Fair Scheduling and Resource Allocation for Wireless Cellular Network with Shared Relays

Yicheng Lin and Wei Yu
Department of Electrical and Computer Engineering
University of Toronto, Toronto, Ontario, Canada
Email: {ylin, weiyu}@comm.utoronto.ca

***Abstract*—This paper examines a shared relaying architecture for intercell interference mitigation in wireless cellular networks, where instead of deploying multiple separate relays within each cell sector, a single relay with multiple antennas is placed at the cell edge and is shared by multiple sectors. To maximize the benefit of shared relaying, resource allocation and the scheduling of users among the adjacent sectors need to be optimized jointly. This paper first provides a degree-of-freedom analysis as a motivation for the shared relay architecture, then formulates and solves a network utility maximization problem for a realistic shared relaying network, where zero-forcing beamforming is used at the relay to separate users spatially and multiple users are scheduled in the frequency domain to maximize frequency reuse. System-level simulations show that the incorporation of the shared relays can significantly improve the overall network utility, and increase the throughput of cell edge users in particular as compared to separate relaying.**

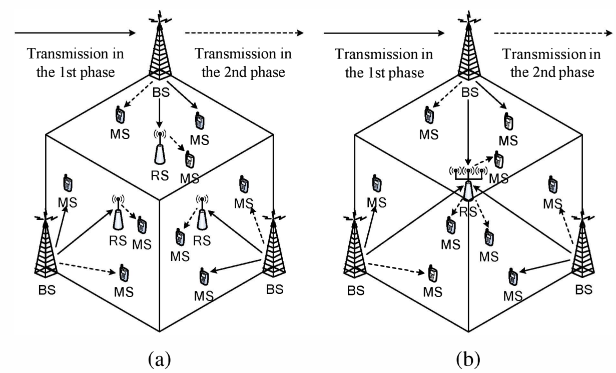

Fig. 1. Relay scenarios. (a) Separate relaying. (b) Shared relaying.

I. Introduction

Conventional two-hop fixed relays provide a convenient solution for extending the coverage of existing cellular infrastructure. The effectiveness of the conventional relaying architecture is, however, also limited by intercell interference. In fact, whenever a relay is deployed close to the cell edge, it has the potential to cause more intercell interference than that in a conventional cellular network. The idea of shared relay, first proposed in [1] and [2], is one way to tackle this problem. In a shared relaying architecture, a multi-antenna relay is placed at the intersection of multiple adjacent sectors and it can be thought of as a coordinated version of multiple separate relays. In the downlink, the shared relay is capable of mitigating intercell interference by spatially separating signals from different base-stations using beamforming techniques, then spatially multiplexing the decoded signals to multiple users. Fig. 1 illustrates a cellular relay network with three sectors serving three users in the downlink, where Fig. 1(a) shows a separate relaying architecture, and Fig. 1(b) shows a shared relaying architecture. Note that as compared to separate relaying, although the shared relay is placed further away from the base-station, its interference mitigation capability more than compensates the increased base-station-to-relay distance, leading to an improved overall network performance.

To truly quantify the benefit of shared relaying, it is important to evaluate the network performance from a system-level perspective. Toward this end, this paper focuses on the scheduling and resource allocation aspects of the shared relaying network. Specifically, we pose network optimization questions such as how the frequencies should be allocated among different links in an OFDM system, and how the frequencies should be reused within each cell. We adopt a network utility maximization framework to characterize the benefit of shared relaying.

There have only been a limited number of works in the literature on the shared relay networks after the initial qualitative description of the concept in IEEE 802.16m [1], [2]. Most notably, [3] shows that shared relaying can approach the gains of local base-station coordination at reduced complexity. In [3], multiple-input multiple-output (MIMO) multiple-access and broadcast techniques are used in the shared relay; the time durations of the two phases are optimally adjusted. The shared relaying concept is further studied in [4], where zero-forcing (ZF) methods are used at the relay, and advanced base-station coordination and two-way relaying techniques are considered. In [5], a joint processing scheme that improves the shared relay strategy of [3] is proposed by letting the base-station and relay send the same message to the corresponding user in the second phase. However, none of these works consider the impact of scheduling, the need for maintaining fairness, and the design of algorithms for dynamic resource allocation and frequency reuse—issues that this paper aims to address.

II. Spatial Degrees of Freedom Analysis

Although it is clear that shared relays are capable of mitigating interference in both the feeder and access links, a quantitative analysis is challenging under complex wireless channels and topologies. This section presents an asymptotic

978-1-4673-3140-1/12/$31.00 ©2012 IEEE

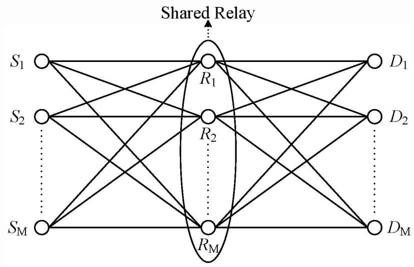

Fig. 2. Abstract model of a shared relaying system

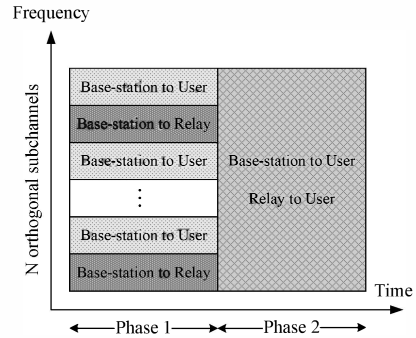

Fig. 3. Half-duplex two-phase relay frame structure

degree-of-freedom[1] analysis to illustrate the theoretical motivation of shared relaying. To this end, we consider an abstract two-hop channel model in Fig. 2, where the shared relay has M antennas to coordinate the transmission of M source-destination pair, while all other nodes have single antenna each. The direct source-to-destination links are ignored since two-hop users typically have weak direct links from the source.

A. Cellular Network Without Relay

The cellular network without relaying can be modeled as an M-user interference channel. Using TDMA, the achievable degree of freedom is $\eta_{s,d} = 1$.

B. Two-Hop Relay Network

Let $N_{s,r}$ and $N_{r,d}$ be the number of frequency resources used in the two hops respectively, and let $C_{s,r}$, $C_{r,d}$ be the corresponding spectral efficiency. To satisfy flow conservation, we have $N_{s,r}C_{s,r} = N_{r,d}C_{r,d}$. For a half-duplex system, the equivalent spectral efficiency of the two-hop link is the ratio of the total data rate and total consumed resources

$$C_{s,r,d} = \frac{N_{s,r}C_{s,r}}{N_{s,r} + N_{r,d}} = \frac{N_{s,r}C_{s,r}}{N_{s,r} + \frac{N_{s,r}C_{s,r}}{C_{r,d}}} = \frac{1}{\frac{1}{C_{s,r}} + \frac{1}{C_{r,d}}}. \tag{1}$$

This is consistent with the result in [6] for multi-hop rate adaptive relaying. Let the degrees of freedom of source-relay and relay-destination links be $\eta_{s,r}$ and $\eta_{r,d}$, respectively. The degree of freedom of the overall two-hop link is therefore

$$\eta_{s,r,d} = \lim_{\mathrm{SNR}\to\infty} \frac{C_{s,r,d}}{\log(\mathrm{SNR})} = \frac{1}{\frac{1}{\eta_{s,r}} + \frac{1}{\eta_{r,d}}}. \tag{2}$$

- *Separate Relaying*: The feeder link and access link are both M-user interference channels, so both have degrees of freedom of either $\eta_{s,r} = \eta_{r,d} = 1$ by TDMA. The number of degrees of freedom of the two-hop network is therefore $\eta^{(\text{separate})}_{s,r,d} = 1/2$ by using (2).
- *Shared Relaying*: The feeder link and access link are a multiple access channel (MAC) and a broadcast channel (BC), respectively. So the degrees of freedom of both links are $\eta_{s,r} = \eta_{r,d} = M$. Using (2), the number of degrees of freedom is $\eta^{(\text{shared})}_{s,r,d} = M/2$ for the two-hop link.

The above analysis shows that a shared relay system can achieve M times the number of degrees of freedom of a separate relay system. To realize this theoretical gain, we next turn to practical scheduling and resource allocation methods for the shared relaying system.

[1]A channel has η degrees of freedom if its capacity can be expressed as $C(\mathrm{SNR}) = \eta \log(\mathrm{SNR}) + o(\log(\mathrm{SNR}))$.

III. General Optimization Framework

A. System Model

Consider a cellular network where each cell is divided into $M = 3$ sectors where users are uniformly distributed. Downlink decode-and-forward strategy is assumed. The base-station-to-user and the relay-to-user links are both used for user scheduling, and are called the *access links.* The base-station-to-relay links, called the *feeder links*, provide the wireless backhaul connection. Users can choose direct transmission from a base-station or indirectly via the relay. Depending on this routing choice, the set of users in sector m, denoted as $\mathcal{K}^{(m)}$, is partitioned into the *one-hop* user set $\mathcal{K}_s^{(m)}$ and the *two-hop* user set $\mathcal{K}_r^{(m)}$. The base-stations, separate relays and users are equipped with a single antenna each, while the shared relay has M antennas covering a cluster of M adjacent coordinated sectors. Only the downlink transmission in the central cluster (the central M sectors) in Fig. 1 are considered in our formulation. The out-of-cluster interference is relatively weak compared to the intra-cluster transmission, but is explicitly modeled in the simulation part.

We adopt a half-duplex OFDMA frame structure as shown in Fig. 3. In the first phase, the base-station transmits to the users and the relay on orthogonal subchannels. In the second phase, the base-station and the relay simultaneously transmit to separate users on all subchannels to maximize frequency reuse. As two-hop users typically have weak direct links, we assume that they do not combine signals of the two phases.

B. Transmit and Receive Strategy

The multi-antenna shared relay in this paper uses linear MIMO processing. In the first phase, minimum-mean-squared-error (MMSE) receiver is used to spatially separate the signals from M base-stations. In the second phase, ZF beamforming is used for transmission from the shared relay to multiple users.

The signal-to-interference-plus-noise ratio (SINR) expressions for various links are described as follows.

1) Base-Station-to-Relay Feeder Link: The SINR of the wireless backhaul connecting the base-station m and the shared relay on subchannel n can be expressed as

$$\gamma_{s_m,r}^{(n)} = \frac{P_{s_m}^{(n)} \left| \left(\mathbf{v}_{s_m,r}^{(n)} \right)^H \mathbf{h}_{s_m,r}^{(n)} \right|^2}{\sigma_r^2 \left\| \mathbf{v}_{s_m,r}^{(n)} \right\|^2 + \sum_{j \neq m} P_{s_j}^{(n)} \left| \left(\mathbf{v}_{s_m,r}^{(n)} \right)^H \mathbf{h}_{s_j,r}^{(n)} \right|^2} \quad (3)$$

where $P_{s_m}^{(n)}$ is the transmit power of base-station m, $\mathbf{h}_{s_m,r}^{(n)} \in \mathbb{C}^{M \times 1}$ is the channel vector from base-station m to the shared relay, $\left(\mathbf{v}_{s_m,r}^{(n)} \right)^H \in \mathbb{C}^{1 \times M}$ is the corresponding receive beamformer, and σ_r^2 is the combined out-of-cluster interference and noise at the shared relay. The optimal MMSE receive beamformer at the shared relay is

$$\mathbf{v}_{s_m,r}^{(n)} = \left(\sigma_r^2 \mathbf{I} + \sum_{j \neq m} P_{s_j}^{(n)} \mathbf{h}_{s_j,r}^{(n)} \left(\mathbf{h}_{s_j,r}^{(n)} \right)^H \right)^{-1} \mathbf{h}_{s_m,r}^{(n)} \quad (4)$$

which suppresses mutual interference among M base-stations and maximizes the receiver SINR.

2) Relay-to-Two-Hop-User Access Link: The shared relay schedules up to M two-hop users in the second phase across M sectors per subchannel. Note that the shared relay is not limited to choosing exactly one user per sector, i.e., the relay may schedule multiple users in one sector while no user in another sector. Let $\mathcal{S}_r^{(n)}$ be the selected user set, so $\left| \mathcal{S}_r^{(n)} \right| \leq M$. ZF beamforming is used to eliminate the mutual interference among users in $\mathcal{S}_r^{(n)}$. The SINR of the link between the shared relay and a scheduled two-hop user k on subchannel n is

$$\begin{aligned} \gamma_{r,k}^{(n)} =& \frac{P_{r,k}^{(n)} \left| \left(\mathbf{h}_{r,k}^{(n)} \right)^H \hat{\mathbf{w}}_{r,k}^{(n)} \right|^2}{\sigma_k^2 + \sum_j P_{s_j}^{(n)} \left| h_{s_j,k}^{(n)} \right|^2 + \sum_{j \in \mathcal{S}_r^{(n)}, j \neq k} P_{r,j}^{(n)} \left| \left(\mathbf{h}_{r,k}^{(n)} \right)^H \hat{\mathbf{w}}_{r,j}^{(n)} \right|^2} \\ \overset{\text{ZF}}{=}& \frac{P_{r,k}^{(n)}}{\left\| \mathbf{w}_{r,k}^{(n)} \right\|^2 \left(\sigma_k^2 + \sum_j P_{s_j}^{(n)} \left| h_{s_j,k}^{(n)} \right|^2 \right)} \end{aligned} \quad (5)$$

where $P_{r,k}^{(n)}$ is the shared relay's power allocation for user k, $\left(\mathbf{h}_{r,k}^{(n)} \right)^H \in \mathbb{C}^{1 \times M}$ is the channel vector from the shared relay to the user k, $\mathbf{w}_{r,k}^{(n)}$ and $\hat{\mathbf{w}}_{r,k}^{(n)} \in \mathbb{C}^{M \times 1}$ are the ZF transmit beamformer and its normalized version, $h_{s_j,k}^{(n)}$ is the channel response between base-station j and user k, and σ_k^2 is the combined out-of-cluster interference and noise at user k. The second equality in (5) comes from the fact that

$$\left(\mathbf{h}_{r,k}^{(n)} \right)^H \hat{\mathbf{w}}_{r,k}^{(n)} = \left(\mathbf{h}_{r,k}^{(n)} \right)^H \frac{\mathbf{w}_{r,k}^{(n)}}{\left\| \mathbf{w}_{r,k}^{(n)} \right\|} = \frac{1}{\left\| \mathbf{w}_{r,k}^{(n)} \right\|} \quad (6a)$$

$$\left(\mathbf{h}_{r,k}^{(n)} \right)^H \hat{\mathbf{w}}_{r,j}^{(n)} = 0. \quad (6b)$$

Note that the intra-cluster interference for the two-hop users come from the base-station frequency reuse; the interference from the relay is eliminated by ZF beamforming.

3) Base-Station-to-One-Hop-User Access Link: The one-hop users can be scheduled by the base-stations in both phases. The SINR of the link between the base-station m and a one-hop user k on subchannel n in the i-th phase is

$$\gamma_{s_m,k}^{(n,i)} = \frac{P_{s_m}^{(n)} \left| h_{s_m,k}^{(n)} \right|^2}{\sigma_k^2 + \sum_{j \neq m} P_{s_j}^{(n)} \left| h_{s_j,k}^{(n)} \right|^2 + \Delta_k^{(n,i)}} \quad (7)$$

where

$$\Delta_k^{(n,i)} = \begin{cases} 0 & i = 1 \quad (8a) \\ \sum_{j \in \mathcal{S}_r^{(n)}} P_{r,j}^{(n)} \left| \left(\mathbf{h}_{r,k}^{(n)} \right)^H \hat{\mathbf{w}}_{r,j}^{(n)} \right|^2 & i = 2 \quad (8b) \end{cases}$$

for the given beamforming vectors $\hat{\mathbf{w}}_{r,j}^{(n)}$ at the shared relay. Note that the inter-user interference term $\Delta_k^{(n,i)}$ only exists in the second phase when the relay is transmitting. This is due to frequency reuse at the relay and at neighboring base-stations.

The SINRs of each link on each subchannel can be mapped to the transmission rate by $r = \log_2(1 + \gamma/\Gamma)$, where $\Gamma = -\ln(5\text{BER})/1.5$ is the SNR gap corresponding to a certain target bit-error-rate (BER). Thus the per-subchannel rate $r_{s_m,r}^{(n)}$, $r_{r,k}^{(n)}$, and $r_{s_m,k}^{(n,i)}$ can be computed using the above relation from their related SINR formula.

C. Utility Maximization Framework

The objective is to maximize the network PF utility [7]:

$$\max \sum_{m=1}^{M} \sum_{k \in \mathcal{K}^{(m)}} \ln(R_k) \quad (9)$$

where R_k is user k's long-term average rate, which is updated using exponential averaging. Proportionally fair scheduling can be implemented in practice using a weighted rate sum maximization formulation [8]:

$$\max \sum_{m=1}^{M} \sum_{k \in \mathcal{K}^{(m)}} \alpha_k r_k, \quad \text{where} \quad \alpha_k = \frac{1}{R_k}. \quad (10)$$

and r_k is user k's instantaneous rate.

IV. Resource Allocation and Scheduling for Shared Relaying

A. Problem Formulation

Users are first partitioned into the one-hop set $\mathcal{K}_s^{(m)}$ and the two-hop set $\mathcal{K}_r^{(m)}$ in each sector m, which is also known as routing. In this paper, we adopt an intra-sector routing metric where users are partitioned based on their received signal strength from base-stations and relays.

Define binary indicator variables $\rho_{s_m,k}^{(n,i)}$ and $\rho_{r,k}^{(n)}$, such that $\rho_{s_m,k}^{(n,i)} = 1$ indicates that the base-station m schedules the user k on subchannel n in the i-th phase, and $\rho_{r,k}^{(n)} = 1$ indicates that the shared relay schedules the user k on subchannel n. The one-hop users are served by the base-station in either or both of the two phases, but only one user is scheduled on each subchannel in every sector in each phase, so

$$\sum_{k\in\mathcal{K}_s^{(m)}} \rho_{s_m,k}^{(n,i)} \le 1,\ \ \rho_{s_m,k}^{(n,i)} \in \{0,1\},\ \ \forall m,n,i\in\{1,2\}. \quad (11)$$

The two-hop users are served by the relay; a maximum of M users are served simultaneously with spatial multiplexing, so

$$\sum_{m=1}^{M}\sum_{k\in\mathcal{K}_r^{(m)}} \rho_{r,k}^{(n)} \le M, \quad \rho_{r,k}^{(n)} \in \{0,1\}, \quad \forall n. \qquad (12)$$

The user rate in (10) can now be expressed as

$$r_k = \begin{cases} \displaystyle\sum_{n=1}^{N}\sum_{i=1}^{2}\rho_{s_m,k}^{(n,i)} r_{s_m,k}^{(n,i)} & k\in\mathcal{K}_s^{(m)} \quad (13a) \\ \displaystyle\sum_{n=1}^{N}\rho_{r,k}^{(n)} r_{r,k}^{(n)} & k\in\mathcal{K}_r^{(m)}. \quad (13b)\end{cases}$$

We assume that the data received by the relay in each phase must be forwarded to the users at the next phase, and no buffering at the relay is considered. This assumption is valid for delay-sensitive services. Then we have the following wireless backhaul constraint: the total *rate demand* of the shared relay for all of its serving users in sector m in the access link, denoted as $R_{r,d}^{(m)}$, should be no larger than the total *rate supply* in the feeder link from the base-station m to this relay, denoted as $R_{s,r}^{(m)}$, i.e.,

$$\begin{aligned} R_{r,d}^{(m)} &\triangleq \sum_{k\in\mathcal{K}_r^{(m)}}\sum_{n=1}^{N}\rho_{r,k}^{(n)} r_{r,k}^{(n)} \\ &\le \sum_{n=1}^{N}\left(1-\sum_{k\in\mathcal{K}_s^{(m)}}\rho_{s_m,k}^{(n,1)}\right) r_{s_m,r}^{(n)} \triangleq R_{s,r}^{(m)}. \end{aligned} \qquad (14)$$

Note that in the first phase, if a subchannel is used for the wireless backhaul transmission, then it is not used for the scheduling of one-hop users of the base-station, i.e., $\rho_{s_m,k}^{(n,1)} = 0$, $\forall k$. Similar backhaul constraint is adopted in previous works for separate relaying system, e.g., [9] for sum rate maximization, and [10] for proportional fairness where the backhaul constraint is considered in a per-user basis.

In this paper, both the base-station and the relay have a fixed power spectral density (PSD). In particular, the shared relay may optimally allocate its per-subchannel power $P_{r,k}^{(n)}$ among all of its scheduled users under the total PSD mask

$$\sum_{m=1}^{M}\sum_{\left\{k\in\mathcal{K}_r^{(m)}\middle|\rho_{r,k}^{(n)}=1\right\}} P_{r,k}^{(n)} = P_r^{\max}, \quad \forall n. \qquad (15)$$

The resource allocation problem can now be formulated as maximizing (10) subject to constraints (11)-(15), where r_k as defined in (13) accounts for all the inter-node interference. The maximization is over the variables of scheduling indicator $\boldsymbol{\rho} = \{\{\rho_{s_m,k}^{(n,i)}\}_{m,k,n,i}, \{\rho_{r,k}^{(n)}\}_{k,n}\}$, the relay's beamforming vector $\mathbf{w} = \{\hat{\mathbf{w}}_{r,k}^{(n)}\}_{k,n}$, and its power allocation $\mathbf{P} = \{P_{r,k}^{(n)}\}_{k,n}$.

B. Resource Allocation and Scheduling

The first step is to dualize (10) with respect to the wireless backhaul constraints (14) for all M sectors:

$$\begin{aligned} g(\boldsymbol{\rho},\mathbf{w},\mathbf{P},\boldsymbol{\lambda}) = &\sum_{m=1}^{M}\left(\sum_{k\in\mathcal{K}_s^{(m)}}\alpha_k r_k + \sum_{k\in\mathcal{K}_r^{(m)}}\alpha_k r_k\right) \\ &+ \sum_{m=1}^{M}\lambda^{(m)}\left(R_{s,r}^{(m)} - R_{r,d}^{(m)}\right) \end{aligned} \qquad (16)$$

where $\lambda^{(m)}$ is the dual variable representing the wireless backhaul price, which coordinates the shared relay's rate demand and supply. Plugging (13) and (14) into (16), we have

$$\begin{aligned} &g(\boldsymbol{\rho},\mathbf{w},\mathbf{P},\boldsymbol{\lambda}) \\ &= \sum_{n=1}^{N}\sum_{m=1}^{M}\Bigg\{\sum_{k\in\mathcal{K}_s^{(m)}}\left[\rho_{s_m,k}^{(n,1)}A_k^{(m,n)} + \rho_{s_m,k}^{(n,2)}B_k^{(m,n)}\right] \\ &\quad + \sum_{k\in\mathcal{K}_r^{(m)}}\rho_{r,k}^{(n)}C_k^{(m,n)}\Bigg\} + \sum_{m=1}^{M}\lambda^{(m)}\sum_{n=1}^{N} r_{s_m,r}^{(n)} \end{aligned} \qquad (17)$$

where $A_k^{(m,n)}$, $B_k^{(m,n)}$, and $C_k^{(m,n)}$ are:

$$A_k^{(m,n)} = \alpha_k r_{s_m,k}^{(n,1)} - \lambda^{(m)} r_{s_m,r}^{(n)} \qquad (18a)$$

$$B_k^{(m,n)} = \alpha_k r_{s_m,k}^{(n,2)} \qquad (18b)$$

$$C_k^{(m,n)} = \left(\alpha_k - \lambda^{(m)}\right) r_{r,k}^{(n)}. \qquad (18c)$$

subject to the rest of the constraints (11)(12)(15). The Lagrangian function (17) can now be decoupled into per-subchannel maximization subproblems where the scheduling indicators are set to be the users with the maximum positive value of $A_k^{(m,n)}$, $B_k^{(m,n)}$, and $C_k^{(m,n)}$.

The term $B_k^{(m,n)}$ is independent of $\lambda^{(m)}$ (unlike $A_k^{(m,n)}$ and $C_k^{(m,n)}$). Consequently the user scheduling at the base-station in the second phase is straightforward, while the user scheduling at the base-station in the first phase and that at the relay require iterative search of the backhaul price $\lambda^{(m)}$. The overall scheduling rule for the base-station and the relay is outlined below:

Algorithm 1: User scheduling for the shared relay system on each subchannel n for given $\lambda^{(m)}$'s.

(a) *Base-station m at the first phase:* Select the user $\hat{k} = \arg\max_{k\in\mathcal{K}_s^{(m)}}\left\{\alpha_k r_{s_m,k}^{(n,1)}\right\}$. If $\alpha_{\hat{k}} r_{s_m,\hat{k}}^{(n,1)} > \lambda^{(m)} r_{s_m,r}^{(n)}$, schedule this user $\hat{k}$; otherwise this subchannel is used for wireless backhaul transmission in the feeder link.

(b) *Shared Relay at the second phase:* The user scheduling at the shared relay should be jointly considered with ZF beamforming and power allocation. The objective is to select up to M users with maximum positive value of $(\alpha_k - \lambda^{(m)})r_{r,k}^{(n)}$. The details will be explained shortly.

(c) *Base-station m at the second phase:* Select the user $\hat{k} = \arg\max_{k \in \mathcal{K}_s^{(m)}} \left\{ \alpha_k r_{s_m,k}^{(n,2)} \right\}$ to schedule.

Part (a) determines the resource allocation competition between the one-hop users and the relay feeder link in the first phase. Part (b) and (c) determine how users are scheduled in the second phase. Since the base-station has a fixed PSD, while the shared relay can allocation powers across beamformer, the relay-to-user rate $r_{r,k}^{(n)}$ contains a fixed level of interference from the base-stations, while the base-station-to-user rate $r_{s_m,k}^{(n,2)}$ is a function of the relay power allocation $P_{r,k}^{(n)}$ and the beamformer $\hat{\mathbf{w}}_{r,k}^{(n)}$ according to (7). Hence, part (b) needs to be executed before part (c).

Part (b) involves selecting users and finding the ZF beamformers and power allocations to maximize the weighted sum rate with weights $\alpha_k - \lambda^{(m)}$. This is a conventional multiuser MIMO problem for which many practical solutions exist. This paper adapts an approach of [11]. The basic idea is to ensure semi-orthogonality among the selected users. First let $\bigcup_m \left\{ k \in \mathcal{K}_r^{(m)} \middle| \alpha_k > \lambda^{(m)} \right\}$ be the initial candidate user set for relay scheduling in all M sectors, as the selected users should have positive values for $C_k^{(m,n)}$. Each step of the algorithm first finds the orthogonal component of the current candidate users' channel with respect to the subspace spanned by the channels of previously selected users, then approximates the per-user weighted rate using the orthogonal component of the channel, equally distributed power, and weights $\alpha_k - \lambda^{(m)}$. Based on the approximated weighted rate, the relay selects up to M users across M sectors in a greedy manner in each subchannel. The detailed description of the relay user selection is omitted here for brevity. Finally, with scheduling fixed, the ZF beamformer $\hat{\mathbf{w}}_{r,k}^{(n)}$ can be obtained vie matrix inversion. Given user scheduling and beamforming results, and since $r_{r,k}^{(n)} = \log_2 \left(1 + \gamma_{r,k}^{(n)} / \Gamma \right)$ where $\gamma_{r,k}^{(n)}$ is given in (5), the optimal power allocation can be calculated by waterfilling, but modified by the weights. For a user in sector m which is scheduled by the shared relay on subchannel n, the transmit power $P_{r,k}^{(n)}$ is

$$P_{r,k}^{(n)} = \left[\frac{\alpha_k - \lambda^{(m)}}{\mu \ln(2)} - \left\| \mathbf{w}_{r,k}^{(n)} \right\|^2 \Gamma \left(\sigma_k^2 + \sum_j P_{s_j}^{(n)} \left| h_{s_j,k}^{(n)} \right|^2 \right) \right]_+ \quad (19)$$

where $[x]_+ = \max(x, 0)$ and the water level μ is chosen to satisfy the power constraint (15).

C. Update of the Lagrangian Dual Variables

The final component of the algorithm is to find the Lagrangian price $\boldsymbol{\lambda} = \left\{ \lambda^{(m)} \right\}_{m=1}^M$ for the dual function

$$q(\boldsymbol{\lambda}) = \max_{\boldsymbol{\rho}, \mathbf{w}, \mathbf{P}} g(\boldsymbol{\rho}, \mathbf{w}, \mathbf{P}, \boldsymbol{\lambda}) \quad (20)$$

such that the wireless backhaul constraints (14) for all the sectors are satisfied, and the dual objective (20) is minimized.

Intuitively, $\lambda^{(m)}$ balances the base-station-to-relay rate supply $R_{s,r}^{(m)}$ and the relay-to-user rate demand $R_{r,d}^{(m)}$. Note that $\lambda^{(m)}$ is upper bounded by $\lambda_{\max}^{(m)} = \max_{k \in \mathcal{K}_r^{(m)}} \alpha_k$, in which case $\alpha_k - \lambda_{\max}^{(m)} \leq 0$ holds for all $k \in \mathcal{K}_r^{(m)}$, consequently relay does not schedule any users and $R_{r,d}^{(m)} = 0$. Meanwhile $\lambda^{(m)}$ is lower bounded by $\lambda_{\min}^{(m)} = 0$, in which case no subchannel is used for relay's feeder link transmission at the first phase, and $R_{s,r}^{(m)} = 0$. To update $\boldsymbol{\lambda}$, first compute $R_{r,d}^{(m)}$ and $R_{s,r}^{(m)}$ in (14) using $\rho_{s_m,k}^{(n,i)}$, $\rho_{r,k}^{(n)}$, $\mathbf{w}$, and $\mathbf{P}$ obtained from part (a) and (b) of *Algorithm 1*, then use subgradient projection, i.e., $\lambda^{(m)} = \left[\lambda^{(m)} - s^{(m)} \left(R_{s,r}^{(m)} - R_{r,d}^{(m)} \right) \right]_{\lambda_{\min}^{(m)}}^{\lambda_{\max}^{(m)}}$ where $s^{(m)}$ is the step size and $[x]_b^a = \min(\max(x, b), a)$.

V. Performance Evaluation

We evaluate the performance of shared relaying as compared to separate relaying in a sectorized cellular networks shown in Fig. 1. The cell radius is 1km, and total bandwidth of 10MHz is divided into 64 orthogonal subchannels. The separate relays are placed at $2/3$ of the cell radius from the base-station [12]. The per-sector user number is denoted as $K = \left| \mathcal{K}^{(m)} \right|$. One tier of out-of-cluster interference is considered, where separate and shared relays are deployed whenever applicable. The base-station's transmit power is $P_s = 46$ dBm. The separate relay's power is P_r, and the comparable shared relay's power is $3P_r$ (as a shared relay with three antennas can be thought of as a combination of three separate relays). The path loss of the access link is $L = 128.1 + 37.6 \log_{10}(d)$ dB (d is in km), with a 8-dB lognormal shadowing and Rayleigh fading. The path loss of the feeder link is $L = 128.1 + 28.8 \log_{10}(d)$ dB (d is in km), with a 4-dB lognormal shadowing and Rician fading with 10-dB Rician factor. The relay receives from its donor base-station with a highly directive antenna in the feeder link ($\theta_{3dB} = 20^o$ [13]), and transmits to users with an omni-directional antenna.

For comparison, the separate relaying is also evaluated with a backhaul-aware scheduler. The resource allocation of the separate relaying is similar to the shared relaying in part (a) and (c) of *Algorithm 1*, while in part (b) each relay m simply schedules one user with the maximum positive weighted rate with the modified weights $\left(\alpha_k - \lambda^{(m)} \right)$. For separate relaying the scheduling of individual relay in each sector is independent, and the rate supply $R_{s,r}^{(m)}$ and and demand $R_{r,d}^{(m)}$ have monotone increasing and decreasing relationships with backhaul price $\lambda^{(m)}$, respectively. Thus it is possible to use bisection search independently in each sector to find the proper $\lambda^{(m)}$ that satisfies the constraint (14), with the same upper and lower bounds of the prices as in the shared relay system.

The performances of different schemes are examined in Table I. It is clear that shared relaying with the proposed algorithm has the best performance in all metrics, where the largest improvement is found for the 5% rate, corresponding to the typical cell edge performance. The throughput cumulative distribution function (CDF) of shared relaying is compared with the separate relaying and no-relay networks in Fig. 4.

It is instructive to analyze the network performance as a function of the relay power as in Fig. 5 and Fig. 6. It is

TABLE I
COMPARISON OF DIFFERENT SCHEMES, WITH $P_r = \frac{1}{3}P_s$ AND $K = 40$.

Scheme \ Metric	Sum Rate[a]	5% Rate[a]	Utility[b]
No Relay	16.33	0.0203	-63.98
Separate Relay	19.23	0.0230	-56.74
Shared Relay	21.10	0.0398	-48.27
Shared over Separate	+10%	+73%	–

[a] measured in Mbps.
[b] computed with (9) as a function of per-user rates in Mbps.

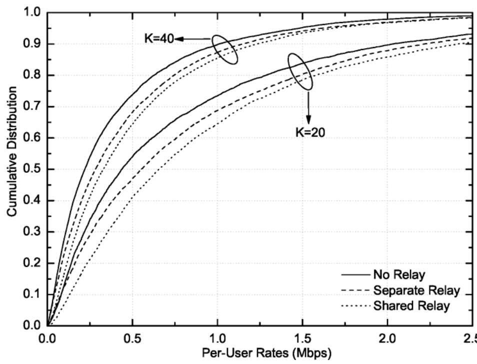

Fig. 4. Comparison of separate relay and shared relay, with $P_r = \frac{1}{3}P_s$

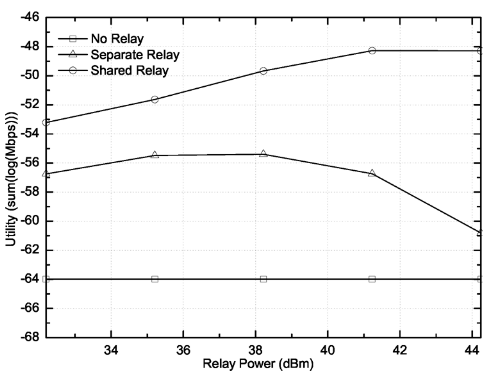

Fig. 5. Per-sector utility as a function of P_r with $K = 40$

observed that the utility of separate relaying first increases with relay power, then decreases significantly with the increased relay power; while the utility of shared relaying increases with relay power and then stays at a constant. Similar trends can be found for cell edge rate (although at very high relay power, the shared relay scenario also sees a decreasing edge rate due to the rising interference from out-of-cluster relays). This overall trend illustrates the shared relay's interference mitigation ability. At high power, the cell-edge performance of separate relaying is hampered by intra-cluster interference, while the shared relay is able to mitigate such interference and thus take advantage of the increased power.

VI. CONCLUSION

This paper illustrates the benefit of shared relaying from the system perspective. To realize the full potential of shared relaying, scheduling and resource allocation algorithms under

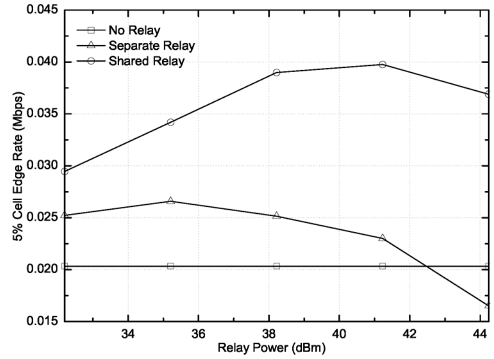

Fig. 6. Per-sector 5% cell edge rate as a function of P_r with $K = 40$

the network utility maximization framework are proposed. The proposed algorithm uses a set of backhaul prices to balance the rate supply and demand at the relay, and to coordinate the backhaul-aware user scheduling at the base-station and the joint scheduling, beamforming, and power allocation at the relay. The simulations illustrate that shared relaying is effective in mitigating intercell interference and in improving the overall system performance in terms of both utility and rates.

REFERENCES

[1] M. Liu, X. Chang, Y. Lu, and H. Si, "A novel network structure based on multi cell coordinated relay," IEEE C802.16m-08/029, Jan. 2008.
[2] Y. Song, H. Yang, J. Liu, L. Cai, D. Li, X. Zhu, K. Wu, and H. Liu, "Relay station shared by multiple base stations for inter-cell interference mitigation," IEEE C802.16m-08/1436r1, Nov. 2008.
[3] S. Peters, A. Panah, K. Truong, and R. Heath, "Relay architectures for 3GPP LTE-advanced," *EURASIP J. Wireless Commun. Netw.*, vol. 2009, pp. 1–14, 2009.
[4] A. Panah, K. Truong, S. Peters, and R. Heath, "Interference management schemes for the shared relay concept," *EURASIP J. Advances in Signal Process.*, vol. 2011, pp. 1–14, 2011.
[5] J. Kim, J. Hwang, and Y. Han, "Joint processing in multi-cell coordinated shared relay network," in *Proc. IEEE Int. Symp. Personal, Indoor and Mobile Radio Commun. (PIMRC)*, Sept. 2010, pp. 702–706.
[6] O. Oyman, N. Laneman, and S. Sandhu, "Multihop relaying for broadband wireless mesh networks: from theory to practice," *IEEE Commun. Magazine*, vol. 45, no. 11, pp. 116–122, Nov. 2007.
[7] F. Kelly, "Charging and rate control for elastic traffic," *European Trans. Telecom.*, vol. 8, no. 1, pp. 33–37, 1997.
[8] E. Chaponniere, P. Black, J. Holtzman, and D. Tse, "Transmitter directed, multiple receiver system using path diversity to equitably maximize throughput," U.S. patent, July 1999.
[9] W. Nam, W. Chang, S. Chung, and Y. Lee, "Transmit optimization for relay-based cellular OFDMA systems," in *Proc. IEEE Int. Conf. Commun. (ICC)*, June 2007, pp. 5714–5719.
[10] Y. Cui, V. Lau, and R. Wang, "Distributive subband allocation, power and rate control for relay-assisted OFDMA cellular system with imperfect system state knowledge," *IEEE Trans. Wireless Commun.*, vol. 8, no. 10, pp. 5096–5102, Oct. 2009.
[11] T. Yoo and A. Goldsmith, "On the optimality of multiantenna broadcast scheduling using zero-forcing beamforming," *IEEE J. Sel. Areas Commun.*, vol. 24, no. 3, pp. 528–541, Mar. 2006.
[12] Y. Liu, R. Hoshyar, X. Yang, and R. Tafazolli, "Integrated radio resource allocation for multihop cellular networks with fixed relay stations," *IEEE J. Sel. Areas Commun.*, vol. 24, no. 11, pp. 2137–2146, Nov. 2006.
[13] M. Salem, A. Adinoyi, M. Rahman, H. Yanikomeroglu, D. Falconer, and Y. Kim, "Fairness-aware radio resource management in downlink OFDMA cellular relay networks," *IEEE Trans. Wireless Commun.*, vol. 9, no. 5, pp. 1628–1639, May 2010.