Resource Allocation with Flexible Channel Cooperation in Cognitive Radio Networks

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

Abstract—We study the resource allocation problem in an OFDMA based cooperative cognitive radio network, where secondary users relay data for primary users in order to gain access to the spectrum. In light of user and channel diversity, we first propose FLEC, a novel flexible channel cooperation scheme. It allows secondary users to freely optimize the use of channels for transmitting primary data along with their own, in order to maximize performance. Further, we formulate a unifying optimization framework based on Nash bargaining solutions to fairly and efficiently allocate resources between primary and secondary networks, in both decentralized and centralized settings. We present an optimal distributed algorithm and a sub-optimal centralized heuristic, and verify their effectiveness via realistic simulations. Under the same framework, we also study conventional identical channel cooperation as the performance benchmark, and propose algorithms to solve the corresponding optimization problems.

Index Terms—Cognitive radio, cooperative communication, resource allocation, Nash bargaining solutions, OFDMA.

1 Introduction

Cognitive radio, with the ability to flexibly adapt its transmission parameters, has been considered a revolutionary technology to open up dynamic access to the under-utilized wireless spectrum [2], [3]. Recently, a new paradigm where primary users (PUs) can leverage secondary users (SUs) for their own transmissions, termed *cooperative cognitive radio networks* (CCRN), is advocated [4], [5]. In CCRN, SUs cooperatively relay data for PUs in order to access the spectrum. Assuming that SUs have better channel conditions to the primary receiver, cooperative relaying can greatly increase the primary transmission rate. Meanwhile, SUs also gain opportunities to access the spectrum, resulting in a "win-win" situation.

A single channel network with only one PU has been considered in [4], [5]. The PU leases its channel to SUs for a fraction of time in exchange for cooperative transmission. SUs allocate a portion out of their time fraction for relaying primary data, and the rest for their own traffic. A Stackelberg game is formulated to determine the optimal time sharing strategy.

In this paper, we investigate cooperative cognitive radio networks from a new perspective. We consider multi-channel cellular networks based on OFDMA, e.g. IEEE 802.16 [6] for the primary network, with multiple SUs assisting multiple PUs on the uplink. Multi-channel networks impose unique challenges of realizing the cooperative paradigm, as we narrate below along with our original contributions.

First, we observe that conventional user cooperation permeated through the literature [7] becomes inefficient when directly applied to multi-channel CCRN. It implicitly postulates that data on one channel has to be relayed on exactly the same channel, which may not be amenable to relaying from a performance perspective. Meanwhile, some other channel may have abundant capacity to incorporate additional data with little cost. In other words, cooperation using the same channel misses the bulk of PU-SU cooperation opportunities, by unnecessarily limiting the space of SU resource allocation to only the temporal dimension.

Our *first* contribution in this paper is a new design for cooperation among SUs and PUs, termed Flexible Channel Cooperation (FLEC), that opens up all dimensions of resource allocation for SUs. It takes advantage of channel and user diversities available in multi-channel networks [8], [9], and allows SUs to freely *optimize* its use of resources, including channels and time slots leased by PUs, as well as power, for relaying primary data along with its own data, as long as all the primary data it received can be delivered.

The basic idea of FLEC works as shown in Fig. 1. We consider the simplified case where time is equally divided into two slots among cooperating users[1]. PUs transmit in the first slot to SUs, and SUs transmit in the second to the primary base station (BS) and to their own access point (AP). A SU strategically optimizes its use of the leased resources. For example, it can use subchannel 1 solely for relaying data aggregated from both subchannel 1 and 2, and use subchannel 2 solely for sending its own data as in Fig. 1. The intuition is that, if subchannel 1 has superior conditions on the SU-BS link but poor conditions on the SU-AP link, it is much more efficient using subchannel 1 to relay data from both subchannels. Such channel *swapping* or *shuffling* results in

- *This work has been presented in part at IEEE INFOCOM, San Diego, CA, March 2010 [1].*
- *The authors are with The Edward S. Rogers Sr. Department of Electrical and Computer Engineering, University of Toronto, Toronto, Ontario M5S 3G4, Canada. Email:* {henryxu, bli}*@eecg.toronto.edu.*

1. Practical justifications for this simplification are as follows. Due to channel diversity, the optimal time sharing strategy is considerably different across the channels for a given pair of PU-SU. From a system perspective, it becomes difficult to structure the uplink bursts in the frames, because transmissions on some subchannels will finish earlier than those on other subchannels.

boosted SU throughput, as well as larger relay capacity for PU, since the overall spectral efficiency is improved. The spectral efficiency gain can in turn be translated into more cooperation opportunities, as well as increased network capacity and better performance.

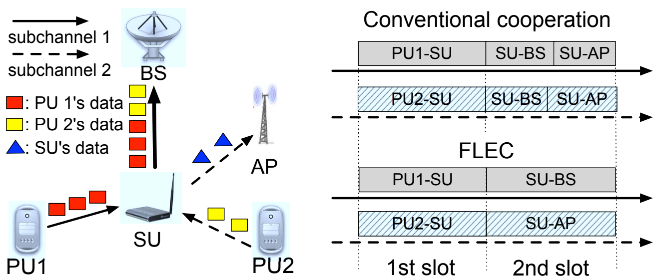


Fig. 1. The motivating scenario for *Flexible Channel Cooperation (FLEC)*.

The preceding description assumes a decentralized setting where the primary and secondary networks are independent. Subchannels are assigned to PUs by the primary BS *a priori* to SU cooperation, and only those assigned to the helped PUs are leased to the respective helping SUs. In a centralized setting where SU cooperation becomes an integral part of the resource allocation performed by the primary BS, it becomes possible to assign any subchannel to a helping SU, to further improve the performance. We also consider the centralized FLEC in our paper, which turns out to be more difficult.

The *second* challenge in multi-channel CCRN is how to schedule the transmissions and allocate resources, in order to maximize performance gains while ensuring fairness among all users. A SU may assist several PUs (as in Fig. 1) simultaneously while a PU may also pair up with several SUs, complicating the resource allocation problem. Moreover, in reality, PUs and SUs are selfish in maximizing their own utility. PUs compete among themselves when one SU resides in a suitable position to relay for all of them; likewise SUs compete among themselves if one channel has good conditions for all of them. Our main objective in this paper, therefore, is to develop *efficient* yet *fair* resource allocation algorithms for FLEC in multi-channel networks, which has not been addressed yet.

To this end, our *second* contribution is a novel unifying optimization framework that jointly considers relay and subchannel assignment, relay strategy optimization, and power control, based on the concept of Nash bargaining [10]. PUs and SUs agree to jointly optimize a social cost function, known as the Nash product, which is essentially the product of utility functions of the cooperating PUs and SUs. The solution concept, known as the Nash bargaining solution (NBS), is a unique Nash equilibrium point that is guaranteed to provide Pareto efficiency with NBS fairness among PUs and SUs, which is a generalized proportional fairness notion [11]. Therefore, gains from cooperation to individual PU and SU are allocated *proportionally* according to their channel conditions, i.e. their contributions to the social welfare gain. These properties make NBS favorable in our problem.

We consider both decentralized and centralized FLEC as introduced above. In the decentralized case, we wish to develop a distributed algorithm that can be performed by users independently with local information only. We tackle this using a dual decomposition technique to transform the global optimization into many per-subchannel problems that can be solved by the respective PUs distributively and optimally. To account for SUs' utility, we rely on the subgradient method [12] to allow PUs to bargain with neighboring SUs autonomously to arrive at the optimal solution for the per-subchannel problem, i.e. the Nash bargaining solution.

In the centralized case, cooperation opportunities are to be carefully invented and engineered, rather than distributively harvested. We identify the inefficiency of subgradient method in this problem, design a three-step heuristic via a decoupling approach, and prove the approximation ratio for the decoupled subchannel assignment algorithm. Both algorithms are highly efficient in that they can meet typical scheduling deadlines of 5–10 ms [6] in OFDMA systems. In addition, we extend our framework to consider resource allocation with conventional identical channel cooperation to complete the analysis. Thus, we believe our work sheds light on the design and implementation of OFDMA based cooperative cognitive radio networks.

The remainder of this paper is structured as follows. Sec. 2 introduces our system models and the concepts of NBS. In Sec. 3 we formulate the resource allocation problem in decentralized setting and present optimal distributed algorithms to solve it. In Sec. 4, we consider the centralized version of the problem and propose practical algorithms with performance guarantees. We study the conventional identical channel cooperation in Sec. 5. We conduct extensive simulations to verify our algorithms in Sec. 6 and summarize related work in Sec. 7. We finally give concluding remarks in Sec. 8.

2 An Optimization Framework

2.1 System Model

We start by introducing the system model. We consider the uplink of a single-cell OFDMA network. We do not consider an ad-hoc network where coordination between PUs and SUs, and synchronization for effective cooperative communications are difficult to achieve especially with multiple PUs. This is in line with previous work on cooperative diversity [7], [13]–[15], and on cooperative cognitive radio networks [4], [5], [16][2].

We do not model the inter-cell interference due to frequency reuse. Inter-cell interference significantly adds to the complexity of the optimization problem, and shall be dealt with as a separate issue on its own right [17]. This simplified interference model is also commonly adopted in related work [4], [5], [13], [14], [18], [19]. The throughput of uplink transmission is typically limited due to the power constraint of PUs. Thus it is better suited to employ cooperation.

A number of SUs are located in the cell and perform cooperative transmission for PUs to access the primary spectrum. We assume that PUs and SUs have infinite backlogged data to send and the OFDM frames are synchronized. Cooperative

2. Although [4] considers an ad-hoc secondary network, its model consists of only one PU who effectively coordinates the time sharing of the spectrum with SUs and the cooperative transmission.

transmissions take place on an OFDM subchannel basis, and transmissions in different subchannels do not interfere with each other. Decode-and-forward multi-hopping [7] is used when SUs relay primary data. Note that our results are readily applicable when other relaying scheme is used. Moreover, higher rates are achievable with more sophisticated coding/decoding schemes, e.g. maximum ratio combining based on the signals received in both slots at the destination (instead of multi-hopping) [7]. Here we focus on decode-and-forward multi-hopping only for simplicity of presentation. Our analysis and algorithms are readily applicable to scenarios with other relaying and coding/decoding schemes.

We model the fading environment by large scale path loss and shadowing, along with small scale frequency-selective Rayleigh fading. The coherence bandwidth is in the order of the width of a few subchannels so that adjacent subchannels have similar channel conditions. Fading between subchannels in different frames is independent, and remains stable during each frame. We assume techniques for channel estimation are employed and full channel side-information (CSI) is available, which makes the optimization possible. Such assumptions about the fading environment and CSI are commonly used as in [13], [14], [18], [19]. Noises are modeled as i.i.d. circularly symmetric complex Gaussian noises $\mathcal{CN}(0, N_0 W)$.

There are K subchannels, N_P primary users and N_S secondary users in the network. Let N be the total number of users, i.e. $N = N_P + N_S$. Let $\mathcal{K} = \{1, 2, \ldots, K\}$ be the set of subchannels, $\mathcal{N_P} = \{1, 2, \ldots, N_P\}$ the set of PUs, $\mathcal{N_S} = \{N_P + 1, N_P + 2, \ldots, N\}$ the set of SUs, and $\mathcal{N} = \mathcal{N_P} \cup \mathcal{N_S}$ the entire set of users. To denote the possibility of direct transmission, i.e. not cooperating with any SU, we denote a void SU as user $N + 1$, and let $\mathcal{N}_{\mathcal{S}}^{+} = \{N_P + 1, N_P + 2, \ldots, N + 1\}$ be the extended set of SUs. One subchannel can only be allocated to one PU, and can only be leased to one SU.

For a given PU $i \in \mathcal{N_P}$, if subchannel $c \in \mathcal{K}$ with bandwidth W and complex channel gain h_i^c is allocated for direct transmission, the achievable throughput is:

$$R_{i,N+1}^c = W \log\left(1 + p_i^c g_i^c\right), \forall c \in \mathcal{K}, i \in \mathcal{N_P}, \qquad (1)$$

where

$$g_i^c = \frac{|h_i^c|^2}{\Gamma N_0 W}.$$

As mentioned, the subscript $(N+1)$ is used to denote the direct transmission mode. Γ is the coding gap to capacity and p_i^c denotes the allocated power. Without loss of generality W equals 1 in the subsequent analysis.

If PU $i \in \mathcal{N_P}$ decided to lease $c \in \mathcal{K}$ to SU $j \in \mathcal{N_S}$ for cooperative transmission, then in the first time slot, the achievable throughput on PU-SU link is

$$R_{i,j}^c = \frac{1}{2} \log\left(1 + 2 p_i^c g_{i,j}^c\right), \forall i \in \mathcal{N_P}, j \in \mathcal{N_S}, c \in \mathcal{K}, \qquad (2)$$

since the effective power and throughput should take into account the two-slot structure of cooperative transmission. For SU j in the second time slot, under FLEC, it can freely decide whether to use c solely for relay, or solely for its own data. For conventional cooperation, it uses c jointly for both purposes in a time sharing manner. W.L.O.G., let $\alpha_j^c \in [0,1]$ denote its relay time sharing strategy. Then j's throughput for relay and its own transmission is as follows, respectively:

$$R_{j,P}^c = \frac{1 - \alpha_j^c}{2} \log\left(1 + 2 p_j^c g_{j,P}^c\right),$$
$$R_j^c = \frac{\alpha_j^c}{2} \log\left(1 + 2 p_j^c g_j^c\right), \forall j \in \mathcal{N_S}, c \in \mathcal{K}. \qquad (3)$$

Note that for $j = N + 1$, i.e. direct transmission for PU, obviously we have

$$R_{N+1,P}^c = R_{N+1}^c = 0, \forall j \in \mathcal{N_S}, c \in \mathcal{K}. \qquad (4)$$

In cases when other relaying and coding schemes are used, for instance amplify-and-forward or compress-and-forward with maximum ratio combining, we only need to change the throughput expressions (3), and our results in this paper are readily applicable. This is left as future work.

With conventional cooperation, $R_{i,j}^c = R_{j,P}^c$ holds for any c PU i leases to SU j. With FLEC, this does not have to hold for every leased subchannel. The only requirement is that SU j should deliver all data from the cooperating PUs, i.e. a total flow conservation requirement as follows:

$$\sum_{c \in \mathcal{K}} \sum_{i \in \mathcal{N_P}} R_{i,j}^c \leq \sum_{c \in \mathcal{K}} R_{j,P}^c, \forall j \in \mathcal{N_S}. \qquad (5)$$

2.2 Basics of Nash bargaining solutions

We present the salient concepts and results from Nash bargaining solutions in this section, which are used in the sequel. For details we refer readers to [10].

The basic setting is as follows: Let $\mathcal{N}$ be the set of players, including PUs and SUs. Let $\mathcal{S}$ be a closed and convex subset of $\mathcal{R}^N$ to represent the set of feasible payoff allocations that players can get if they all work together. Let $R_n^{\min}$ be the minimal payoff that the n-th player would expect; otherwise, he will not cooperate. Suppose $\{R_n \in \mathcal{S} | R_n \geq R_n^{\min}, \forall n \in \mathcal{N}\}$ is a nonempty bounded set. Define $\mathbf{R}^{\min} = (R_1^{\min}, \ldots, R_N^{\min})$, then the pair $(\mathcal{S}, \mathbf{R}^{\min})$ is called a N-person bargaining problem.

Within the feasible set $\mathcal{S}$, we first define the notion of Pareto optimality as a selection criterion in a typical game setting.

Definition 1: The point $(R_1, \ldots, R_N)$ is said to be *Pareto optimal* if and only if there is no other allocation R_n' such that $R_n' \geq R_n, \forall n \in \mathcal{N}$, and $R_n' > R_n, \exists n \in \mathcal{N}$, i.e. there exists no other allocation that leads to superior performance for some user without inferior performance for some other user.

The question that arises is: at which of infinitely many Pareto optimal points should we operate the system? A possible further criterion is the fairness of resource sharing. In this paper, we use the NBS fairness axioms from game theory.

Definition 2: $\bar{\mathbf{r}}$ is a **NBS**, i.e. $\bar{\mathbf{r}} = \phi(\mathcal{S}, \mathbf{R}^{\min})$, if the following axioms are satisfied [10]:

1) *Individual Rationality*: $\bar{R}_n \geq R_n^{\min}, \forall n \in \mathcal{N}$
2) *Feasibility*: $\bar{\mathbf{r}} \in \mathcal{S}$
3) *Pareto Optimality*
4) *Independence of Irrelevant Alternatives*: If $\bar{\mathbf{r}} \in \mathcal{S}' \subset \mathcal{S}$, $\bar{\mathbf{r}} = \phi(\mathcal{S}, \mathbf{R}^{\min})$, then $\bar{\mathbf{r}} = \phi(\mathcal{S}', \mathbf{R}^{\min})$

5) *Independence of Linear Transformations*: For any linear scale transformation ψ, $\psi(\phi(\mathcal{S}', \mathbf{R}^{\min})) = \phi(\psi(\mathcal{S}), \psi(\mathbf{R}^{\min}))$.
6) *Symmetry*: If $\mathcal{S}$ is invariant under all exchanges of players, then $\phi_i(\mathcal{S}, \mathbf{R}^{\min}) = \phi_{i'}(\mathcal{S}, \mathbf{R}^{\min}) \forall i, i'$.

Axioms 4-6 are called axioms of fairness. The irrelevant alternative axiom asserts that eliminating the feasible solutions that would not have been chosen should not affect the NBS solution. Axiom 5 asserts that the bargaining solution is scale invariant. The symmetry axiom asserts that if the feasible ranges for all players are completely symmetric, then all users have the same solution.

The following theorem shows that there is exactly one NBS that satisfies the above axioms.

Theorem 1: There is a unique solution function $\phi(\mathcal{S}, \mathbf{R}^{\min})$ that satisfies all axioms in *Definition 2* such that [10]

$$\phi(\mathcal{S}, \mathbf{R}^{\min}) \in \underset{\mathbf{R} \in \mathcal{S}, \mathbf{R} \succeq \mathbf{R}^{\min}}{\operatorname{argmax}} \prod_{n \in \mathcal{N}} \left(R_n - R_n^{\min} \right). \quad (6)$$

It has been proved that, when $R_n^{\min} = 0$ for all n, NBS fairness reduces to proportional fairness [18]. Note that in our problem, $R_n^{\min}$ for PUs will surely be non-zero since they get positive throughput if not cooperate, while that for SUs will be zero. Therefore NBS fairness here is different than proportional fairness. In general, the intuitive idea is that after the minimal requirements are met for all users, the rest of the resources are allocated *proportionally* to users according to their conditions.

2.3 An Optimization Framework Based on NBS

For our problem, we wish to consider *long-term* NBS fairness, which depends on the average throughput gain from cooperation over a relatively long period of time. For elastic traffic, long-term fairness not only faithfully reflects users' perceived performance, but also gives more flexibility to exploit time diversity of wireless channels. As discussed above, the cooperative game in an OFDMA cooperative cognitive radio networks can be formulated as follows.

Each user, being primary or secondary, has $\bar{R}_n$, the average total throughput summed across all subchannels, as its objective function. It is bounded above and has a non-empty, closed, and convex support. $\bar{\mathbf{R}}^{\min}$ is an N-dimensional vector that represents the minimal average performance requirements as in Sec. 2.2. For PUs, the minimal requirement will be the optimal average throughput they could obtain should they choose not to cooperate with SUs, given by a multi-user uplink scheduling algorithm [20]. For SUs, their minimal requirement that can be obtained without cooperation is clearly zero. $\mathbf{S}$ is the feasible set of resource allocation that satisfies $\bar{R}_n > \bar{R}_n^{\min}, \forall n$.

The problem, then, is to find the NBS, i.e., to solve the optimization problem (6) with $\bar{R}_n$ and $\bar{R}_n^{\min}$. The product terms in (6) make it difficult to solve. Mathematically, it is equivalent to solving the following:

$$\max_{\bar{\mathbf{R}} \in \mathcal{S}, \mathbf{R} \succeq \mathbf{R}^{\min}} \sum_{n \in \mathcal{N}} \ln \left(\bar{R}_n - \bar{R}_n^{\min} \right). \quad (7)$$

Notice that this is a long-term utility maximization problem whose optimum is achieved over a period of time. For the scheduling and resource allocation problem, it has to be solved in each scheduling epoch because channel conditions change over time. Therefore it is important to identify the instantaneous objective function we optimize in each epoch in order to arrive at long-term utility optimum. From the seminal paper of [21], it has been shown that maximizing the aggregate marginal utility $\sum U'(\bar{R}_n) \cdot R_n$ at each epoch exactly achieves long-term utility maximization. Therefore, separating the terms for PUs and SUs, the basic resource allocation framework for OFDMA cooperative cognitive radio networks at each epoch is:

$$\max_{\mathbf{R} \in \mathcal{S}, \mathbf{R} \succeq \mathbf{R}^{\min}} \sum_{i \in \mathcal{N}_{\mathcal{P}}} \frac{R_i - R_i^{\min}}{\bar{R}_i - \bar{R}_i^{\min}} + \sum_{j \in \mathcal{N}_{\mathcal{S}}} \frac{R_j}{\bar{R}_j}. \quad (8)$$

$R_i, \bar{R}_i, R_j, \bar{R}_j$ denote the instantaneous and average throughput for PU i and SU j at current epoch, respectively. Both $\bar{R}_i$ and $\bar{R}_j$ can be readily obtained by applying the exponential moving averaging technique. $R_i^{\min}, \bar{R}_i^{\min}$ are the instantaneous and average throughput requirement respectively, which can be obtained by running a multi-user scheduling algorithm at each epoch [20], and using exponential moving averaging technique.

Note that without considering long-term performance, the optimization must guarantee fairness in each epoch. However, when a time window is used, the fairness requirement is relaxed to the time window length. This provides more flexibility to improve the spectral efficiency, by making the current resource allocation related to previous ones. The term $\bar{R}_n - \bar{R}_n^{\min}$ in the denominator of (8) serves as a weight factor to adjust the priority of user n. If the user has an unfairly large throughput gain from cooperation from previous epochs, it may need to contribute more to others in the current epoch. Therefore the long-term fairness model encourages users to contribute more when channel conditions are better, and in turn gain more when it needs more help. In general it helps to achieve better system performance while enforcing the fairness notion over long run.

A final remark is that our optimization framework maximizes throughput gains without considering QoS requirements for both PUs and SUs for reasons of both tractability and conciseness. QoS requirements, such as minimum delay, bit error rate, etc., are usually specific to multimedia applications such as mobile video streaming, and is not addressed in this work that targets a general data transmission application. They can be incorporated as additional constraints into the optimization framework, and new algorithms can be developed as a possible direction of future work.

3 An Optimal Distributed Algorithm

3.1 Problem Formulation

We first consider a decentralized setting where the secondary network is independent from the primary network, and cannot be controlled by the primary BS. Thus, BS allocates resources to PUs *a priori* to any cooperative transmission, and SUs have to "negotiate" distributively with PUs in order to have cooperation taking place. In other words, cooperative transmission serves as an add-on component to the existing primary network, and is *opportunistically* harvested. This may correspond to the most immediate implementation scenario of CCRN that does not call for any change in the existing primary infrastructure, and therefore is of practical interest.

In this case, PU channel assignment is done separately by the BS, and is not part of the optimization. The resource allocation problem, including relay assignment, SU subchannel assignment, SU relay strategy optimization using FLEC, and PU-SU power control within the basic framework in Sec. 2.3 can be expressed succinctly as:

$$\begin{aligned} \max_{\mathbf{R},\mathbf{P},\boldsymbol{\alpha}} \quad & \sum_{i\in\mathcal{N_P}} \frac{R_i - R_i^{\min}}{\bar{R}_i - \bar{R}_i^{\min}} + \sum_{j\in\mathcal{N_S}} \frac{R_j}{\bar{R}_j} \\ \text{s.t.} \quad & \mathbf{0} \preceq \mathbf{P}\cdot\mathbf{1}^T \preceq \mathbf{p}^{\max}, \\ & \mathbf{R} \succeq \mathbf{R}^{\min}, \mathbf{R} \in \mathcal{C}(\mathbf{P},\boldsymbol{\alpha}), \end{aligned} \tag{9}$$

where $\mathbf{p}^{\max} = [p_1^{\max}, \dots, p_N^{\max}]^T$ is the power constraint vector, $\mathbf{P}$ is an $N \times K$ matrix such that P_n^c denotes the power expended by user n in subchannel c, $\boldsymbol{\alpha}$ is an $N_S \times K$ matrix such that α_j^c denotes the FLEC strategy of SU j on c, and $\mathcal{C}(\cdot)$ denotes the achievable rate region given $\mathbf{P}$ and α (Eq. (1)–(4)), with the flow conservation constraint at each SU (Eq. (5)). Since only one PU and one SU can be active on each subchannel, the column vector $\mathbf{P^c}$ has at most two non-zero entries, and it also specifies relay and subchannel assignments.

3.2 Dual Decomposition

The decentralized problem (9) is essentially a mixed integer program, with the objective function being neither convex nor concave. However, in an OFDMA system with many narrow subchannels, the optimal solution is always a convex function of $\mathbf{p}^{\max}$, because if two sets of throughputs using two different sets of $\mathbf{P}$ and $\boldsymbol{\alpha}$ are achievable individually, their linear combination is also achievable by a frequency-division multiplexing of the two sets of strategies. In particular, using the duality theory of [22], the following is true:

Proposition 1: The decentralized resource allocation problem (9) has zero duality gap in the limit as the number of OFDM subchannels goes to infinity, even though the discrete selection of subchannels, SUs and relay strategies are involved.

This proposition allows us to solve non-convex problems in their dual domain. Note that although the proposition requires the number of subchannels to go to infinity, in reality the duality gap is very close to zero as long as the number of subchannels is large [13].

Introduce Lagrangian multiplier vectors $\boldsymbol{\lambda}, \boldsymbol{\mu}, \boldsymbol{\nu}$ to the power, individual rationality, and flow conservation constraints. The Lagrangian becomes

$$\begin{aligned} L(\mathbf{R},\mathbf{P},\boldsymbol{\alpha},\boldsymbol{\lambda},\boldsymbol{\mu},\boldsymbol{\nu}) = & \sum_{i\in\mathcal{N_P}} \frac{R_i - R_i^{\min}}{\bar{R}_i - \bar{R}_i^{\min}} + \sum_{j\in\mathcal{N_S}} \frac{R_j}{\bar{R}_j} \\ & + \sum_{n\in\mathcal{N}} \lambda_n \left(p_n^{\max} - \sum_{c\in\mathcal{K}} p_n^c \right) + \sum_{i\in\mathcal{N_P}} \mu_i \left(R_i - R_i^{\min} \right) \\ & + \sum_{j\in\mathcal{N_S}} \nu_j \left(\sum_{c\in\mathcal{K}} R_{j,P}^c - \sum_{c\in\mathcal{K}} \sum_{i\in\mathcal{N_P}} R_{i,j}^c \right) \end{aligned} \tag{10}$$

The dual function becomes

$$g(\boldsymbol{\lambda},\boldsymbol{\mu},\boldsymbol{\nu}) = \begin{cases} \max\limits_{\mathbf{R},\mathbf{P},\boldsymbol{\alpha}} & L(\mathbf{R},\mathbf{P},\boldsymbol{\alpha},\boldsymbol{\lambda},\boldsymbol{\mu},\boldsymbol{\nu}) \\ \text{s.t.} & \text{Eq. (1)–(4)} \end{cases} \tag{11}$$

We know from convex optimization theory that as long as we can solve the maximization problem denoted by the dual function $g(\boldsymbol{\lambda},\boldsymbol{\mu},\boldsymbol{\nu})$, we can obtain the optimal solution of the dual problem by minimizing $g(\boldsymbol{\lambda},\boldsymbol{\mu},\boldsymbol{\nu})$ subject to the constraint that $\boldsymbol{\lambda},\boldsymbol{\mu}$ and $\boldsymbol{\nu}$ are non-negative. Thus we focus on solving the dual function in the following.

To solve $g(\boldsymbol{\lambda},\boldsymbol{\mu},\boldsymbol{\nu})$ with given $\boldsymbol{\lambda},\boldsymbol{\mu},\boldsymbol{\nu}$, it is equivalent to solving the same problem with the following objective:

$$\begin{aligned} \sum_{c\in\mathcal{K}} \Bigg(& \sum_{i\in\mathcal{N_P}} \sum_{j\in\mathcal{N_S^+}} \left(\frac{1}{\bar{R}_i - \bar{R}_i^{\min}} + \mu_i \right) R_{i,j}^c + \sum_{j\in\mathcal{N_S}} \frac{R_j^c}{\bar{R}_j} \\ & - \sum_{i\in\mathcal{N_P}} \lambda_i p_i^c - \sum_{j\in\mathcal{N_S}} \lambda_j p_j^c + \sum_{j\in\mathcal{N_S}} \nu_j \Big(R_{j,P}^c - \sum_{i\in\mathcal{N_P}} R_{i,j}^c \Big) \Bigg), \end{aligned}$$

where the term $\sum_n \lambda_n p_n^{\max} - \sum_i (\frac{1}{\bar{R}_i - \bar{R}_i^{\min}} + \mu_i) R_i^{\min}$ from the original objective is ignored for $\boldsymbol{\lambda}$ is given. Notice that in the first term of the objective, j could be $N+1$ which corresponds to the possibility of direct transmission.

Therefore, the problem can be decomposed into K per-subchannel problems. Recall that each subchannel is already assigned to a PU by the BS, the per-subchannel problem then reduces to finding the optimal helping SU, relay strategy, and resource allocation, and can be shown alternatively as follows:

$$\begin{aligned} \max_{j,p_i^c,p_j^c,\alpha_j^c} \quad & \left(\frac{1}{\bar{R}_i - \bar{R}_i^{\min}} + \mu_i \right) R_{i,j}^c + \frac{R_j^c}{\bar{R}_j} - \lambda_i p_i^c - \lambda_j p_j^c \\ & + \nu_j \left(R_{j,P}^c - R_{i,j}^c \right) \\ \text{s.t.} \quad & \text{Eq. (1)–(4)}, i = F(c), \alpha_j^c = \{0,1\}, \end{aligned} \tag{12}$$

where i is the primary user of subchannel c determined by the conventional multiuser scheduling denoted as $F(\cdot)\colon \mathcal{K} \to \mathcal{N_P}$.

3.3 Solving the Per-Subchannel Problem

The previous sections show that in a decentralized setting with per-user power constraint and per-SU total flow constraint, the resource allocation problem (9) can be solved optimally and efficiently in the dual domain. However, this hinges upon efficient solutions to the per-subchannel problem (12), which is required to solve the dual function $g(\boldsymbol{\lambda},\boldsymbol{\mu},\boldsymbol{\nu})$. In this section, we show the per-subchannel maximization problem can be solved efficiently via exhaustive search.

The main idea is to consider p_n^c as the optimizing variable and express $R_{i,j}^c, R_j^c, R_{j,P}^c$ in terms of p_i^c, p_j^c. The per-subchannel problem is essentially a joint optimization of transmission strategy, relay assignment, and relay strategy. For each subchannel c, its PU i needs to decide whether to use direct or cooperative transmission, which SU to cooperate with, while the chosen SU j needs to optimize its relay strategy denoted by the time sharing parameter $\alpha_j^c \in \{0,1\}$. Therefore, the exhaustive search is performed over a finite set defined by

- PU transmission strategies: {direct, cooperative}
- SU relay assignment: $j, j \in \mathcal{N_S}$

- SU relaying strategies: {primary data only ($\alpha_j^c = 0$), its own data only ($\alpha_j^c = 1$)}

We derive optimal solutions $\tilde{p}_i^c, \tilde{p}_j^c, \tilde{\alpha}_j^c$ under direct or cooperative transmission modes for any combination of subchannel c with its PU i and the SU j in the following.

3.3.1 Direct Transmission

If PU i chooses direct transmission, the problem becomes

$$\max_{p_i^c} \left(\frac{1}{(\bar{R}_i - \bar{R}_i^{\min})} + \mu_i \right) \log(1 + p_i^c g_i^c) - \lambda_i p_i^c \tag{13}$$

the solution of which is readily available by simple calculus:

$$\tilde{p}_i^c = \left[\frac{1}{\lambda_i \left(\bar{R}_i - \bar{R}_i^{\min}\right)} + \frac{\mu_i}{\lambda_i} - \frac{1}{g_i^c} \right]^+ . \tag{14}$$

3.3.2 Cooperative Transmission

Substituting the rate formulas (2)–(3) into (12) and regrouping the terms, the objective (12) becomes

$$\frac{\log(1 + 2p_i^c g_{i,j}^c)}{2(\bar{R}_i - \bar{R}_i^{\min})} + \frac{(\mu_i - \nu_j)\log(1 + 2p_i^c g_{i,j}^c)}{2} - \lambda_i p_i^c$$
$$+ \frac{\alpha_j^c \log(1 + 2p_j^c g_j^c)}{2\bar{R}_j} + \frac{\nu_j (1 - \alpha_j^c)\log(1 + 2p_j^c g_{j,P}^c)}{2} - \lambda_j p_j^c \tag{15}$$

The first three terms, denoted as $b_c(j, \lambda_i, \mu_i, \nu_j)$, represent PU i's benefit by having SU j as its relay, discounted by possible violation of flow conservation with price ν_j and power expenditure with price λ_i. $b_c(j, \lambda_i, \mu_i, \nu_j)$ can be easily optimized by i as only p_i^c is involved:

$$\tilde{p}_i^c = \frac{1}{2} \left[\frac{1}{\lambda_i(\bar{R}_i - \bar{R}_i^{\min})} + \frac{(\mu_i - \nu_j)}{\lambda_i} - \frac{1}{g_{i,j}^c} \right]^+ . \tag{16}$$

The last three terms, denoted as $b_j(c, \lambda_j, \nu_j)$, represent SU j's benefits from transmitting either its own or PU i's data on subchannel c, discounted by the power expenditure with price λ_j. Two optimizing variables α_j^c and p_j^c are involved here.

Maximization of $b_j(c, \lambda_j, \nu_j)$ can be done by setting α_j^c to 0 and 1, deriving the optimal p_j^c respectively as shown (17), and comparing the objective values. Ties can be broken arbitrarily.

$$\tilde{p}_j^c = \begin{cases} \frac{1}{2} \left[\frac{\nu_j}{\lambda_j} - \frac{1}{g_{j,P}^c} \right]^+, & \text{when } \tilde{\alpha}_j^c = 0, \\ \frac{1}{2} \left[\frac{1}{\lambda_j \bar{R}_j} - \frac{1}{g_j^c} \right]^+, & \text{when } \tilde{\alpha}_j^c = 1. \end{cases} \tag{17}$$

To summarize, the per-subchannel problem (12) can be efficiently solved via exhaustive search over a finite set defined by the transmission strategies, SUs, and SU relay strategies with FLEC as discussed above. The size of this discrete set is very limited, making it feasible for a practical network. The entire procedure can be summarized as follows:

Subroutine 1: Exhaustive search for solving (12) for a given subchannel c and its PU i:

- Every SU j maximizes $b_j(c, \lambda_j, \nu_j)$ using (17), and obtains $\tilde{p}_j^c, \tilde{\alpha}_j^c$. It then sends its optimal utility $\tilde{b}_j(c, \lambda_j, \nu_j)$, and ν_j, to its neighboring PUs.
- Every PU solves for $\tilde{p}_i^c$ using (14) for direct transmission.
- Every PU solves the joint utility maximization (15) distributively using (14), (16) and $\tilde{b}_j(c, \lambda_j, \nu_j)$ to get $\tilde{p}_i^c$ for cooperative transmission for each j. Then find the optimal $\tilde{j}$ that maximizes the joint utility.
- Choose the transmission mode with better joint utility. The corresponding optimal resource allocation $\tilde{j}, \tilde{p}_i^c, \tilde{p}_j^c, \tilde{\alpha}_j^c$ is then fixed.

Note that message exchange between PU and SUs are necessary here. Specifically, ν_j and the optimal value of SU's benefits $\tilde{b}_j(c, \lambda_j, \nu_j)$ needs to be passed to PU i.

3.4 An Optimal Distributed Algorithm

We have shown that the dual function can be decomposed into K per-subchannel problems, the optimal solutions of which can be obtained efficiently through exhaustive search. Then, the primal problem (9) can be optimally solved by minimizing the dual objective:

$$\begin{array}{ll} \min & g(\boldsymbol{\lambda}, \boldsymbol{\mu}, \boldsymbol{\nu}) \\ \text{s.t.} & \boldsymbol{\lambda}, \boldsymbol{\mu}, \boldsymbol{\nu} \succeq \mathbf{0}. \end{array} \tag{18}$$

Subgradient method can be used to solve this dual problem. The updating rules are as follows:

$$\lambda_n^{(l+1)} = \left[\lambda_n^{(l)} + \delta_n^{(l)} \left(\sum_{c \in \mathcal{K}} \tilde{p}_n^c - p_n^{\max} \right) \right]^+ , \tag{19}$$

$$\mu_i^{(l+1)} = \left[\mu_i^{(l)} + \epsilon_n^{(l)} \left(R_i^{\min} - R_i \right) \right]^+ , \tag{20}$$

$$\nu_j^{(l+1)} = \left[\nu_j^{(l)} + \kappa_j^{(l)} \left(\sum_{c \in \mathcal{K}} \sum_{i \in \mathcal{N}_{\mathcal{P}}} \tilde{R}_{i,j}^c - \sum_{c \in \mathcal{K}} \tilde{R}_{j,P}^c \right) \right]^+ \tag{21}$$

$\tilde{p}_n^c$ denotes the optimal power allocation for user $n, n \in \mathcal{N}$. Following a diminishing step size rule for choosing $\boldsymbol{\delta^{(l)}}, \boldsymbol{\epsilon^{(l)}}, \boldsymbol{\kappa^{(l)}}$, the subgradient method above is guaranteed to converge to the optimal dual variables [12]. The optimal primal variables can then be easily found.

Observe that, because of the dual decomposition, dual optimization by subgradient method can be done in a *distributed* fashion. First, in each iteration, the per-subchannel problems (12) can be solved simultaneously by the PU of the subchannel exchanging information with neighboring SUs as in *Subroutine* 1, though the objective jointly involves PU's and SU's benefits.

Second, subgradient updates can also be distributively performed by each primary and secondary users. The algorithm can be perceived as an iterative bargaining process. The dual variable ν_j is exchanged between PUs and SUs and serves as a relay price signal to coordinate the level of cooperation. When the relay traffic demand $\sum_c \sum_i \tilde{R}_{i,j}^c$ from PUs exceeds the supply $\sum_c \tilde{R}_{j,P}^c$ from j, i.e. PUs over-exploit j, j increases its relay price ν_j for the next round of bargaining to suppress the excessive demand, as shown in (21). Similarly, if j has redundant relay capacity $\sum_c \tilde{R}_{j,P}^c > \sum_c \sum_i \tilde{R}_{i,j}^c$, it will decrease the relay price ν_j to attract more relay traffic and therefore obtain more channels to use. The process continues until it converges to the optimal resource allocation.

The interpretation of other dual variables λ_n and μ_i is also worth mentioning. For each user, λ_n is easily understood as a price signal to regulate its power consumption. μ_i for each PU is used to ensure that the resource allocation is individual rational, i.e. it is beneficial for each PU in that the total throughput obtained from cooperation R_i is larger than $R_i^{\min}$. When $R_i < R_i^{\min}$, μ_i will be increased as in (20), and so will $\tilde{p}_i^c$ as in (16). Therefore, R_i^c will be larger in the next iteration. Both dual variables are kept privately and updated independently with only local information.

Algorithm 1 *Optimal Distributed Bargaining for FLEC*

1. The primary BS runs a multiuser scheduling algorithm to determine $R_i^{\min}$ for PUs without cooperation.
2. Each primary user initializes $\lambda_i^{(0)}, \mu_i^{(0)}$. Each secondary user initializes both power and relay prices $\lambda_j^{(0)}, \nu_j^{(0)}$.
3. Given $\boldsymbol{\lambda^{(l)}}, \boldsymbol{\mu^{(l)}}, \boldsymbol{\nu^{(l)}}$, each PU i coordinates with each neighboring SU j concurrently to solve the per-subchannel resource allocation problem (12) using *Subroutine 1*.
4. SU j bargains by performing a subgradient update for the relay price $\nu_j^{(l)}$ as in (21). PU i updates $\mu_i^{(0)}$ as in (20). Each user also updates the power price $\lambda_n^{(l)}$ as in (19).
5. Return to step 3 until convergence.
6. Every user updates $\bar{R}_n$ from its total throughput R_n in this epoch. Every PU i updates $\bar{R}_i^{\min}$ from $R_i^{\min}$ in Step 1. They will be used for resource allocation in next epoch.

The complete bargaining algorithm is shown in Algorithm 1. Since it optimally solves the dual problem (18), it optimally solves the primal problem (9) according to Proposition 1.

Theorem 2: The distributed bargaining algorithm as shown in Algorithm 1 always converges, and when it converges its solution *optimally* solves the decentralized resource allocation problem (9).

We now analyze the amount of message exchanges and complexity here. For a pair of PU-SU, two messages $\nu_j, \tilde{b}_j(i, \lambda_j, \nu_j)$ need to be exchanged for each c. They can easily be piggybacked in the probing packets from SU to PU to measure the channel gain, resulting in zero message exchange overhead. The complexity of solving K per-subchannel problems by exhaustive search is $O(KN_S)$. The complexity of the subgradient update is polynomial in the dimension of the dual problem, which is K. Therefore, the complete algorithm has complexity polynomial in KN_S. While this may render it infeasible for real-time scheduling within 5–10 ms when the network scales, the distributed nature of the algorithm makes it possible for each PU to *concurrently* solve the per-subchannel problem, reducing the complexity to only $O(N_S)$. Also, each user can update their own prices as dual variables independently. Further, in reality, only a few SUs residing in the neighborhood of the PU can potentially help and thus have to be considered. Therefore from the network point of view, each round of bargaining has complexity $O(1)$.

Careful readers may be concerned with the slow convergence of the subgradient updates, especially when the problem scales up. We comment that since each PU only needs to bargain with neighboring SUs, the convergence complexity is limited by the size of the neighborhood and does not scale up with the problem size. Also, only scalar dual variables need to be updated for each user. We observe in simulations in Sec. 6.3 that the algorithm converges within about 20 iterations in most cases.

4 A Centralized Heuristic Algorithm

We now proceed to the centralized setting. Recall that in the decentralized setting, the subchannel assignment to PUs is done by the BS without considering the possibility of cooperative transmission, and thus is not part of the optimization. This enables efficient development of distributed algorithms, but is sub-optimal in general. Here we consider the scenario where the SU cooperative transmission becomes an integral part of primary BS scheduling, and SUs abide by the scheduling decisions, provided that the resource allocation is fair as reflected by the NBS fairness. With centralized FLEC, we have an additional dimension to optimize: *global subchannel assignment* for both PU and SU.

4.1 Motivation for Developing Heuristics

The problem can be formulated in a similar way as the decentralized problem (9), and optimally solved via dual decomposition and subgradient update. However, it is computationally prohibitive to do so. Since the BS can now assign any subchannel used by any helped PU to any helping SU, at each iteration, the per-subchannel problem now becomes:

$$\begin{aligned}\max_{i,j,i',j',p_i^c,p_{j'}^c,\alpha_{j'}^c} \quad & \left(\frac{1}{\bar{R}_i - \bar{R}_i^{\min}} + \mu_i\right) R_{i,j}^c + \frac{R_{j'}^c}{\bar{R}_{j'}} - \lambda_i p_i^c \\ & -\lambda_{j'} p_{j'}^c + \nu_{j'}\left(R_{j',P}^c - R_{i',j'}^c\right) \\ \text{s.t.} \quad & \text{Eq. (1)–(4)}, \alpha_{j'}^c = \{0,1\}\end{aligned} \tag{22}$$

Compared to the decentralized version (12), there are additional variables i, i', j' to optimize, which represents the global subchannel assignment. Specifically, i is the PU assigned to use c and j is its helping SU, while j' is the SU assigned to use c and i' is the PU whose data is relayed by j'. Note that i (j) needs not to be equal to i' (j'). The solution of this problem thus has to exhaustively search all possible combinations of PU-SU pairs for each subchannel, which has a complexity of $O(KN_P^2N_S^2)$ since distributed concurrent optimization is not possible.

Moreover, because of the global impact of centralized subchannel assignment, the speed of convergence of dual variables $\boldsymbol{\lambda}, \boldsymbol{\mu}, \boldsymbol{\nu}$ scales up with the size of the dual problem which scales quadratically with N_P and N_S, instead of being independent of the dual problem size as in the decentralized case. Note that although the convergence of subgradient method is guaranteed, the speed of convergence is not, and often depends heavily on problem conditioning and scaling [12]. From our computational experiences, the convergence of subgradient updates is too slow to be useful for practical use as will be shown in Sec. 6.3.

Given that complexity has been significantly increased, we focus on developing efficient heuristics in this section, which reduce the complexity while maintaining good performance.

Nevertheless, the slow subgradient based centralized algorithm, called *Centralized Optimization for FLEC* hereafter, is used to derive the optimal performance as a benchmark as in Sec. 6.4.

4.2 Overview of the Heuristic Algorithm

To make the problem more tractable, we decouple it to three orthogonal dimensions: relay assignment, subchannel assignment, and power control. First, we derive optimal relay assignment using bipartite matching, assuming that each SU is only able to help one distinct PU and one PU can only be matched to one SU. This simplification is reasonable as it ensures a certain level of fairness. Then we assume that power is equally distributed, and derive an subchannel assignment algorithm. Even with optimal relay and equal power assignment, this turns out to be an NP-hard problem. We propose a suboptimal algorithm based on randomized rounding and prove its approximation ratio. Finally, power allocation is solved to maximize performance with the given subchannel assignment. Be reminded that as an initialization step, the BS first performs a multi-user scheduling [20] to determine $R_i^{\min}, \bar{R}_i^{\min}$ for PUs before the three component algorithms run. The entire heuristic algorithm is called *Centralized Heuristic for FLEC* hereafter.

We do not claim that our heuristic design is the only choice here. In fact other heuristic designs are entirely possible. For example, one may choose to solve the subchannel assignment first, then relay assignment, and finally power control. It is also possible to jointly solve two of the three orthogonal dimensions. For example one may choose to solve the joint problem of relay and subchannel assignment and then compute the power allocation based on the solution of the joint problem. These possibilities are beyond the scope of this paper and left as future research, since they have different formulations and require different solutions. We do not claim that our heuristic design is the best, although simulation studies in Sec. 6 show that it improves performance significantly compared to the conventional identical channel cooperation.

4.3 Relay Assignment

Here, we model each user n as having an *imaginary* channel with a normalized channel gain to noise ratio $\bar{g}_n^c = \frac{1}{K}\sum_c g_n^c$ and power $p_n^{\max}$. Then the optimal FLEC strategy reduces to simple time-sharing on this channel. Assuming each SU can only help one distinct PU and one PU can only be matched to one SU, the optimal relay assignment under the basic framework in Sec. 2.3 can be determined by:

$$\max_{x_{i,j}\in\{0,1\}} \sum_{i\in\mathcal{N}_\mathcal{P}} \sum_{j\in\mathcal{N}_\mathcal{S}^+} x_{i,j}\left(\frac{\hat{R}_{i,j}-R_i^{\min}}{\bar{R}_i-\bar{R}_i^{\min}}+\frac{\hat{R}_{j,i}}{\bar{R}_j}\right)$$

$$\text{s.t.}\quad \hat{R}_{i,j}=\frac{1}{2}\min\left(\log(1+2p_i^{\max}\bar{g}_{i,j}),\log(1+2p_j^{\max}\bar{g}_{j,P})\right),$$

$$\hat{R}_{j,i}=\frac{1}{2}\log(1+2p_j^{\max}\bar{g}_j)\left[1-\frac{\log(1+2p_i^{\max}\bar{g}_{i,j})}{\log(1+2p_j^{\max}\bar{g}_{j,P})}\right]^+,$$

$$\forall i\in\mathcal{N}_\mathcal{P}, j\in\mathcal{N}_\mathcal{S},$$

$$\hat{R}_{i,N+1}=\log(1+p_i^{\max}\bar{g}_i),\hat{R}_{N+1,i}=0,\bar{R}_{N+1}=\infty.$$

$$\sum_{i\in\mathcal{N}_\mathcal{P}} x_{i,j}=1,\forall j\in\mathcal{N}_\mathcal{S},\ \sum_{j\in\mathcal{N}_\mathcal{S}} x_{i,j}=1,\forall i\in\mathcal{N}_\mathcal{P}$$

$x_{i,j}$ is the binary variable denoting the relay assignment of SU j to PU i. Thus, for PU i, its overall cooperative throughput $R_{i,j}$ when matched to SU j is the minimum of the two hops PU-SU and SU-BS. When i chooses $j=N+1$, i.e. direct transmission, the throughput is calculated from the Shannon formula. For SU j, the overall cooperative throughput $R_{j,i}$ when matched to PU i is implied from the time-sharing strategy, since it must relay all primary traffic whenever possible.

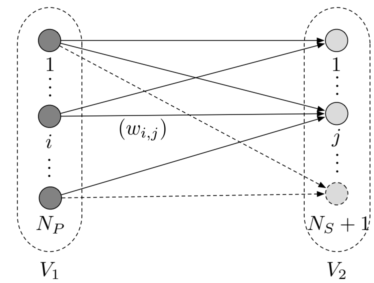


Fig. 2. Weighted bipartite matching for optimal relay assignment.

The above relay assignment is a weighted bipartite matching problem that can be optimally solved. To see this, construct a graph $A=(V_1\times V_2,E)$ where V_1 and V_2 correspond to the set of PUs and SUs respectively as shown in Fig. 2. We patch a void vertex to V_2 to incorporate the direct transmission. The edge set E corresponds to $N_P(N_S+1)$ edges connecting all possible pairs of users in the two vertex sets. Each edge (i,j) carries a weight, $w_{i,j}$, where

$$w_{i,j}=\frac{\hat{R}_{i,j}-R_i^{\min}}{\bar{R}_i-\bar{R}_i^{\min}}+\frac{\hat{R}_{j,i}}{\bar{R}_j}.$$

For edges connecting PUs to the void SU that we patched, the edge weights have captured the maximum marginal utility given by direct transmission. Observe that A is bipartite, optimal relay assignment is then equivalent to finding maximum weighted bipartite matching on A. The Hungarian algorithm is a popular polynomial-time algorithm to solve it optimally [23].

4.4 Subchannel Assignment

For PUs using direct transmission as determined by optimal relay assignment, they do not share resources with SUs, and as such cannot benefit from SU cooperation. Therefore they use the same subchannels as allocated in the initialization step. For the set of PUs $\mathcal{N}_\mathcal{P}^\mathcal{R}$ that use cooperative transmission, the set of their allocated subchannels $\mathcal{K}^\mathcal{R}$ in the initialization step will be collected and re-assigned by the following algorithm. For each PU i and its unique helping SU $j(i)$, we assume they will use power $\bar{p}_i=\frac{p_i^{\max}}{K_i}, \bar{p}_{j(i)}=\frac{p_{j(i)}^{\max}}{K_i}$ respectively on each subchannel, where K_i is the number of subchannels allocated to i in the initialization step [20]. Such an equal power assumption is widely used and leads to subchannel assignment algorithms with near-optimal performance, as reported extensively [20], [24] and will be shown in Sec. 6.4.

The subchannel assignment problem can be formulated as in (23): where $w_{i,j(i)}^{c_1}$ denotes the marginal utility (normalized to a large value $w_{\max}$) obtained by PU i on being assigned c_1

on PU-SU link in the first time slot (i.e. $w^{c_1}_{i,j(i)} = 0.5\log(1+2\bar{p}_i g^{c_1}_{i,j})/(\bar{R}_i - \bar{R}^{\min}_i)$), and $w^{c_2}_{j(i),P}$ denotes the marginal utility of assigning c_2 for $j(i)$ on SU-BS link in the second slot (i.e. $w^{c_2}_{j(i),P} = 0.5\log(1+2\bar{p}_{j(i)} g^{c_2}_{j(i),P})/(\bar{R}_i - \bar{R}^{\min}_i)$. $w^{c_2}_j$ denotes the normalized marginal utility of SU j on being assigned c_2 for its own data $0.5\log(1+2\bar{p}_j g^{c_2}_j)/\bar{R}_j$, and a_i, b_j denote the aggregate marginal utility (flow) achieved by PU i and SU j respectively. $x^{c_1}_i$ is the binary variable denoting whether c_1 is assigned to PU i in the first time slot, $y^{c_2}_i$ denotes whether c_2 is assigned to i's helper SU $j(i)$ for relaying in the second time slot, and $y^{c_2}_j$ denotes whether c_2 is assigned to SU j for its own transmission in the second time slot.

$$\max_{x^{c_1}_i, y^{c_2}_i, y^{c_2}_j} \sum_{i\in\mathcal{N}^{\mathcal{R}}_{\mathcal{P}}} a_i + \sum_{j\in\mathcal{N}_{\mathcal{S}}} b_j \tag{23}$$
$$\text{s.t.} \sum_{c_2\in\mathcal{K}^{\mathcal{R}}} y^{c_2}_j \cdot w^{c_2}_j = b_j, \forall j \in \mathcal{N}_{\mathcal{S}},$$
$$\sum_{c_1\in\mathcal{K}^{\mathcal{R}}} x^{c_1}_i \cdot w^{c_1}_{i,j(i)} = a_i, \sum_{c_2\in\mathcal{K}^{\mathcal{R}}} y^{c_2}_i \cdot w^{c_2}_{j(i),P} = a_i, \forall i \in \mathcal{N}^{\mathcal{R}}_{\mathcal{P}},$$
$$\sum_{i\in\mathcal{N}^{\mathcal{R}}_{\mathcal{P}}} x^{c_1}_i = 1, \sum_{i\in\mathcal{N}^{\mathcal{R}}_{\mathcal{P}}} y^{c_2}_i + \sum_{j\in\mathcal{N}_{\mathcal{S}}} y^{c_2}_j = 1, \forall c_1, c_2 \in \mathcal{K}^{\mathcal{R}},$$

Theorem 3: The subchannel assignment problem under the above IP formulation is NP-hard.

Proof: The problem can be reduced from type-dependent multiple knapsack problems (MKP), where each set of knapsacks (users) belongs to a different type (time slot and primary/secondary). The profit of allocating an item (subchannel) depends not only on the knapsacks but also the type of them. The one-type MKP is known to be NP-hard and even hard to approximate [25]. Therefore our problem is NP-hard. □

Given the hardness of the problem, we present a rounding based algorithm to solve it as shown in Algorithm 2. It ensures that each subchannel is assigned to at most one user for both slots. We now capture the performance of the algorithm.

Theorem 4: Algorithm 2 provides an approximation ratio of $1 - \sqrt{\frac{4cN_S}{K^R}\ln(K^R)}$ with high probability, where K^R is the cardinality of the subchannel set $\mathcal{K}^{\mathcal{R}}$.

Proof: Refer to the Appendix in the supplementary materials for a detailed proof. □

Therefore, its performance becomes better when there is a larger magnitude of available subchannels to users in the system. Since the number of subchannels in a practical OFDMA system is much bigger than that of users, Algorithm 2 can be expected to provide good performance.

4.5 Power Control

After all the subchannels are allocated as above, power can be allocated to each user optimally. For PUs with direct transmission, optimal power allocation is a simple water-filling solution. For PUs with cooperative transmission, optimal power allocation is performed on a per-pair basis with their unique helping SUs. With subchannels allocated and their use on an SU determined, power allocation on each pair of PU-SU is a standard convex optimization problem and can be readily solved by KKT conditions. We omit the details here.

Algorithm 2 *Rounding-based Subchannel Assignment*

1. Formulate the problem using the IP above. Solve its LP relaxation with $x^{c_1}_i, y^{c_2}_i, y^{c_2}_j$ being relaxed to $[0,1]$. Let the LP solutions be $\hat{x}^{c_1}_i, \hat{y}^{c_2}_i, \hat{y}^{c_2}_j$ and $\hat{a}_i, \hat{b}_j$.
2. Adopt the following procedure to round the fractional solutions, $\hat{x}^{c_1}_i, \hat{y}^{c_2}_n$, to integral values, $\tilde{x}^{c_1}_i, \tilde{y}^{c_2}_n$, where $n \in \{i \in \mathcal{N}^{\mathcal{R}}_{\mathcal{P}}\} \cup \{j \in \mathcal{N}_{\mathcal{S}}\}$.
   - For every c_2, round $y^{c_2}_n$ to 1 ($\tilde{y}^{c_2}_n$) with probability $\hat{y}^{c_2}_n$. If $\tilde{n}$ is the user to whom c_2 is assigned, then $\hat{y}^{c_2}_n = 0, \forall n \neq \tilde{n}$.
   - Update $\tilde{a}_i = \frac{\sum \tilde{y}^{c_2}_i w^{c_2}_{j(i),P}}{1-\delta}, \tilde{b_j} = \frac{\sum \tilde{y}^{c_2}_j w^{c_2}_j}{1-\delta}$, where δ is a constant derived in the Appendix. Run the LP again on $x^{c_1}_i$ only. Let $\bar{x}^{c_1}_i$ be the solutions of the new LP.
   - For c_1, round $x^{c_1}_i$ to 1 ($\tilde{x}^{c_1}_i$) with probability $\bar{x}^{c_1}_i$. If $\tilde{i}$ is the PU c_1 is assigned to, then $\tilde{x}^{c_1}_i = 0, \forall i \neq \tilde{i}$.

5 Identical Channel Cooperation

In previous sections, we have addressed the resource allocation problem with FLEC in both decentralized and centralized settings. In this section, we present solutions for resource allocation with conventional identical channel cooperation (CC), which makes our analysis complete. The motivation to study CC here is that it can serve as the performance benchmark for our flexible channel cooperative scheme. Also, due to implementation and complexity considerations, FLEC may not be feasible in certain scenarios, whereas CC is comparatively easier to implement due to its simplicity. Similar to FLEC, we also consider both decentralized and centralized CC.

5.1 Decentralized CC

5.1.1 Problem Formulation

Scheduling and resource allocation of decentralized CC can be similarly formulated as that of FLEC in Sec. 3.1. The key difference is that, the per-subchannel flow conservation constraints need to be satisfied for each subchannel, instead of only total flow conservation (5) for FLEC. Mathematically,

$$\sum_{i\in\mathcal{N}_{\mathcal{P}}} R^c_{i,j} \leq R^c_{j,P}, \forall j \in \mathcal{N}_{\mathcal{S}}, c \in \mathcal{K}. \tag{24}$$

In addition, we have the following for the time sharing strategy:

$$\alpha^c_j \in [0,1), \forall c \in \mathcal{K}, j \in \mathcal{N}_{\mathcal{S}} \tag{25}$$

instead of $\alpha^c_j = \{0,1\}$ for FLEC. $\alpha^c_j \neq 1$ since j must relay in CC. The problem can then be presented as:

$$\max_{\mathbf{R},\mathbf{P},\boldsymbol{\alpha}} \sum_{i\in\mathcal{N}_{\mathcal{P}}} \frac{R_i - R^{\min}_i}{\bar{R}_i - \bar{R}^{\min}_i} + \sum_{j\in\mathcal{N}_{\mathcal{S}}} \frac{R_j}{\bar{R}_j}$$
$$\text{s.t. } \mathbf{P}\cdot\mathbf{1}^T \preceq \mathbf{p}^{\max}, \mathbf{R} \succeq \mathbf{R}^{\min}, (1)-(4), (24)-(25). \tag{26}$$

From an intuition level, CC has more flexible time sharing strategy, but requires the relay transmission to be on the identical subchannel. FLEC is more flexible in terms of channel sharing strategy for cooperative transmission, but the time sharing strategy is restricted. From an optimization point of view, CC has $N_S K$ per-subchannel flow conservation constraints, while FLEC only has N_S total flow conservation constraints, where N_S is the number of SUs and K the number of subchannels as in Sec. 2.1. Given that K is typically on the order of hundreds in a practical OFDMA system, CC formulation has far more constraints than FLEC. Because these flow conservation constraints directly impact SU's throughput for relay and its own transmission as shown in (3), they are active constraints that directly impact the optimization objective. Thus, it is not a surprise that FLEC outperforms CC in both decentralized and centralized settings, as we will show in Sec. 6.

In general, (26) can also be solved in the dual domain by taking advantage of convexity through frequency sharing in OFDMA networks. The same set of techniques, including dual decomposition, exhaustive search for the per-subchannel problems, and subgradient method to update the dual variables, can be applied in the same way as in Sec. 3.2. However, due to the per-subchannel flow constraint, a different dual decomposition technique is used to efficiently achieve the dual optimum via the subgradient method, making the analysis different. We show the differences in details in the following.

5.1.2 Dual Decompsition

Recall in the decentralized FLEC problem (9), we relax power, flow, and individual rationality constraints, so they can be decoupled into per-subchannel constraints, and the dual variable updates can be understood as coordinating these constraints such that, when combined together, they are satisfied at the end of the process. In the decentralized resource allocation problem of CC (26), the flow constraint is already in the decoupled form to be satisfied for each subchannel. Thus, we only need to relax the total power and individual rationality constraints.

As discussed, the Lagrangian can be written as:

$$L(\mathbf{R},\mathbf{P},\boldsymbol{\alpha},\boldsymbol{\lambda},\boldsymbol{\mu}) = \sum_{i\in\mathcal{N_P}} \frac{R_i - R_i^{\min}}{\bar{R}_i - \bar{R}_i^{\min}} + \sum_{j\in\mathcal{N_S}} \frac{R_j}{\bar{R}_j} + \sum_{n\in\mathcal{N}} \lambda_n \left(p_n^{\max} - \sum_{c\in\mathcal{K}} p_n^c\right) + \sum_{i\in\mathcal{N_P}} \mu_i \left(R_i - R_i^{\min}\right). \quad (27)$$

The dual function is

$$g(\boldsymbol{\lambda},\boldsymbol{\mu}) = \begin{cases} \max\limits_{\mathbf{R},\mathbf{P},\boldsymbol{\alpha}} & L(\mathbf{R},\mathbf{P},\boldsymbol{\alpha},\boldsymbol{\lambda},\boldsymbol{\mu}) \\ \text{s.t.} & (1)-(4),(24)-(25), \end{cases} \quad (28)$$

By expanding the term R_i, R_j, ignoring constant terms, and realizing that each subchannel is already assigned to a PU, the per-subchannel problem can be written as

$$\begin{aligned} \max_{j,p_i^c,p_j^c,\alpha_j^c} \quad & \frac{R_{i,j}^c}{\bar{R}_i - \bar{R}_i^{\min}} + \mu_i R_{i,j}^c + \frac{R_j^c}{\bar{R}_j} - \lambda_i p_i^c - \lambda_j p_j^c \\ \text{s.t.} \quad & (1)-(4),(25), R_{i,j}^c \le R_{j,P}^c, i = F(c) \end{aligned} \quad (29)$$

where i is the primary user of subchannel c determined by the conventional multiuser scheduling denoted as $F(\cdot)\colon \mathcal{K} \to \mathcal{N_P}$.

5.1.3 Solutions to the Per-subchannel Problem

Exhaustive search can also be used to solve the per-subchannel problem, as in Sec. 3.3. As we have seen, to enable such search we need to derive optimal solutions $\tilde{p}_i^c, \tilde{p}_j^c, \tilde{\alpha}_j^c$ under direct and cooperative transmission modes for any combination of subchannel c with its PU i and the SU j. Readily we can see that for direct transmission, the optimal solution $\tilde{p}_i^c$ is the same as in (14). However, for cooperative transmission, the derivations are different from the previous analysis.

The first observation is that, maximization of the problem is achieved with the inequality of the flow conservation constraint achieved as equality. This can be easily verified by observing that increasing $R_{j,P}^c$ any further beyond $R_{i,j}^c$ will not increase the utility of PU i. On the other hand, it will decrease the utility of SU j in the objective function, since j will inevitably have less resources to improve its own throughput.

With this observation, and by substituting the rate formulas (2)–(3), we need to solve the following:

$$\begin{aligned} \max_{p_i^c,p_j^c,\alpha_j^c} \quad & \left(\frac{1}{2(\bar{R}_i - \bar{R}_i^{\min})} + \mu_i\right) \log(1 + 2p_i^c g_{i,j}^c) - \lambda_i p_i^c \\ & + \frac{\alpha_j^c \log(1 + 2p_j^c g_j^c)}{2\bar{R}_j} - \lambda_j p_j^c \\ \text{s.t.} \quad & \frac{\log(1 + 2p_i^c g_{i,j}^c)}{2} = \frac{(1-\alpha_j^c)\log(1 + 2p_j^c g_{j,P}^c)}{2}, \\ & \alpha_j^c \in [0,1). \end{aligned} \quad (30)$$

Essentially, this is a constrained non-linear maximization with respect to two variables with standard solution methods. But it turns out quite difficult to obtain a closed form solution. We resort to numerical methods to obtain solutions efficiently.

The entire procedure to solve the per-suchannel problem for CC thus can be summarized as follows:

Subroutine 2: Exhaustive search for solving (29) for a given subchannel c and its PU i:

- Solve for $\tilde{p}_i^c$ using (14) for direct transmission.
- Contact each neighboring SU j to obtain λ_j, g_j^c and $\bar{R}_j$. Solve the joint utility maximization problem (30) numerically to get optimal $\tilde{p}_i^c, \tilde{p}_j^c, \tilde{\alpha}_j^c$ for each j. Then find the optimal $\tilde{j}$ that maximizes the joint utility.
- Choose the transmission mode with larger utility. Output the corresponding optimal resource allocation $\tilde{j}, \tilde{p}_i^c, \tilde{p}_j^c, \tilde{\alpha}_j^c$.

The complete resource allocation for CC, denoted as Distributed Bargaining for CC, is shown in Algorithm. 3.

Algorithm 3 *Distributed Bargaining for CC*

1. The primary BS runs a multiuser scheduling algorithm to determine $R_i^{\min}$ for PUs without cooperation.
2. Each user initializes its power price $\lambda_n^{(0)}$. Each PU initializes the dual variable $\mu_i^{(0)}$.
3. Given $\boldsymbol{\lambda}^{(l)}$, each PU i solves the per-subchannel resource allocation problem (29) using *Subroutine* 2.
4. Each user n bargains by performing a subgradient update for the price λ_n as in (19). Each PU i also updates μ_i as in (20).
5. Return to step 3 until convergence.
6. Every user updates $\bar{R}_n$ from its total throughput R_n in this epoch. Every PU i updates $\bar{R}_i^{\min}$ from $R_i^{\min}$ in Step 1. They will be used for resource allocation in next epoch.

5.2 Centralized CC

Finally we consider resource allocation of centralized CC, which takes into account subchannel assignment to PUs and SUs. By the same argument as in Sec. 4, our focus is on developing efficient heuristics with short running time. We follow the same approach in developing Centralized Heuristics for FLEC and divide the problem into three dimensions, i.e. relay assignment, subchannel assignment, and power control. Readily we can see that the same relay assignment algorithm based on maximum weighted bipartite matching can be used here, since we would have an exactly the same problem formulation with only total flow conservation constraints, when all the channels are combined to form an imaginary channel as in Sec. 4.3. It is also straightforward that optimal power allocation follows the famous water-filling solution, given the relay and subchannel assignment. The only difference then lies in solving the subchannel assignment, which turns out to be much easier. The entire algorithm is referred to as Centralized Heuristics for CC thereafter.

5.2.1 *Subchannel Assignment*

As in Sec. 4.4, we only consider the set of PUs $\mathcal{N}_{\mathcal{P}}^{\mathcal{R}}$ that are assigned with an unique helping SU each. Their allocated subchannels $\mathcal{K}^{\mathcal{R}}$ in the initialization step is re-assigned by the channel assignment algorithm. The same assumptions are inherited, that each PU i and its unique helping SU $j(i)$ use equal power $\bar{p}_i = \frac{p_i^{\max}}{K_i}, \bar{p}_{j(i)} = \frac{p_{j(i)}^{\max}}{K_i}$ respectively on each subchannel, where K_i is the number of subchannels allocated to i in the initialization step.

From the per-subchannel flow conservation constraint (24), optimal time sharing $\bar{\alpha}_{j(i)}^c$ can be uniquely determined under equal power allocation $\bar{p}_i, \bar{p}_{j(i)}$ on each subchannel. Specifically, from (24),

$$\bar{\alpha}_{j(i)}^c = 1 - \frac{\log\left(1 + 2\bar{p}_i g_{i,j(i)}^c\right)}{\log\left(1 + 2\bar{p}_{j(i)} g_{j(i),P}^c\right)}, \forall c \in \mathcal{K}^{\mathcal{R}} \tag{31}$$

when c is allocated to $i, j(i)$. Thus the corresponding optimal utility $u_{i,j(i)}^c$ is simply

$$u_{i,j(i)}^c = \frac{\log\left(1 + 2\bar{p}_i g_{i,j(i)}^c\right)}{2(\bar{R}_i - \bar{R}_i^{\min})} + \frac{\bar{\alpha}_{j(i)}^c \log\left(1 + 2\bar{p}_{j(i)} g_{j(i)}^c\right)}{2\bar{R}_{j(i)}} \tag{32}$$

The subchannel assignment problem can be casted as:

$$\max_{x_i^c} \quad \sum_{c \in \mathcal{K}^{\mathcal{R}}} \sum_{i \in \mathcal{N}_{\mathcal{P}}^{\mathcal{R}}} x_i^c u_{i,j(i)}^c$$
$$\text{s.t.} \quad \sum_{i \in \mathcal{N}_{\mathcal{P}}^{\mathcal{R}}} x_i^c = 1, \forall c \in \mathcal{K}^{\mathcal{R}}$$

The constraint is such that each subchannel is only allocated to one pair of PU-SU. This can be easily solved by assigning each subchannel c to a PU i that has the largest $u_{i,j(i)}^c$. That is, $\tilde{i}^c = \text{argmax}_{i \in \mathcal{N}_{\mathcal{P}}^{\mathcal{R}}} u_{i,j(i)}^c$.

5.3 Discussions

Through the analysis in this section, we can see that our NBS resource allocation framework is readily applicable to identical channel cooperation. In general, CC makes the problem easier to solve compared to that with FLEC. The reason is the straightforward time sharing cooperation strategy that simplifies the scenario and reduces the degrees of optimization freedom. However, optimality is sacrificed simply because time sharing accounts for only a subset of all possible cooperation strategies under FLEC, as we have already seen from the analysis and will be verified in the simulation studies in the next secion.

6 Performance Evaluation

To evaluate the performance of FLEC with the proposed resource allocation algorithms, we adopt empirical parameters to model the fading environment. There are 128 subchannels centered at 2.5 GHz, each with 312.5 kHz bandwidth. Channel gain can be decomposed into a large-scale log normal shadowing component with standard deviation of 5.8 and path loss exponent of 4, and a small-scale Rayleigh fading component. The inherent frequency selectivity is captured by an exponential power delay profile with delay spread 1.257 νs as reported via extensive measurements [26]. The entire 40 MHz channel is partitioned into blocks of size equal to the coherence bandwidth $B_c \approx 795.6$ KHz. Three independent Rayleigh waveforms are generated for each block using the modified Jakes fading model and a weighted sum is taken to calculate the SNR. A scheduling epoch is of 5 ms duration, and an evaluation period consists of 1000 scheduling epochs. The number of PUs is set to 60, and the number of SUs varies.

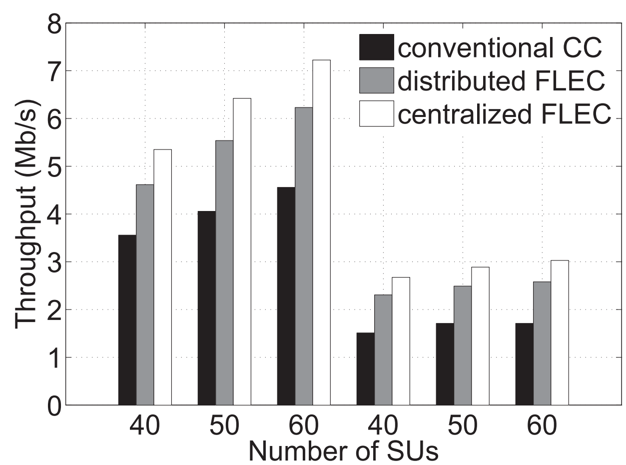


Fig. 3. Overall throughput performances. The first three bars represent PU throughput, and the last three bars represent SU throughput.

6.1 Overall performance of FLEC

We first evaluate the overall performance of distributed and centralized FLEC compared with conventional identical channel cooperation ("CC" in the figures). We use *Centralized Heuristic for CC* to derive CC performance as the benchmark here. In Fig. 3, we plot the average throughput of both PUs (first three bars) and SUs (last three bars). We can see that *Distributed Bargaining for FLEC* and *Centralized Heuristic for FLEC* as in Sec. 3-4 provide 20–40% and 30–60% improvement, respectively. It clearly demonstrates the advantage of FLEC. A similar trend is also observed for SUs, although the improvement becomes marginal when the number of SUs scales up. The reason is that, though a larger number of SUs provides more and better cooperation for PUs and thus improves their throughput, it results in fewer channels leased to each SU, and a lower degree of optimization freedom.

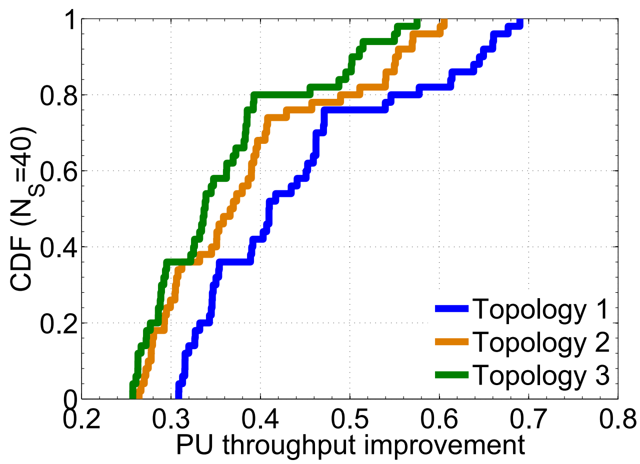
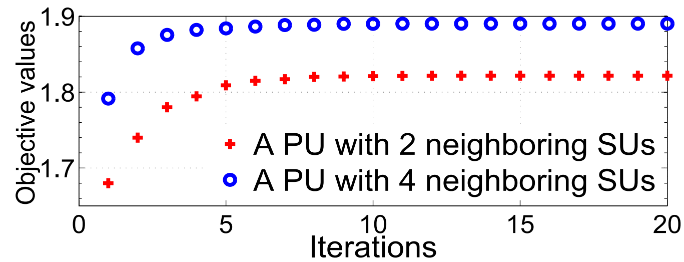
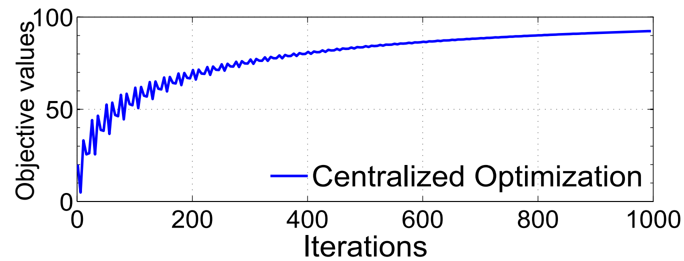


Fig. 4. Effects of topology on PU throughput improvement.



(a) Fast convergence of *Distributed Bargaining for FLEC.*



(b) Slow convergence of *Centralized Optimization for FLEC.*

Fig. 5. Convergence of the algorithms ($N_S = 50$).

TABLE 1
Running time of component algorithms in *Centralized Heuristic for FLEC*.

Component algorithm	Ave. running time (ms)	Min.	Max.
relay assignment	0.34	0.32	0.35
subchannel assignment	0.26	0.2	0.29
power allocation	0.19	0.17	0.2

6.2 Effects of Topology

Next we investigate the effects of topology on FLEC. We choose *Distributed Bargaining* as the representative algorithm, and evaluate three representative topologies, where N_S equals 40 and the average distance from PU to BS is controlled to be 0.8, 0.65, and 0.5 of the cell radius (topology 1, 2, and 3 respectively). We observe from Fig. 4 that while cooperation always results in some improvement in PU throughput, scenarios dominated by high path loss and poor shadowing benefit the most (topology 1), as more cooperation opportunities can be explored. SU's throughput also becomes better in these scenarios, which we do not show here due to space limit. This observation justifies the deployment of SU cooperation for throughput enhancement in primary networks with high path loss and limited coverage.

6.3 Practicality of FLEC algorithms

In this section we are concerned with the practicality of the FLEC algorithms. First, we study the convergence of our distributed algorithm for decentralized FLEC. Fig. 5 shows the convergence of *Distributed Bargaining* for two randomly chosen PUs with different number of neighboring SUs. It is clear that the distributed algorithm converges within 20 iterations, validating its feasibility in practice. The reason for the fast convergence, as discussed in Sec. 3.4, is mainly the limited size of neighborhood. With distributed and concurrent operations, it is indeed suitable for practical implementation.

We then study the algorithms for the centralized problem. We first observe that *Centralized Optimization for FLEC* does not converge even after 1000 iterations in Fig. 5. This echoes our concern about the complexity of centralized subgradient update of two vector dual variables in Sec. 4, and justifies our motivation to design efficient heuristics.

To understand the practicality of *Centralized Heuristic*, we observe the running time of its component algorithms in our simulations. Table 1 summarizes the average as well as the minimum and maximum running time of the three component algorithms. We can see that all of them are on the order of milliseconds on an Intel Xeon Quad-core CPU running at 3 GHz with 2 GB memory and without any multi-threading. Therefore the usual scheduling deadlines of 5–10 ms [6] can be satisfactorily met. Through the discussions here, we summarize that both the *Distributed Bargaining* and *Centralized Heuristic* are practical in terms of running time.

6.4 Near-optimality of *Centralized Heuristic*

We now evaluate the performance loss of *Centralized Heuristic* compared with that of *Centralized Optimization*. Recall that *Centralized Optimization* is developed via the same methodology of dual decomposition and subgradient update as used in *Distributed Bargaining* in Sec. 4. As seen from Fig. 6, with respect to the average throughput of both PU and SU, *Centralized Heuristic* losses about 5% in all cases. We also evaluate the performance loss with different values of $p_i^{\max}$, and find

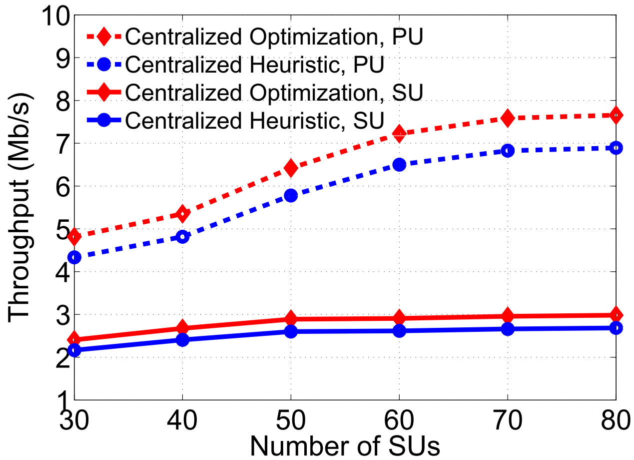


Fig. 6. Near-optimality of *Centralized Heuristic*.

that the gap widens when $p_i^{\max}$ decreases. The reason is that our subchannel assignment is based on the assumption of equal power allocation, which becomes invalid when $p_i^{\max}$ is small, and affects the performance of our *Centralized Heuristic*. Numerical details are not presented because of limited space. Due to slow convergence of *Centralized Optimization*, we may conclude that *Centralized Heuristic* achieves a good tradeoff between performance and complexity, and is amenable to practical implementations.

7 Related Work

A plethora of work has been done on spectrum sharing based on cognitive radio [2]. Generally, they fall into three paradigms, *interweave*, *underlay*, and *overlay* [27]. The interweave paradigm insists that SUs should only transmit when PUs are not, while the underlay paradigm allows SUs to transmit concurrently with PUs provided that their signals do not cause harmful interference. Essentially, in both cases SUs are transparent to PUs. The *overlay* paradigm, which is the focus of this paper, assumes PUs have side information about SUs, and leverages them to improve primary network performance. However, most existing works focus on information theoretic analysis [28].

In networking literature, [4] first proposes the idea of cooperative cognitive radio network, where the secondary users can earn spectrum access in exchange for cooperation with the primary user. A Stackelberg game is formulated where the primary user acts as the leader and determines the optimal time sharing strategy in maximizing its transmission rate. [5] considers a slightly different setting where the traffic demand of primary user is taken into account, and the utility function includes a revenue component from secondary users. [29] considers the game of one PU and multiple SUs in which the PU decides the portion of access time and the SU decide the relay power level. In [30] a priority queueing system model is developed, and in [16] a credit-based spectrum sharing scheme is studied for cooperative cognitive radio network. These work adopt a single shared channel setting with a single primary user and an ad-hoc network of secondary users. On contrary, in this paper we consider a multi-channel setting where the OFDMA based primary and secondary networks co-locate, which represents a more practical network scenario and has not been considered before.

Resource allocation with cooperative diversity has been extensively studied in general wireless networks [31]–[33]. Specifically, our paper is more related to work in cognitive radio or cooperative OFDMA networks. For the former, most work [34]–[36] consider maximizing SUs' throughput with constrained interference to PUs. In other words, they all consider the underlay paradigm. For the latter, most related to our work are [13] and [14]. [13] addresses the problem with a joint consideration of relay assignment, channel allocation, relay strategy optimization, and power control. Our previous work [14] considers the problem with a novel network coding based cooperation strategy, and proposes approximation algorithms with performance guarantees. Compare to these work, we consider the performance of primary and secondary users *jointly*, and apply the concept of Nash bargaining solutions [10] to ensure both parities benefit from cooperation fairly.

The application of Nash bargaining to multi-criteria optimization is not new in the networking field. [37] applies it to ensure fairness in a network flow control problem. Kelly in the seminal work [11] has also shown that Nash bargaining ensures proportional fairness in a TCP setting. NBS has been also applied to allocate resources in cooperative OFDMA networks [18], [19]. These works do not consider the inefficiency of conventional cooperation methods in the context of multi-channel CCRN, and only heuristics without any performance bounds are given. Finally, our conference version of this paper [1] does not study the identical channel cooperation in details.

8 Concluding Remarks

This work represents an early attempt to study OFDMA cooperative cognitive radio networks. The central question addressed is how to effectively exploit secondary user cooperation when conventional cooperation method becomes inefficient in this scenario, which has not yet been explored. We propose FLEC, a flexible channel cooperation design to allow SUs to customize the use of leased resources in order to maximize performance. We develop a unifying optimization framework based on Nash bargaining solutions to address the resource allocation problem with FLEC, where relay assignment, subchannel assignment, relay strategy optimization and power control intricately interplay with one another. An optimal distributed algorithm as well as an efficient centralized heuristic with near-optimal performance are proposed. We also extend our framework to consider resource allocation with conventional cooperation.

References

[1] H. Xu and B. Li, "Efficient resource allocation with flexible channel cooperation in OFDMA cognitive radio networks," in *Proc. IEEE INFOCOM*, 2010.

[2] S. Haykin, "Cognitive radio: Brain-empowered wireless communications," *IEEE J. Sel. Areas Commun.*, vol. 23, no. 2, pp. 201–220, February 2005.

[3] I. Akyildiz, W.-Y. Lee, M. Vuran, and S. Mohanty, "Next generation/dynamic spectrum access/cognitive radio wireless networks: A survey," *Elsevier Comput. Netw.*, vol. 50, pp. 2127–2159, September 2006.

[4] O. Simeone, I. Stanojev, S. Savazzi, Y. Bar-Ness, U. Spagnolini, and R. Pickholtz, "Spectrum leasing to cooperating secondary ad hoc networks," *IEEE J. Sel. Areas Commun.*, vol. 26, no. 1, pp. 203–213, January 2008.

[5] J. Zhang and Q. Zhang, "Stackelberg game for utility-based cooperative cognitive radio networks," in *Proc. ACM MobiHoc*, 2009.

[6] IEEE Standard, "802.16^{TM}: Air interface for fixed wireless access systems," 2005.

[7] A. Sendonaris, E. Erkip, and B. Aazhang, "User cooperation diversity — Part I: System description," *IEEE Trans. Commun.*, vol. 51, no. 11, pp. 1927–1938, November 2003.

[8] M. Andrews and L. Zhang, "Scheduling algorithms for multi-carrier wireless data systems," in *Proc. ACM MobiCom*, 2007.

[9] S. Deb, V. Mhatre, and V. Ramaiyan, "Wimax relay networks: Opportunistic scheduling to exploit multiuser diversity and frequency selectivity," in *Proc. ACM MobiCom*, 2008.

[10] G. Owen, *Game Theory*. Academic Press, 2001.

[11] F. P. Kelly, A. K. Maulloo, and D. K. H. Tan, "Rate control for communication networks: Shadow prices, proportional fairness and stability," *J. Operat. Res. Soc.*, vol. 49, no. 3, pp. 237–252, March 1998.

[12] S. Boyd and A. Mutapcic, "Subgradient methods," Lecture notes of EE364b, Stanford University, Winter Quarter 2006-2007. http://www.stanford.edu/class/ee364b/notes/subgrad_method_notes.pdf.

[13] T. C.-Y. Ng and W. Yu, "Joint optimization of relay strategies and resource allocations in cooperative cellular networks," *IEEE J. Sel. Areas Commun.*, vol. 25, no. 2, pp. 328–339, February 2007.

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

[15] Z. Zhang, J. Shi, H.-H. Chen, M. Guizani, and P. Qiu, "A cooperation strategy based on Nash bargaining solution in cooperative relay networks," *IEEE Trans. Veh. Technol.*, vol. 57, no. 4, pp. 2570–2577, July 2008.

[16] D. Niyato and P. Wang, "Credit-based spectrum sharing for cognitive mobile multihop relay networks," in *Proc. IEEE WCNC*, 2010.

[17] Y. Xiang, J. Luo, and C. Hartmann, "Inter-cell interference mitigation through flexible resource reuse in OFDMA based communication networks," in *Proc. 13th European Wireless Conference (EW)*, 2007.

[18] Z. Han, Z. J. Ji, and K. J. R. Liu, "Fair multiuser channel allocation for OFDMA networks using Nash bargaining solutions and coalitions," *IEEE Trans. Commun.*, vol. 53, no. 8, pp. 1366–1376, August 2005.

[19] K.-D. Lee and V. Leung, "Fair allocation of subcarrier and power in an OFDMA wireless mesh network," *IEEE J. Sel. Areas Commun.*, vol. 24, no. 11, pp. 2051–2060, November 2006.

[20] J. Huang, V. Subramanian, R. Agrawal, and R. Berry, "Joint scheduling and resource allocation in uplink OFDM systems for broadband wireless access networks," *IEEE J. Sel. Areas Commun.*, vol. 27, no. 2, pp. 226–234, February 2009.

[21] G. Song and Y. G. Li, "Cross-layer optimization for OFDM wireless network — Part I and II," *IEEE Trans. Wireless Commun.*, vol. 4, no. 2, March 2005.

[22] W. Yu and R. Lui, "Dual methods for nonconvex spectrum optimization of multicarrier systems," *IEEE Trans. Commun.*, vol. 54, no. 7, pp. 1310–1322, July 2006.

[23] C. Papadimitriou and K. Steiglitz, *Combinatorial Optimization: Algorithms and Complexity*. Prentice Hall, 1998.

[24] J. Jang and K. B. Lee, "Transmit power adaptation for multi-user OFDM systems," *IEEE J. Sel. Areas Commun.*, vol. 21, no. 2, pp. 171–179, February 2003.

[25] L. Fleischer, M. Goemans, V. Mirrokni, and M. Sviridenko, "Tight approximation algorithms for maximum general assignment problems," in *Proc. ACM SODA*, 2006.

[26] IEEE 802.16 task group, "Channel models for fixed wireless applications," July 2001.

[27] A. Goldsmith, S. A. Jafar, I. Maric, and S. Srinivasa, "Breaking spectrum gridlock with cognitive radios: An information theoretic perspective," *Proc. IEEE*, vol. 97, no. 5, pp. 894–914, May 2009.

[28] I. Marić, R. Yates, and G. Kramer, "Capacity of interference channels with partial transmitter cooperation," *IEEE Trans. Inf. Theory*, vol. 53, no. 10, pp. 3536–3548, October 2007.

[29] H. Wang, L. Gao, X. Gan, X. Wang, and E. Hossain, "Cooperative spectrum sharing in cognitive radio networks: A game-theoretic approach," in *Proc. IEEE ICC*, 2010.

[30] C. Zhang, X. Wang, and J. Li, "Cooperative cognitive radio with priority queueing analysis," in *Proc. IEEE ICC*, 2009.

[31] A. Bletsas, A. Khisti, D. Reed, and A. Lippman, "A simple cooperative diversity method based on network path selection," *IEEE J. Sel. Areas Commun.*, vol. 24, no. 3, pp. 659–672, March 2006.

[32] Y. Zhao, R. Adve, and T. Lim, "Improving amplify-and-forward relay networks: Optimal power allocation versus selection," *IEEE Trans. Wireless Commun.*, vol. 6, no. 8, pp. 3114–3123, August 2007.

[33] L. Le and E. Hossain, "Cross-layer optimization frameworks for multihop wireless networks using cooperative diversity," *IEEE Trans. Wireless Commun.*, vol. 7, no. 7, pp. 2592–2602, July 2008.

[34] R. Urgaonkar and M. Neely, "Opportunistic scheduling with reliability guarantees in cognitive radio networks," *IEEE Trans. Mobile Comput.*, vol. 8, no. 6, pp. 766–777, June 2009.

[35] L. B. Le, P. Mitran, and C. Rosenberg, "Queue-aware subchannel and power allocation for downlink OFDM-based cognitive radio networks," in *Proc. IEEE WCNC*, 2009.

[36] H. Kim and K. G. Shin, "Asymmetry-aware real-time distributed joint resource allocation in IEEE 802.22 WRANs," in *Proc. IEEE INFOCOM*, 2010.

[37] R. Mazumdar, L. G. Mason, and C. Doulgligeris, "Fairness in network optimal flow control: Optimality of product forms," *IEEE Trans. Commun.*, vol. 39, no. 5, pp. 775–782, May 1991.

Hong Xu received his B.Engr degree from the Department of Information Engineering, The Chinese University of Hong Kong, and his M.A.Sc. degree from the Department of Electrical and Computer Engineering, University of Toronto in 2010. Since 2009, he has been a Ph.D. candidate in the Department of Electrical and Computer Engineering, University of Toronto. His research interests include large-scale networked systems, cloud computing, network economics, and wireless networking. His current focus is on the design and analysis of market based resource allocation in cloud computing using economics and optimization theory. He has also worked on resource allocation in cognitive radio networks and cellular networks using cross-layer optimization and mechanism design.

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

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

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