Kuhn-Tucker (KKT) conditions when deriving the optimal solution in Section III-E.

Thus, the dual problem is given by

$$\min_{\lambda,\boldsymbol{\delta},\boldsymbol{\beta}\geq 0} \max_{\mathcal{P},\mathcal{S}} \quad \mathcal{L}(\lambda, \boldsymbol{\delta}, \boldsymbol{\beta}, \mathcal{P}, \mathcal{S}). \qquad (27)$$

E. Distributed Solution - Subproblem for Each Relay

By dual decomposition, the dual problem in (27) can be decomposed into a master problem and several subproblems. The dual problem can be solved iteratively where in each iteration each relay solves its own subproblem without using CSI from the other relays and passes its local solution to the BS for solving the master problem. The subproblem to be solved at relay m is given by

$$\begin{aligned}
&\max_{\mathcal{P},\mathcal{S}} \quad \mathcal{L}_m(\lambda, \boldsymbol{\delta}, \boldsymbol{\beta}, \mathcal{P}, \mathcal{S}) \quad \text{with} \quad \mathcal{L}_m(\lambda, \boldsymbol{\delta}, \boldsymbol{\beta}, \mathcal{P}, \mathcal{S}) \\
&= \sum_{k\in\mathcal{U}_m} \sum_{i=1}^{n_F} \sum_{t=1}^{4} (w_k+\delta_k) s^{[t,i]}_{m,k} \widetilde{C}^{[t,i]}_{m,k} \\
&-\lambda \sum_{k\in\mathcal{U}_m} \sum_{i=1}^{n_F} \sum_{n=1}^{N} \sum_{t=1}^{4} \widetilde{P}^{[t,i]}_{m,k,n} - \sum_{k\in\mathcal{U}_m} \sum_{i=1}^{n_F} \sum_{t=1}^{4} \beta_i s^{[t,i]}_{m,k}, \qquad (28)
\end{aligned}$$

where the Lagrange multipliers λ, $\boldsymbol{\delta}$, and $\boldsymbol{\beta}$ are provided by the BS.

Using standard optimization techniques and the KKT conditions, the optimal power allocation for subcarrier i on spatial channel n via relay m for different transmission strategies can be obtained as summarized in Table II. On the other hand, $P^{[t,i]*}_{SR_m,k,n}$ can be calculated from $P^{[t,i]*}_{SR_m,k,n} = P^{[t,i]*}_{m,k,n} - P^{[t,i]*}_{RD_m,k,n}, \forall t, n$. To obtain the optimal subcarrier allocation, we calculate the marginal benefit $Q^{[t,i]}_{m,k}$ [38] of each transmission strategy by solving $\frac{\partial \mathcal{L}_m}{\partial s^{[t,i]}_{m,k}}\Big|_{P^{[t,i]}_{m,k,n}=P^{[t,i]*}_{m,k,n}} = Q^{[t,i]}_{m,k} - \frac{\beta_i}{w_k+\delta_k}$, which yields

$$\begin{aligned}
Q^{[1,i]}_{m,k} &= C^{[1,i]*}_{m,k} - \sum_{n=1}^{N} \frac{P^{[1,i]*}_{m,k,n}\gamma^{[i]}_{RD_m,k,n}\gamma^{[i]}_{SR_m,n}/(1+\Gamma^{[1,i]*}_{m,k,n})}{\ln(2)(2\Gamma^{[1,i]*}_{m,k,n}\gamma^{[i]}_{LI_m,n} + \Upsilon^{[i]}_{m,k,n})}, \\
Q^{[t,i]}_{m,k} &= C^{[t,i]*}_{m,k} - \sum_{n=1}^{N} \frac{\Gamma^{[t,i]*}_{m,k,n}}{2\ln(2)(1+\Gamma^{[t,i]*}_{m,k,n})}, \ t = 2,3,4, \qquad (29)
\end{aligned}$$

where $C^{[t,i]*}_{m,k}$ and $\Gamma^{[t,i]*}_{m,k,n}$ are obtained by substituting the optimal powers into the corresponding capacity equations and SNR/SINR expressions in Table I, respectively. Thus, the subcarrier selection determined by relay m is given by

$$s^{[t,i]*}_{m,k} = \begin{cases} 1 & \text{if } Q^{[t,i]}_{m,k} = \max\limits_{a,b} Q^{[a,i]}_{m,b} \text{ and } Q^{[a,i]}_{m,b} \geq \frac{\beta_i}{w_b+\delta_b} \\ 0 & \text{otherwise.} \end{cases} \qquad (30)$$

The dual variables β_i act as the global price in using subcarrier i in the system. On the other hand, dual variables δ_k and w_k not only affect the power allocation by changing the water-level of user k in the power allocation equations in Table II, but also force the scheduler to assign more subcarriers to *delay sensitive* users and higher priority users by lowering the price in the selection process. It can be seen from Table II and (30) that relay m, $m \in \{1,\dots,M\}$, only requires the CSI of its own BS-to-relay link, the CSI of its own relay-to-user links, and the dual variables λ, β_i, $i \in \{1,\dots,n_F\}$, and δ_k, $k \in \{1,\dots,K\}$, supplied by the BS.

F. Solution of the Master Dual Problem at the BS

For solving the master problem at the BS, each relay calculates the local resource usage and passes this information, i.e., $C_{m,k}^{[t,i]*}, P_{SR_m,k,n}^{[t,i]*}$, and $P_{RD_m,k,n}^{[t,i]*}$, to the BS. The gradient method is used to solve the minimization in the master problem in (27)

$$\begin{aligned}\beta_i(j+1) &= \left[\beta_i(j) - \xi_1(j) \times \left(1 - \sum_{m=1}^{M}\sum_{k\in\mathcal{U}_m}\sum_{t=1}^{4} s_{m,k}^{[t,i]}\right)\right]^+, \forall i\\ \lambda(j+1) &= \Bigg[\lambda(j) - \xi_2(j) \times \Bigg(P_T \\ &\quad - \sum_{m=1}^{M}\sum_{k\in\mathcal{U}_m}\sum_{i=1}^{n_F}\sum_{t=1}^{4}\sum_{n=1}^{N}\widetilde{P}_{m,k,n}^{[t,i]}\Bigg)\Bigg]^+,\\ \delta_k(j+1) &= \left[\delta_k(j) - \xi_3(j) \times \left(\rho_m^{(k)} - R^{(k)}\right)\right]^+, \ \forall k \in \mathcal{D}_m,\end{aligned} \tag{31}$$

where j is the iteration index, $\xi_1(j)$, $\xi_2(j)$, and $\xi_3(j)$ are positive step sizes. Convergence to the optimal solution is guaranteed if the chosen step sizes satisfy the general conditions stated in [39, Chapter 1.2]. By combining the gradient update equations at the BS and the subcarrier selection criterion in (30) at the relays, a selected subcarrier will be occupied by one user with one transmission strategy only eventually.

G. Algorithm Complexity and Optimization Problem Feasibility

Algorithm Complexity: According to Theorem 1, the optimum transmitter precoding and receiver post-processing matrices diagonalize the MIMO-OFDMA relay channels into N parallel channels. Then, by exploiting the properties of the parallel channels, we transform the original problem into a concave optimization problem with respect to the optimization variables. As a results, dual decomposition can be used to solve the transformed problem. Thus, the computational complexity of the proposed algorithm can be expressed as [40], [41]: $\mathcal{O}(N^3 \times n_F \times K) + \mathcal{O}(N \times K \times n_F \times t_{tot} \times l_{iter})$. Variables t_{tot} and l_{iter} are the total number of transmission strategies and the number of iterations, respectively. The terms $\mathcal{O}(N^3 \times n_F \times K)$ and $\mathcal{O}(N \times K \times n_F \times t_{tot} \times l_{iter})$ represent the complexity of the SVD operation and the complexity of solving the dual problem in (27), respectively. Note that even with the optimal diagonal structure, the complexity of the original optimization problem in (21) is $\mathcal{O}(N^3 \times n_F \times K) + \mathcal{O}(N \times t_{tot}^{K^{n_F}})$, which is prohibitively high compared to the complexity of the transformed problem.

Optimization Problem Feasibility: The considered problem becomes infeasible if the resource allocator is unable to meet the data rate requirements of at least one of the delay sensitive users. In other words, the feasibility of optimization problem (21) is directly related to the question whether the total system power P_T satisfies the condition $P_T \geq P_{margin}$, where P_{margin} is the minimum total power needed to guarantee the required rates R_k of all delay sensitive users, in the absence of non-delay sensitive users [42], [43]. On the other hand, finding P_{margin} is equivalent to the well known margin adaptation problem [35], [43]. Different algorithms have been proposed for finding P_{margin} for multi-carrier systems [35], [43] and can be used for verifying the feasibility of our problem.

IV. Simulation Results

In this section, we evaluate the system performance using simulations. Each cell is modeled as two concentric ring-shaped discs as shown in Fig. 1. The outer boundary and the inner boundary have radii of 1 km and 500 m, respectively. There are $M = 3$ relays equally distributed on the inner cell boundary for assisting the transmission and K active users are uniformly distributed between the inner and the outer boundaries. Unless specified otherwise, there are 3 *delay sensitive* users with data rate requirement $R^{(k)} = 0.1$ bit/s/Hz in the system, while the remaining users are *non-delay sensitive*. The number of subcarriers is $n_F = 128$ with carrier center frequency 2.5 GHz, system bandwidth $\mathcal{B} = 5$ MHz, and $w_k = 1, \forall k$. Each subcarrier has a bandwidth of 39 kHz and a noise variance $N_0 = -128$ dBm. The 3GPP path loss model is used. The small scale fading coefficients of the relay-to-user links are modeled as independent and identically distributed (i.i.d.) Rayleigh random variables. On the other hand, a strong line of sight communication channel between the BS and the relays is expected since they are placed in relatively high positions in practice and the number of blockages between them are limited. Hence, the small scale fading coefficients of the BS-to-relay links are modelled as i.i.d. Rician random variables with Rician factor $\kappa = 6$ dB. The weighted average system throughput is obtained by counting the number of packets which are successfully decoded by the users averaged over both macroscopic and microscopic fading.

A. Convergence of Distributed Algorithm

Figs. 3 and 4 illustrate the evolution of the Lagrange multipliers λ and δ_1 of the distributed algorithm over time for different maximum transmit powers P_T, $K = 15$ users, $M = 3$ relays, $N = 4$ antennas, and an average loop interference power $\mathcal{E}\{\gamma_{LI_m,n}^{[i]}\} = 29$ dB. Positive constant step sizes $\xi_1(t), \xi_2(t)$, and $\xi_3(t)$, which were optimized for fast convergence, were adopted. The results in Figs. 3 and 4 were averaged over 1000 independent adaptation processes. For a practical transmit power region, it can be observed that the distributed iterative algorithm converges fast and typically achieves above 95% of the optimal value within 5 iterations. Note that for finding the optimal dual variables in Figs. 3 and 4, we assume that there is a central unit in which global CSI is available. Since the dual problem is convex in nature, numerical solvers can be used to find the optimal dual variables. Two commonly used methods are gradient based algorithms (with sufficiently large number of iterations) and the ellipsoid method. Besides, the duality gap between the solution of the primal and the dual of the considered problem

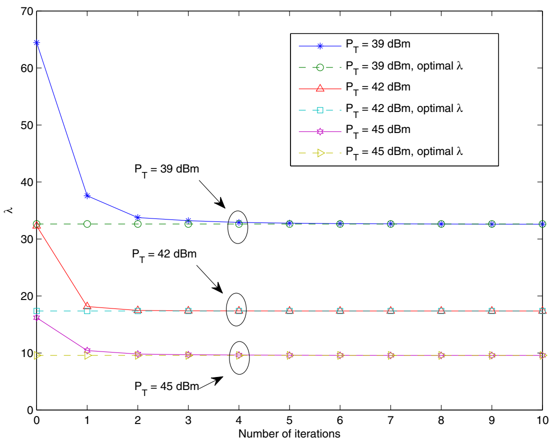


Fig. 3. Lagrange multiplier λ versus number of iterations for $N = 4$ antennas, $K = 15$ users, and $M = 3$ relays.

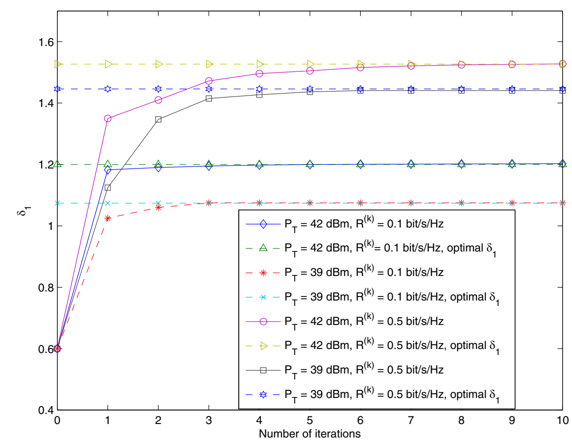


Fig. 4. Lagrange multiplier δ_1 versus number of iterations for $N = 4$ antennas, $K = 15$ users, and $M = 3$ relays for different transmit power levels and data rate requirements for *delay sensitive users*.

is negligible since the transformed problem is convex, cf. Fig. 5.

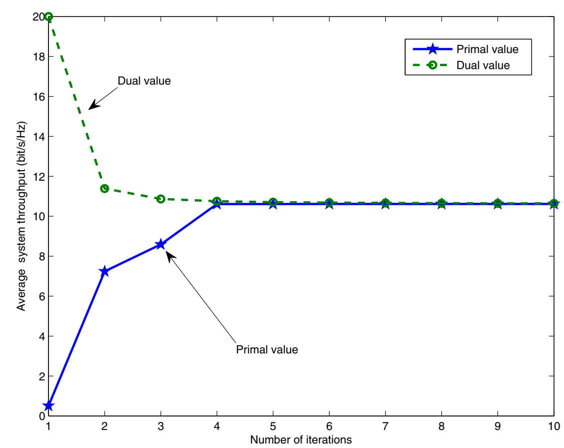


Fig. 5. Average system throughput for primal problem and dual problem versus number of iterations for a maximum transmit power $P_T = 48$ dBm, $K = 15$ users, $M = 3$ relays, $N = 4$ antennas, $w_k = 1\,\forall k$, and an average loop interference power $\mathcal{E}\{\gamma_{LI_m,n}^{[i]}\} = 29$ dB.

B. Transmission Strategy Selection Probability

Figs. 6 and 7 illustrate the transmission strategy selection probability versus P_T for different loop interference power levels. The number of iterations for the proposed distributed algorithm is 5. For a weak to moderate average loop interference power, i.e., $\mathcal{E}\{\gamma_{LI_m,n}^{[i]}\} \leq 26$ dB, as demonstrated in Fig. 6, DF-FD relaying, i.e., $t = 1$, dominates the system performance when there are $N = 4$ antennas. This is because DF-FD relaying has a better spectral efficiency by allowing the BS and the relays to transmit simultaneously in two phases, while the HD relays use two phases to transmit one message. However, when $N = 1$, DF-HD relaying becomes more attractive since DF-FD relaying with a single antenna is more likely to suffer from both strong interference and weak desired channel gain, compared to the case of multiple antennas. On the other hand, when the average loop interference is strong, $\mathcal{E}\{\gamma_{LI_m,n}^{[i]}\} = 29$ dB, as depicted in Fig. 7, DF-HD relaying becomes the dominant strategy and the selection probability of DF-FD relaying drops dramatically, since interference is avoided in DF-HD relaying by sacrificing spectral efficiency. Besides, it can be observed that DF relaying is more attractive than AF in the considered practical operating region as the DF protocol avoids noise and/or self-interference amplification by signal regeneration.

C. Average System Throughput Versus Transmit Power

Fig. 8 depicts the average weighted system throughput versus the total transmit power for a total of 15 users with strong average residual loop interference. The number of iterations for the proposed iterative resource allocation algorithm is 5 and 10. It can be seen that the performance difference between 5 iterations and 10 iterations is negligible which confirms the practicality of our proposed iterative resource allocation algorithm. On the other hand, the system performance scales with the number of antennas almost linearly. This is because the multiple antennas in the transceivers provide extra degrees of freedom for resource allocation. Nevertheless, the performance gain due to multiple antennas is limited in the low transmit power regime, where the loop interference dominates the system performance.

Fig. 9 shows the individual performances of DF-FD, DF-HD, AF-FD, and AF-HD relaying. The resource allocation algorithms for these relaying protocols can be obtained by restricting the hybrid relaying resource allocation algorithm to select the corresponding objective function only. The number of iterations is 5. It can be observed that DF-FD and DF-HD relaying outperform the AF relaying protocols for moderate to strong loop interference powers. This is because the AF relays not only amplify the thermal noise power in the case of HD relaying, but also the loop interference in case of FD relaying. On the other hand, AF-FD and DF-FD relaying outperform HD relaying for weak to moderate average loop interference powers (e.g. $\mathcal{E}\{\gamma_{LI_m,n}^{[i]}\} \leq 26$ dB), which suggests that a

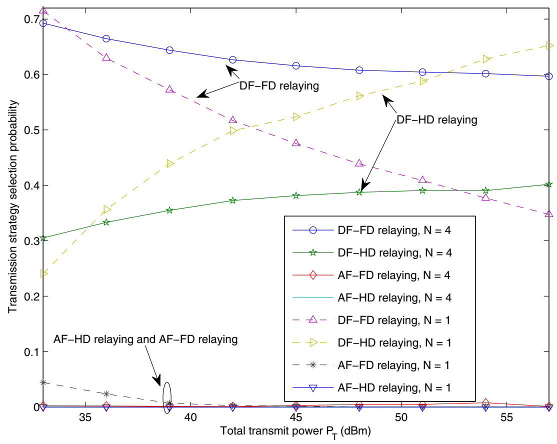


Fig. 6. Transmission strategy selection probability versus total transmit power for $K = 15$ users and different numbers of antennas N for a moderate average loop interference power at each relay, i.e., $\mathcal{E}\{\gamma_{LI_m,n}^{[i]}\} = 26$ dB.

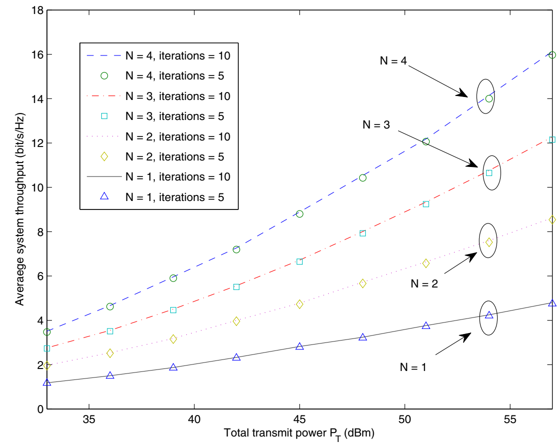


Fig. 8. Average system throughput versus total transmit power for a strong average residual loop residual loop interference power of $\mathcal{E}\{\gamma_{LI_m,n}^{[i]}\} = 29$ dB, different numbers of antennas N, and $K = 15$ users.

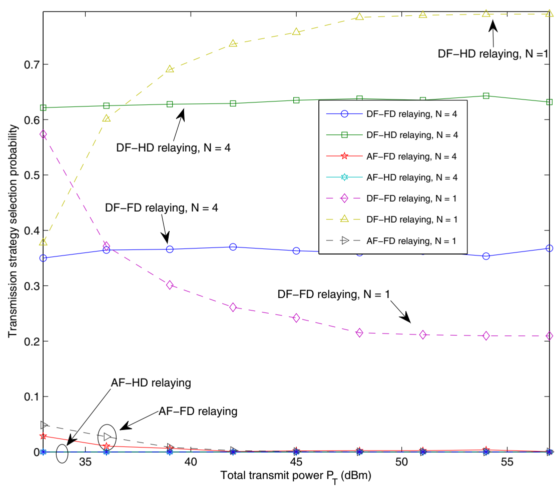


Fig. 7. Transmission strategy selection probability versus total transmit power for $K = 15$ users and different numbers of antennas N for a strong average loop interference power at each relay, i.e., $\mathcal{E}\{\gamma_{LI_m,n}^{[i]}\} = 29$ dB.

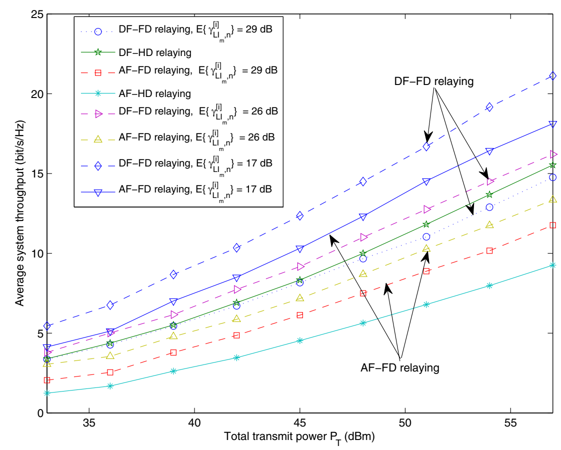


Fig. 9. Average system throughput of different relaying protocols versus total transmit power for different residual loop interference powers $\mathcal{E}\{\gamma_{LI_m,n}^{[i]}\}$, $K = 15$ users, and $N = 4$ antennas.

large performance gain can be achieved by DF-FD relaying if the average loop interference can be sufficiently reduced.

D. Average System Throughput Versus Number of Users

Fig. 10 shows the average system throughput versus the number of users for different numbers of antennas N and different data rate requirements for the delay sensitive users. The number of iterations for the proposed algorithm is 5. It can be observed that the average system throughput increases with the number of users. This is because as the number of users increase, the proposed distributed algorithm has a higher chance to select users who have strong channels in both hops. This effect is known as multi-user diversity (MUD). However, the scheduler looses degrees of freedom for scheduling and resource allocation as the data rate requirements of the *delay sensitive* users become more stringent, since most of the resources are consumed by these users regardless of their possibly poor channel qualities.

V. Conclusions

In this paper, we formulated the dynamic resource allocation and scheduling design for MIMO-OFDMA system with full duplex and hybrid relaying as a non-convex and combinatorial optimization problem, in which heterogeneous users and the loop-interference cancellation error were taken into consideration. By exploiting the structures of the optimal precoding and post-processing matrices, the matrix optimization problem was decomposed into scalar optimization problems. An efficient iterative distributed resource allocation algorithm with closed-form power and subcarrier allocation policies was derived by dual decomposition. Simulation results not only showed that the performance of the proposed iterative algorithm approaches the optimal performance within a small number of iterations, but also demonstrated the possible performance gains obtained by FD MIMO-relaying compared to HD relaying.

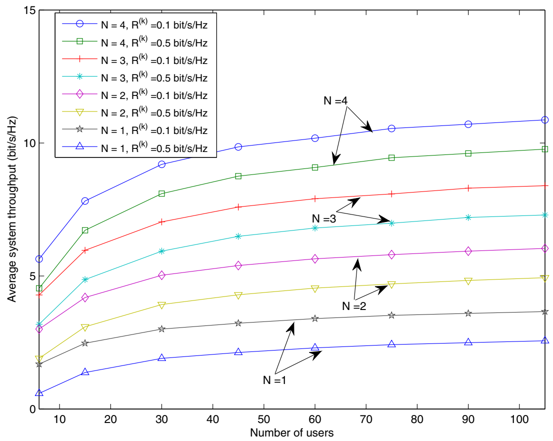


Fig. 10. Average system throughput versus the number of users for an average residual loop interference power of $\mathcal{E}\{\gamma^{[i]}_{LI_{m},n}\} = 29$ dB, different numbers of antennas N, and different data rate requirements for delay sensitive users.

APPENDIX

A. Proof of Theorem 1

Without loss of generality, we consider vectors $\mathbf{x} \in \mathbb{R}^N$ and $\mathbf{y} \in \mathbb{R}^N$ with elements $x_{(1)} \geq x_{(2)} \geq \ldots \geq x_{(N)}$ and $y_{(1)} \geq y_{(2)} \geq \ldots \geq y_{(N)}$ sorted in descending order, respectively. We first state some known results from majorization theory.

Definition 1: [44, Chapter 1.A.1] For any $\mathbf{x}, \mathbf{y} \in \mathbb{R}^N$, $\mathbf{y}$ majorizes $\mathbf{x}$ if

$$\sum_{i=1}^{k} x_{(i)} \leq \sum_{i=1}^{k} y_{(i)}, \quad 1 \leq k \leq N-1 \quad \text{and} \tag{32}$$

$$\sum_{i=1}^{N} x_{(i)} = \sum_{i=1}^{N} y_{(i)}, \tag{33}$$

which can be denoted by $\mathbf{x} \prec \mathbf{y}$.

Definition 2: [44, Chapter 1.A.2] For any $\mathbf{x}, \mathbf{y} \in \mathbb{R}^N$, $\mathbf{y}$ weakly majorizes $\mathbf{x}$ if

$$\sum_{i=1}^{k} x_{(i)} \leq \sum_{i=1}^{k} y_{(i)}, \quad 1 \leq k \leq N, \tag{34}$$

which can be represented by $\mathbf{x} \prec_w \mathbf{y}$.

Theorem 2: [44, Chapter 9.B.1] For an $N \times N$ Hermitian matrix $\mathbf{A}$,

$$\text{diag}(\mathbf{A}) \prec \text{eig}(\mathbf{A}). \tag{35}$$

Lemma 1: [44, Chapter 9.H.2] Let $\mathbf{Z}$ and $\mathbf{Y}$ be two $N \times N$ matrices and $\mathbf{M} = \mathbf{Z}^H \mathbf{Y} \mathbf{Z}$. Then the following is true:

$$\sigma_{\mathbf{M}} \prec_w (\sigma_{\mathbf{Z}} \odot \sigma_{\mathbf{Y}} \odot \sigma_{\mathbf{Z}}), \tag{36}$$

where $\sigma_{\mathbf{M}}$, $\sigma_{\mathbf{Z}}$, and $\sigma_{\mathbf{Y}}$ denote $N \times 1$ vectors containing the singular values of matrices $\mathbf{M}$, $\mathbf{Z}$, and $\mathbf{Y}$, arranged in descending order, respectively. $\odot$ denotes the element-wise Schur product operation of two vectors.

Lemma 2: [44, Chapter 9.H.1.h] Let $\mathbf{Z}$ and $\mathbf{Y}$ be two $N \times N$ matrices, then

$$\text{tr}(\mathbf{ZY}) \geq \sum_{i=1}^{N} \mu_i(\mathbf{Z}) \mu_{N-i+1}(\mathbf{Y}). \tag{37}$$

Furthermore, we define the following matrices for the sake of notational simplicity:

$$\mathbf{H}^{[i]}_{SR_m} \mathbf{B}^{[t,i]}_{m,k} = \mathbf{U}_A \mathbf{\Lambda}_A^{\frac{1}{2}} \mathbf{Q}_A, \tag{38}$$

$$\mathbf{H}^{[i]}_{R_m,k} \mathbf{F}^{[t,i]}_{m,k} = \mathbf{X}^{[t,i]}_{m,k} \left(\mathbf{A}^{[t,i]}_{m,k} + \mathbf{K}^{[t,i]}_{m,k}\right)^{-(1/2)}, \tag{39}$$

$$\mathbf{A}^{[t,i]}_{m,k} = \mathbf{H}^{[i]}_{SR_m} \mathbf{B}^{[t,i]}_{m,k} (\mathbf{B}^{[t,i]}_{m,k})^H (\mathbf{H}^{[i]}_{SR_m})^H = \mathbf{U}_A \mathbf{\Lambda}_A \mathbf{U}_A^H, \tag{40}$$

$$\mathbf{X}^{[t,i]}_{m,k} = \mathbf{H}^{[i]}_{R_m,k} \mathbf{F}^{[t,i]}_{m,k} \left(\mathbf{A}^{[t,i]}_{m,k} + \mathbf{K}^{[t,i]}_{m,k}\right)^{(1/2)} = \mathbf{U}_X \mathbf{\Lambda}_X \mathbf{V}_X^H, \tag{41}$$

where $\mathbf{U}_A \mathbf{\Lambda}_A^{\frac{1}{2}} \mathbf{Q}_A$, $\mathbf{U}_A \mathbf{\Lambda}_A \mathbf{U}_A^H$, and $\mathbf{U}_X \mathbf{\Lambda}_X \mathbf{V}_X^H$ are the SVDs of matrices $\mathbf{H}^{[i]}_{SR_m} \mathbf{B}^{[t,i]}_{m,k} \in \mathbb{C}^{N \times N}$, $\mathbf{A}^{[t,i]}_{m,k} \in \mathbb{C}^{N \times N}$, and $\mathbf{X}^{[t,i]}_{m,k} \in \mathbb{C}^{N \times N}$, respectively. $\mathbf{U}_A$, $\mathbf{U}_X$, and $\mathbf{V}_X^H$ are $\mathbb{C}^{N \times N}$ unitary matrices. $\mathbf{Q}_A \in \mathbb{C}^{N \times N}$ is an arbitrary unitary matrix which can be varied by rotating precoding matrix $\mathbf{B}^{[t,i]}_{m,k}$. In the first part of the proof, we follow a similar approach as in [45]. We prove that the optimal precoding and post-processing matrices used at the BS and relays, respectively, should *jointly diagonalize* the MIMO channels on each subcarrier for minimizing the end-to-end MMSE. Then, based on the optimal structure of the end-to-end MMSE matrix, we construct the optimal precoding and post-processing matrices used at the BS and relays.

The following proof exploits with matrix inversion lemma, i.e., $(\mathbf{A} + \mathbf{BCD})^{-1} = \mathbf{A}^{-1} - \mathbf{A}^{-1} \mathbf{B}(\mathbf{D}\mathbf{A}^{-1}\mathbf{B} + \mathbf{C}^{-1})^{-1}\mathbf{D}\mathbf{A}^{-1}$ [46].

1) The Optimal Precoding and Post-Processing Matrices for AF-FD Relaying: The MMSE matrix $\mathbf{E}^{*[3,i]}_{m,k}$ for $t = 3$ can be written as $\mathbf{E}^{*[3,i]}_{m,k}$

$$\begin{aligned}
&= \left(\mathbf{I}_N + (\mathbf{\Gamma}^{[3,i]}_{m,k})^H (\mathbf{\Theta}^{[i]}_{m,k})^{-1} \mathbf{\Gamma}^{[3,i]}_{m,k}\right)^{-1} \\
&= \mathbf{I}_N - (\mathbf{\Gamma}^{[3,i]}_{m,k})^H \left(\mathbf{\Gamma}^{[3,i]}_{m,k} (\mathbf{\Gamma}^{[3,i]}_{m,k})^H + \mathbf{\Theta}^{[i]}_{m,k}\right)^{-1} \mathbf{\Gamma}^{[3,i]}_{m,k} \\
&= \mathbf{I}_N - (\mathbf{\Gamma}^{[3,i]}_{m,k})^H \Big(\mathbf{H}^{[i]}_{R_m,k} \mathbf{F}^{[3,i]}_{m,k} \left(\mathbf{A}^{[3,i]}_{m,k} + \mathbf{K}^{[3,i]}_{m,k}\right) (\mathbf{F}^{[3,i]}_{m,k})^H \\
&\quad \times (\mathbf{H}^{[i]}_{R_m,k})^H + \mathbf{I}_N\Big)^{-1} \mathbf{\Gamma}^{[3,i]}_{m,k} \\
&= \mathbf{I}_N - (\mathbf{H}^{[i]}_{SR_m} \mathbf{B}^{[3,i]}_{m,k})^H (\mathbf{A}^{[3,i]}_{m,k} + \mathbf{K}^{[3,i]}_{m,k})^{-\frac{1}{2}} (\mathbf{X}^{[3,i]}_{m,k})^H \\
&\quad \times \left(\mathbf{X}^{[3,i]}_{m,k} (\mathbf{X}^{[3,i]}_{m,k})^H + \mathbf{I}_N\right)^{-1} \\
&\quad \times \mathbf{X}^{[3,i]}_{m,k} (\mathbf{A}^{[3,i]}_{m,k} + \mathbf{K}^{[3,i]}_{m,k})^{-\frac{1}{2}} (\mathbf{H}^{[i]}_{SR_m} \mathbf{B}^{[3,i]}_{m,k}) \\
&= \mathbf{I}_N - (\mathbf{H}^{[i]}_{SR_m} \mathbf{B}^{[3,i]}_{m,k})^H (\mathbf{A}^{[3,i]}_{m,k} + \mathbf{K}^{[3,i]}_{m,k})^{-\frac{1}{2}} \\
&\quad \times \mathbf{V}_X \mathbf{\Lambda}_X \mathbf{U}_X^H \left(\mathbf{X}^{[3,i]}_{m,k} (\mathbf{X}^{[3,i]}_{m,k})^H + \mathbf{I}_N\right)^{-1} \\
&\quad \times \mathbf{U}_X \mathbf{\Lambda}_X \mathbf{V}_X^H (\mathbf{A}^{[3,i]}_{m,k} + \mathbf{K}^{[3,i]}_{m,k})^{-\frac{1}{2}} (\mathbf{H}^{[i]}_{SR_m} \mathbf{B}^{[3,i]}_{m,k}) \\
&= \mathbf{I}_N - (\mathbf{H}^{[i]}_{SR_m} \mathbf{B}^{[3,i]}_{m,k})^H (\mathbf{A}^{[3,i]}_{m,k} + \mathbf{K}^{[3,i]}_{m,k})^{-\frac{1}{2}} \mathbf{V}_X \\
&\quad \times ((\mathbf{\Lambda}_X)^{-2} + \mathbf{I}_N)^{-1} \mathbf{V}_X^H (\mathbf{A}^{[3,i]}_{m,k} + \mathbf{K}^{[3,i]}_{m,k})^{-\frac{1}{2}} (\mathbf{H}^{[i]}_{SR_m} \mathbf{B}^{[3,i]}_{m,k}) \\
&= \mathbf{I}_N - \mathbf{Q}_A^H \mathbf{\Lambda}_A^{\frac{1}{2}} (\mathbf{\Lambda}_D) \mathbf{U}_A^H \mathbf{V}_X \left((\mathbf{\Lambda}_X)^{-2} + \mathbf{I}_N\right)^{-1} \\
&\quad \times \mathbf{V}_X^H \mathbf{U}_A (\mathbf{\Lambda}_D) \mathbf{\Lambda}_A^{\frac{1}{2}} \mathbf{Q}_A \\
&= \mathbf{I}_N - \mathbf{Q}_A^H \mathbf{\Lambda}_A^{\frac{1}{2}} (\mathbf{\Lambda}_D) \mathbf{Q}_F^H \left((\mathbf{\Lambda}_X)^{-2} + \mathbf{I}_N\right)^{-1} \mathbf{Q}_F (\mathbf{\Lambda}_D) \mathbf{\Lambda}_A^{\frac{1}{2}} \mathbf{Q}_A \\
&= \mathbf{I}_N - \mathbf{P}_{AF},
\end{aligned} \tag{42}$$

where $\mathbf{Q}_F^H = \mathbf{U}_A^H \mathbf{V}_X$, $\mathbf{\Lambda}_D = (\mathbf{\Lambda}_{A_{m,k}} + \mathbf{K}^{[3,i]}_{m,k})^{-\frac{1}{2}}$, and $\mathbf{P}_{AF} = \mathbf{Q}_A^H \mathbf{\Lambda}_A^{\frac{1}{2}} (\mathbf{\Lambda}_D) \mathbf{Q}_F^H \left((\mathbf{\Lambda}_X)^{-2} + \mathbf{I}_N\right)^{-1} \mathbf{Q}_F (\mathbf{\Lambda}_D) \mathbf{\Lambda}_A^{\frac{1}{2}} \mathbf{Q}_A$

is an $N \times N$ matrix. Now, we apply Theorem 2 to matrix $\mathbf{P}_{AF}$ which yields

$$\text{diag}(\mathbf{P}_{AF}) \prec \text{eig}(\mathbf{P}_{AF}). \quad (43)$$

If we set $\mathbf{Q}_A^H = \mathbf{I}_N$ and $\mathbf{Q}_F^H = \mathbf{I}_N$ such that $\mathbf{P}_{AF}$ is a diagonal matrix, then by applying Lemma 1, (43) can be further expressed as

$$\begin{aligned}\text{diag}(\mathbf{P}_{AF}) \prec \text{eig}(\mathbf{P}_{AF}) &= \sigma_{\mathbf{P}_{AF}} \prec_w (\sigma_{\mathbf{D}_B} \odot \sigma_{\mathbf{D}_A} \odot \sigma_{\mathbf{D}_B}) \\ &= \text{diag}(\mathbf{D}_B^2 \mathbf{D}_A), \end{aligned} \quad (44)$$

where $\mathbf{D}_A = \Big((\mathbf{\Lambda}_X)^{-2} + \mathbf{I}_N\Big)^{-1}$ and $\mathbf{D}_B = \mathbf{\Lambda}_A^{\frac{1}{2}} \mathbf{\Lambda}_D$. In other words, the sum of diagonal elements of matrix $\mathbf{P}_{AF}$ is maximized if $\mathbf{P}_{AF}$ is a diagonal matrix. Therefore, by setting $\mathbf{Q}_A^H = \mathbf{I}_N$ and $\mathbf{Q}_F^H = \mathbf{I}_N$, the sum of diagonal elements of the MMSE matrix is minimized.

On the other hand, using the SVD matrix of $\mathbf{H}_{SR_m}^{[i]}$ (23) in (38), the transmit power at the BS for user k in subcarrier i using transmission strategy t is given by

$$\begin{aligned} &\text{tr}(\mathbf{B}_{m,k}^{[3,i]}(\mathbf{B}_{m,k}^{[3,i]})^H) \\ = \quad &\text{tr}\Big((\mathbf{\Lambda}_{SR_m}^{[i]})^{-2}(\mathbf{U}_{SR_m}^{[i]})^H \mathbf{U}_A \mathbf{\Lambda}_A \mathbf{U}_A^H \mathbf{U}_{SR_m}^{[i]}\Big) \\ \times \quad &\text{tr}\Big((\mathbf{\Lambda}_{SR_m}^{[i]})^{-2} \mathbf{\Lambda}_A\Big), \end{aligned} \quad (45)$$

where (45) is obtained from Lemma 2 which suggests that the transmit power is minimized if $\mathbf{U}_A = \mathbf{U}_{SR_m}^{[i]}$. So, the optimal matrix $\mathbf{B}_{m,k}^{[3,i]}$ can be expressed as

$$\mathbf{B}_{m,k}^{[3,i]} = \mathbf{V}_{SR_m}^{[i]} (\mathbf{\Lambda}_{SR_m}^{[i]})^{-1} \mathbf{\Lambda}_A^{\frac{1}{2}} \mathbf{Q}_A = \mathbf{V}_{SR_m}^{[i]} \mathbf{\Lambda}_{B_m,k}^{[3,i]}, (46)$$

where $\mathbf{Q}_A = \mathbf{I}_N$ and $\mathbf{\Lambda}_{B_m,k}^{[3,i]} = (\mathbf{\Lambda}_{SR_m}^{[3,i]})^{-1} \mathbf{\Lambda}_A^{\frac{1}{2}}$. Similarly, for calculating the optimal precoding matrix used at the relays, we substitute (23) into (38), which yields

$$\begin{aligned} &\tilde{\mathbf{F}}_{m,k}^{[3,i]} \quad (47) \\ = \quad &(\mathbf{V}_{R_m,k}^{[i]})^H \mathbf{F}_{m,k}^{[3,i]}] \\ = \quad &(\mathbf{\Lambda}_{R_m,k}^{[i]})^{-1} (\mathbf{U}_{R_m,k}^{[i]})^H \mathbf{U}_X \mathbf{\Lambda}_X \mathbf{V}_X^H (\mathbf{A}_{m,k}^{[3,i]} + \mathbf{K}_{m,k}^{[3,i]})^{-1/2}. \end{aligned}$$

The power consumed at relay m on subcarrier i can be written as

$$\begin{aligned} &\text{tr}\Big(\tilde{\mathbf{F}}_{m,k}^{[3,i]} (\mathbf{A}_{m,k}^{[3,i]} + \mathbf{K}_{m,k}^{[3,i]}) (\tilde{\mathbf{F}}_{m,k}^{[3,i]})^H\Big) \\ = \quad &\text{tr}\Big(\mathbf{F}_{m,k}^{[3,i]} (\mathbf{A}_{m,k}^{[3,i]} + \mathbf{K}_{m,k}^{[3,i]}) (\mathbf{F}_{m,k}^{[3,i]})^H\Big) \\ = \quad &\text{tr}\Big((\mathbf{\Lambda}_{R_m,k}^{[i]})^{-2} (\mathbf{U}_{R_m,k}^{[i]})^H \mathbf{U}_X \mathbf{\Lambda}_X^2 \mathbf{U}_X^H \mathbf{U}_{R_m,k}^{[i]}\Big) \\ = \quad &\text{tr}\Big((\mathbf{\Lambda}_{R_m,k}^{[i]})^{-2} \mathbf{\Lambda}_X^2\Big), \end{aligned} \quad (48)$$

where the last equality in (48) is obtained by applying Lemma 2 which reveals that minimal power transmission is obtained if $\mathbf{U}_X = \mathbf{U}_{R_m,k}^{[i]}$ and $\mathbf{V}_X^H = \mathbf{U}_A = \mathbf{U}_{SR_m}^{[i]}$. So, the optimal matrix $\mathbf{F}_{m,k}^{[3,i]}$ is given by

$$\begin{aligned} \mathbf{F}_{m,k}^{[3,i]} &= \mathbf{V}_{R_m,k}^{[i]} (\mathbf{\Lambda}_{R_m,k}^{[i]})^{-1} \mathbf{\Lambda}_X \mathbf{V}_X^H (\mathbf{A}_{m,k}^{[3,i]} + \mathbf{K}_{m,k}^{[3,i]})^{-1/2} \\ &= \mathbf{V}_{R_m,k}^{[i]} (\mathbf{\Lambda}_{R_m,k}^{[i]})^{-1} \mathbf{\Lambda}_X \mathbf{\Lambda}_D (\mathbf{U}_{SR_m}^{[i]})^H \\ &= \mathbf{V}_{R_m,k}^{[i]} \mathbf{\Lambda}_{F_m,k}^{[3,i]} (\mathbf{U}_{SR_m}^{[i]})^H. \end{aligned} \quad (49)$$

Clearly, $\mathbf{B}_{m,k}^{[t,i]}$ and $\mathbf{F}_{m,k}^{[t,i]}$ as given in (46) and (49) diagonalize the end-to-end equivalent channel.

2) The Optimal Precoding and Post-Processing Matrices for DF-FD Relaying: For DF-relaying protocols, the channel capacity on subcarrier i is always limited by its bottleneck link. The MMSE matrix at the relay in (11) can be written as

$$\begin{aligned} &\mathbf{\Delta}_{m,k}^{*[1,i]} \\ = \quad &\mathbf{I}_N - (\mathbf{H}_{SR_m}^{[i]} \mathbf{B}_{m,k}^{[1,i]})^H \Big(\mathbf{H}_{SR_m}^{[i]} \mathbf{B}_{m,k}^{[1,i]} (\mathbf{H}_{SR_m}^{[i]} \mathbf{B}_{m,k}^{[1,i]})^H + \\ &\mathbf{\Upsilon}_{m,k}^{[1,i]}\Big)^{-1} \mathbf{H}_{SR_m}^{[i]} \mathbf{B}_{m,k}^{[1,i]} \\ = \quad &\mathbf{I}_N - \mathbf{Q}_A^H \mathbf{\Lambda}_A^{\frac{1}{2}} \mathbf{U}_A^H \Big(\mathbf{U}_A \mathbf{\Lambda}_A \mathbf{U}_A^H + \mathbf{\Upsilon}_{m,k}^{[1,i]}\Big)^{-1} \mathbf{U}_A \mathbf{\Lambda}_A^{\frac{1}{2}} \mathbf{Q}_A \\ = \quad &\mathbf{I}_N - \mathbf{Q}_A^H \mathbf{\Lambda}_A^{\frac{1}{2}} \mathbf{\Lambda}_{A+\Upsilon} \mathbf{\Lambda}_A^{\frac{1}{2}} \mathbf{Q}_A = \mathbf{I}_N - \mathbf{Q}_A^H \mathbf{\Lambda}_Z \mathbf{Q}_A \\ = \quad &\mathbf{I}_N - \mathbf{P}_{DF}, \end{aligned} \quad (50)$$

where $\mathbf{P}_{DF} = \mathbf{Q}_A^H \mathbf{\Lambda}_A^{\frac{1}{2}} \mathbf{\Lambda}_{A+\Upsilon} \mathbf{\Lambda}_A^{\frac{1}{2}} \mathbf{Q}_A$, $\mathbf{\Lambda}_{A+\Upsilon} = \mathbf{\Lambda}_A + \mathbf{\Upsilon}_{m,k}^{[1,i]}$, $\mathbf{\Lambda}_Z = \mathbf{\Lambda}_A^{\frac{1}{2}} \mathbf{\Lambda}_{A+\Upsilon} \mathbf{\Lambda}_A^{\frac{1}{2}}$, and $\mathbf{\Upsilon}_{m,k}^{[1,i]}$ is a diagonal matrix which was defined in (12). By applying the same technique as in (44), the diagonal elements of matrix $\mathbf{P}_{DF}$ are majorized if $\mathbf{Q}_A^H = \mathbf{I}_N$ such that $\mathbf{P}_{DF}$ is a diagonal matrix. In other words, the MSE is minimized if $\mathbf{P}_{DF}$ is a diagonal matrix. On the other hand, we can treat the link of relay m to user k as a point-to-point MIMO communication channel due to the intermediate decoding and re-encoding processes at relay m. Therefore, we can directly apply the results in [33] to diagonalize the user channels with the help of $\mathbf{F}_{m,k}^{[t,i]}$.

By following the same steps as in (45), we find that the optimal structure of precoding matrix $\mathbf{B}_{m,k}^{[t,i]}$ for DF-FD relaying has the same form as in (46), i.e., $\mathbf{B}_{m,k}^{[t,i]} = \mathbf{V}_{SR_m}^{[i]} \mathbf{\Lambda}_{B_m,k}^{[t,i]}$. On the other hand, for the link of relay m to user k on subcarrier i, the optimal structure of $\mathbf{F}_{m,k}^{[t,i]}$ should contain the N eigenvectors of $\mathbf{H}_{R_m,k}^{[i]} (\mathbf{H}_{R_m,k}^{[i]})^H$ for channel diagonalization, i.e., $\mathbf{F}_{m,k}^{[t,i]} = \mathbf{V}_{R_m,k}^{[i]} \mathbf{\Lambda}_{F_m,k}^{[t,i]}$, where $\mathbf{\Lambda}_{F_m,k}^{[t,i]} \in \mathbb{C}^{N \times N}$ is a diagonal matrix with $\text{diag}(\mathbf{\Lambda}_{F_m,k}^{[t,i]}) = \Big\{\sqrt{P_{RD_m,k,1}^{[t,i]}} \sqrt{P_{RD_m,k,2}^{[t,i]}} \ldots \sqrt{P_{RD_m,k,N}^{[t,i]}}\Big\}$, which represents the transmit power in each eigenchannel for the relay m to user k link on subcarrier i.

Remark 1: To prove Theorem 1 for AF-HD relaying, we can follow the proof in [45] or use the same approach as above and set the interference cancellation error to $\sigma_e^2 = 0$.

B. Proof of Proposition 1

For $t = 1$, i.e., DF-FD relaying, by varying the power allocation variables in each hop, the maximum capacity for $t = 1$ on subcarrier i on spatial channel n is achieved when the amounts of information received at the relay and the destination are identical, i.e., $\frac{P_{SR_m,k,n}^{[1,i]} \gamma_{SR_m,n}^{[i]}}{\gamma_{LI_m,n}^{[i]} P_{RD_m,k,n}^{[1,i]} + 1} = P_{RD_m,k,n}^{[1,i]} \gamma_{RD_m,k,n}^{[i]}$. Therefore, we can set $P_{SR_m,k,n}^{[1,i]} = P_{m,k,n}^{[1,i]} - P_{RD_m,k,n}^{[1,i]}$ in the above equality and solve for $P_{RD_m,k,n}^{[1,i]}$ which yields

$$P_{RD_m,k,n}^{[1,i]} = \frac{\sqrt{(\Upsilon_{m,k,n}^{[i]})^2 + 4\Psi_{m,k,n}^{[i]} P_{m,k,n}^{[1,i]}} - \Upsilon_{m,k,n}^{[i]}}{2\gamma_{LI_m,n}^{[i]} \gamma_{RD_m,k,n}^{[i]}}. \quad (51)$$

Then, the effective SINR can be calculated by multiplying $P^{[1,i]}_{RD_m,k,n}$ in (51) with $\gamma^{[i]}_{RD_m,k,n}$. $C^{[2,i]}_{m,k}$ can be obtained in a similar manner. On the other hand, for the AF protocol, we assume a high SNR in order to obtain a tractable result. Then, the SINR/SNR for $t = 3, 4$ can be approximated as

$$\Gamma^{[3,i]}_{m,k,n} \approx \frac{\frac{P^{[3,i]}_{SR_m,k,n}\gamma^{[i]}_{SR_m,n}}{P^{[3,i]}_{RD_m,k,n}\gamma^{[i]}_{LI_m,n}} P^{[3,i]}_{RD_m,k,n}\gamma^{[i]}_{RD_m,k,n}}{P^{[3,i]}_{RD_m,k,n}\gamma^{[i]}_{RD_m,k,n} + \frac{P^{[3,i]}_{SR_m,k,n}\gamma^{[i]}_{SR_m,n}}{P^{[3,i]}_{RD_m,k,n}\gamma^{[i]}_{LI_m,n}}} \text{ and}$$

$$\Gamma^{[4,i]}_{m,k,n} \approx \frac{\gamma^{[i]}_{SR_m,n} P^{[4,i]}_{SR_m,k,n} P^{[4,i]}_{RD_m,k,n}\gamma^{[i]}_{RD_m,k,n}}{\gamma^{[i]}_{SR_m,n} P^{[4,i]}_{SR_m,k,n} + P^{[4,i]}_{RD_m,k,n}\gamma^{[i]}_{RD_m,k,n}}, \quad (52)$$

respectively. Now, we use $P^{[t,i]}_{SR_m,k,n} = P^{[t,i]}_{m,k,n} - P^{[t,i]}_{RD_m,k,n}$ in the SINR/SNR expressions and take the derivative of $\Gamma^{[t,i]}_{m,k,n}$ w.r.t. $P^{[t,i]}_{SR_m,k,n}$ for $t = 3, 4$. Again, we can express $P^{[t,i]}_{RD_m,k,n}$ in terms of $P^{[t,i]}_{m,k,n}$ and the channel coefficients, e.g. $P^{[4,i]}_{RD_m,k,n} = \frac{P^{[4,i]}_{m,k,n}(\gamma^{[i]}_{SR_m,n} - \sqrt{\gamma^{[i]}_{SR_m,n}\gamma^{[i]}_{RD_m,k,n}})}{\gamma^{[i]}_{SR_m,n} - \gamma^{[i]}_{RD_m,k,n}}$ for $t = 4$. Then, $C^{[4,i]}_{m,k}$ can be obtained by substituting the new expression for $P^{[t,i]}_{RD_m,k,n}$ into (52) for $t = 4$. For deriving $C^{[3,i]}_{m,k}$ in Table I, we follow the same steps as for $t = 4$, and further approximate the resulting SINR as

$$\Gamma^{[3,i]}_{m,k,n} \approx \frac{(P^{[3,i]}_{m,k,n})^2\gamma^{[i]}_{SR_m,n}\gamma^{[i]}_{RD_m,k,n}}{P^{[3,i]}_{m,k,n}\gamma^{[i]}_{SR_m,n} + 2\sqrt{(P^{[3,i]}_{m,k,n})^3\Psi^{[i]}_{m,k,n}}} \approx \frac{\sqrt{\gamma^{[i]}_{RD_m,k,n}\gamma^{[i]}_{SR_m,n}P^{[3,i]}_{m,k,n}}}{2\sqrt{\gamma^{[i]}_{LI_m,n}}} \quad (53)$$

for $P^{[3,i]}_{m,k,n} \to \infty$ to make the power allocation tractable.

References

[1] H. Zhu and J. Wang, "Chunk-based resource allocation in OFDMA systems—part I: chunk allocation," *IEEE Trans. Commun.*, vol. 57, pp. 2734–2744, Sep. 2009.

[2] V. Lau, W. K. Ng, and D. Hui, "Asymptotic tradeoff between cross-layer goodput gain and outage diversity in OFDMA systems with slow fading and delayed CSIT," *IEEE Trans. Wireless Commun.*, vol. 7, pp. 2732–2739, July 2008.

[3] D. Lopez-Perez, A. Valcarce, G. de la Roche, and J. Zhang, "OFDMA femtocells: a roadmap on interference avoidance," *IEEE Commun. Mag.*, vol. 47, pp. 41–48, Sep. 2009.

[4] M. K. Awad and S. Xuemin, "OFDMA based two-hop cooperative relay network resources allocation," in *Proc. 2008 IEEE Int. Conf. Commun.*, pp. 4414–4418.

[5] D. W. K. Ng and R. Schober, "Cross-layer scheduling for OFDMA amplify-and-forward relay networks," *IEEE Trans. Veh. Technol.*, vol. 59, pp. 1443–1458, Mar. 2010.

[6] I. Hammerstrom and A. Wittneben, "Power allocation schemes for amplify-and-forward MIMO-OFDM relay links," *IEEE Trans. Wireless Commun.*, vol. 6, pp. 2798–2802, Aug. 2007.

[7] E. Calvo, J. Vidal, and J. R. Fonollosa, "Optimal resource allocation in relay-assisted cellular networks with partial CSI," *IEEE Trans. Signal Process.*, vol. 57, pp. 2809–2823, July 2009.

[8] W. Nam, W. Chang, S.-Y. Chung, and Y. H. Lee, "Transmit optimization for relay-based cellular OFDMA systems," in *Proc. 2007 IEEE Int. Commun. Conf.*, pp. 5714–5719.

[9] B. Gedik and M. Uysal, "Impact of imperfect channel estimation on the performance of amplify-and-forward relaying," *IEEE Trans. Wireless Commun.*, vol. 8, pp. 1468–1479, Mar. 2009.

[10] I. Krikidis, J. Thompson, and S. Mclaughlin, "On the diversity order of non-orthogonal amplify-and-forward over block-fading channels," *IEEE Trans. Wireless Commun.*, vol. 8, pp. 1890–1900, June 2010.

[11] D. W. K. Ng and R. Schober, "Cross-layer scheduling design for OFDMA two-way amplify-and-forward relay networks," in *Proc. 2010 IEEE Wireless Commun. Netw. Conf.*, pp. 1–6.

[12] X. J. Zhang and Y. Gong, "Adaptive power allocation in two-way amplify-and-forward relay networks," in *Proc. 2009 IEEE Int. Commun. Conf.*

[13] M. Chen and A. Yener, "Power allocation for multi-access two-way relaying," in *Proc. 2009 IEEE Int. Commun. Conf.*

[14] H. Ju, E. Oh, and D. Hong, "Catching resource-devouring worms in next-generation wireless relay systems: two-way relay and full-duplex relay," *IEEE Commun. Mag.*, vol. 47, pp. 58–65, Sep. 2009.

[15] "Full duplex configuration of Un and Uu subframes for type I relay," 3GPP TSG RAN WG1 R1-100139, Tech. Rep., Jan. 2010.

[16] "Text proposal on inband full duplex relay for TR 36.814," 3GPP TSG RAN WG1 R1-101659, Tech. Rep., Feb. 2010.

[17] T. Riihonen, S. Werner, and R. Wichman, "Hybrid full-duplex/half-duplex relaying with transmit power adaptation," *IEEE Trans. Wireless Commun.*, vol. 10, pp. 3074–3085, Sep. 2011.

[18] H. Ju, E. Oh, and D. Hong, "Improving efficiency of resource usage in two-hop full duplex relay systems based on resource sharing and interference cancellation," *IEEE Trans. Wireless Commun.*, vol. 8, pp. 3933–3938, Aug. 2009.

[19] Y. Fan, H. V. Poor, and J. S. Thompson, "Cooperative multiplexing in full-duplex multi-antenna relay networks," in *Proc. 2008 IEEE Global Telecommun. Conf.*, pp. 1–5.

[20] T. Kwon, S. Lim, S. Choi, and D. Hong, "Optimal duplex mode for DF relay in terms of the outage probability," *IEEE Trans. Veh. Technol.*, vol. 59, pp. 3628–3634, Sep. 2010.

[21] D. W. K. Ng and R. Schober, "Dynamic resource allocation in OFDMA systems with full and hybrid-duplex relaying," in *Proc. 2011 IEEE Int. Commun. Conf.*

[22] T. Riihonen, S. Werner, and R. Wichman, "Optimized gain control for single-frequency relaying with loop interference," *IEEE Trans. Wireless Commun.*, vol. 8, pp. 2801–2806, June 2009.

[23] M. Jain, J. I. Choi, T. Kim, D. Bharadia, K. Srinivasan, S. Seth, P. Levis, S. Katti, and P. Sinha, "Practical, real-time, full duplex wireless," in *Proc. 2011 Int. Conf. Mobile Comput. Netw.*

[24] J. I. Choi, M. Jain, K. Srinivasan, P. Levis, and S. Katti, "Achieving single channel, full duplex wireless communication," in *Proc. 2010 Int. Conf. Mobile Comput. Netw.*

[25] M. A. Khojastepour, K. Sundaresan, S. Rangarajan, X. Zhang, and S. Barghi, "The case for antenna cancellation for scalable full duplex wireless communications," in *Proc. 2011 ACM Workshop Hot Topics Netw.*

[26] M. Duarte and A. Sabharwal, "Full-duplex wireless communications using off-the-shelf radios: Feasibility and first results," in *Proc. 2010 Asilomar Conf. Signals, Syst. Computers*, pp. 1558–1562.

[27] Y. Yu and Y. Hua, "Power Allocation for a MIMO relay system with multiple-antenna users," *IEEE Trans. Signal Process.*, vol. 58, pp. 2823–2835, May 2010.

[28] C. Li, X. Wang, L. Yang, and W.-P. Zhu, "A joint source and relay power allocation scheme for a class of MIMO relay systems," *IEEE Trans. Signal Process.*, vol. 57, pp. 4852–4860, Dec. 2009.

[29] Y. Fu, L. Yang, and W.-P. Zhu, "A nearly optimal amplify-and-forward relaying scheme for two-hop MIMO multi-relay networks," *IEEE Commun. Lett.*, vol. 14, pp. 229–231, Mar. 2010.

[30] J. Andrews, A. Ghosha, and R. Muhamed, *Fundamentals of WiMAX: Understanding Broadband Wireless Networking*, 1st edition. Prentice Hall PTR, 2007.

[31] Y. Liu, M. Tao, B. Li, and H. Shen, "Optimization framework and graph-based approach for relay-assisted bidirectional OFDMA cellular networks," *IEEE Trans. Wireless Commun.*, vol. 9, pp. 3490–3500, Nov. 2010.

[32] C. Shannon, "Communication in the presence of noise," *Proc. IRE*, vol. 37, pp. 10–21, Jan. 1949.

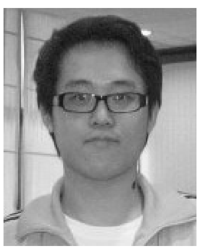
[33] D. P. Palomar, J. M. Cioffi, and M. A. Lagunas, "Joint Tx-Rx beamforming design for multicarrier MIMO channels: a unified framework for convex optimization," *IEEE Trans. Signal Process.*, vol. 51, pp. 2381–2401, Sep. 2003.

[34] N. Kim, Y. Lee, and H. Park, "Performance analysis of MIMO system with linear MMSE receiver," *IEEE Trans. Wireless Commun.*, vol. 7, pp. 4474–4478, Nov. 2008.

[35] C. Y. Wong, R. S. Cheng, K. B. Letaief, and R. D. Murch, "Multiuser OFDM with adaptive subcarrier, bit, and power allocation," *IEEE J. Sel. Areas Commun.*, vol. 17, pp. 1747–1758, Oct. 1999.

[36] W. Yu and R. Lui, "Dual methods for nonconvex spectrum optimization of multicarrier systems," *IEEE Trans. Commun.*, vol. 54, pp. 1310–1321, July 2006.

[37] S. Boyd and L. Vandenberghe, *Convex Optimization*. Cambridge University Press, 2004.
[38] W. Yu and J. M. Cioffi, "FDMA capacity of Gaussian multiple-access channels with ISI," *IEEE Trans. Commun.*, vol. 50, pp. 102–111, Jan. 2002.
[39] D. P. Bertsekas, *Nonlinear Programming*, 2nd edition. Athena Scientific, 1999.
[40] I. Wong and B. Evans, "Optimal downlink OFDMA resource allocation with linear complexity to maximize ergodic rates," *IEEE Trans. Wireless Commun.*, vol. 7, pp. 962–971, Mar. 2008.
[41] Z. Shen, R. Chen, J. Andrews, R. Heath, and B. Evans, "Low complexity user selection algorithms for multiuser MIMO systems with block diagonalization," *IEEE Trans. Signal Process.*, vol. 54, pp. 3658–3663, Sep. 2006.
[42] M. Tao, Y.-C. Liang, and F. Zhang, "Resource allocation for delay differentiated traffic in multiuser OFDM systems," *IEEE Trans. Wireless Commun.*, vol. 7, pp. 2190–2201, June 2008.
[43] N. Hassan and M. Assaad, "Resource allocation in multiuser OFDMA system: feasibility and optimization study," in *Proc. 2009 IEEE Wireless Commun. Netw. Conf.*, pp. 1–6.
[44] A. W. Marshall and I. Olkin, *Inequalities: Theory of Majorization and its Applications*. Academic Press, 1979.
[45] Y. Rong, X. Tang, and Y. Hua, "A unified framework for optimizing linear nonregenerative multicarrier MIMO relay communication systems," *IEEE Trans. Signal Process.*, vol. 57, pp. 4837–4851, Dec. 2009.
[46] R. A. Horn and C. R. Johnson, *Matrix Analysis*. Cambridge University Press, 1985.



Derrick Wing Kwan Ng (S'06) received the bachelor degree with first class honors and master of philosophy (M.Phil.) degree in electronic engineering from the Hong Kong University of Science and Technology (HKUST) in 2006 and 2008, respectively. He is currently working toward the Ph.D. degree in the University of British Columbia (UBC). In the summer of 2011 and spring of 2012, he was a visiting scholar at the Centre Tecnològic de Telecomunicacions de Catalunya - Hong Kong (CTTC-HK). His research interests include cross-layer optimization for wireless communication systems, resource allocation in OFDMA wireless systems, and communication theory. He received the Best Paper Awards at the IEEE Global Telecommunication Conference (Globecom) 2011 and the IEEE Third International Conference on Communications and Networking in China 2008. He was awarded the IEEE Student Travel Grants for attending IEEE WCNC 2010, IEEE ICC 2011, and IEEE Globecom 2011. He was also the recipient of the 2009 Four Year Doctoral Fellowship from the UBC, Sumida & Ichiro Yawata Foundation Scholarship in 2008, and R&D Excellence Scholarship from the Center for Wireless Information Technology in HKUST in 2006. He has served as an editorial assistant of the IEEE TRANSACTIONS ON COMMUNICATIONS since Jan. 2012.

Ernest S. Lo (S'02-M'08) is the Chief Representative of the Centre Tecnològic de Telecomunicacions de Catalunya - Hong Kong (CTTC-HK). Prior to this, he was a Croucher Postdoc Fellow at Stanford University, and received his Ph.D., M.Phil., and B.Eng. (1st Hons.) from Hong Kong University of Science and Technology. His current research interests cover new design opportunities for existing and future wireless multiuser networks for enhancement of flexibility and spectral & energy efficiencies. He has also worked on resource allocation, channel coding, and wireless system-level design.

Dr. Lo has received two Best Paper Awards, one at IEEE ICC'07, Glasgow, and another one at IEEE GlobeCom'11, Houston. He was also the award winner of the Croucher Fellowship in 2008. He contributed to the standardization of the IEEE 802.22 cognitive radio WRAN system and holds several pending and granted US and China patents. Some of them were successfully transferred to companies. He served as an Editorial Assistant of the IEEE TRANSACTIONS ON WIRELESS COMMUNICATIONS when it was founded and has been a TPC member of various conferences, including IEEE PIMRC'09, IEEE ICC'10, IEEE GLOBECOM'10, IEEE ICC'11, IEEE GLOBECOM'11, ICNC'12, and IEEE ICC'12, et al. He was also honored as an Exemplary Reviewer of IEEE COMMUNICATIONS LETTERS.

Robert Schober (M'01-SM'08-F'10) was born in Neuendettelsau, Germany, in 1971. He received the Diplom (Univ.) and the Ph.D. degrees in electrical engineering from the University of Erlangen-Nuermberg in 1997 and 2000, respectively. From May 2001 to April 2002, he was a Postdoctoral Fellow at the University of Toronto, Canada, sponsored by the German Academic Exchange Service (DAAD). Since May 2002, he has been with the University of British Columbia (UBC), Vancouver, Canada, where he is now a Full Professor and Canada Research Chair (Tier II) in Wireless Communications. His research interests fall into the broad areas of communication theory, wireless communications, and statistical signal processing.

Dr. Schober received the 2002 Heinz Maier-Leibnitz Award of the German Science Foundation (DFG), the 2004 Innovations Award of the Vodafone Foundation for Research in Mobile Communications, the 2006 UBC Killam Research Prize, the 2007 Wilhelm Friedrich Bessel Research Award of the Alexander von Humboldt Foundation, and the 2008 Charles McDowell Award for Excellence in Research from UBC. In addition, he received best paper awards from the German Information Technology Society (ITG), the European Association for Signal, Speech and Image Processing (EURASIP), IEEE ICUWB 2006, the International Zurich Seminar on Broadband Communications, and European Wireless 2000. Dr. Schober is also the Area Editor for Modulation and Signal Design for the IEEE TRANSACTIONS ON COMMUNICATIONS.