Joint TDD Backhaul and Access Optimization in Dense Small-Cell Networks

Mehrdad Shariat, Emmanouil Pateromichelakis, Atta ul Quddus, and Rahim Tafazolli

***Abstract*—This paper addresses the problem of joint backhaul (BH) and access link optimization in dense small-cell networks with a special focus on time-division duplexing (TDD) mode of operation in BH and access link transmission. Here, we propose a framework for joint radio resource management, where we systematically decompose the problem in BH and access links. To simplify the analysis, the procedure is tackled in two stages. At the first stage, the joint optimization problem is formulated for a point-to-point scenario where each small cell is simply associated with a single user. It is shown that the optimization can be decomposed into separate power and subchannel allocation in both BH and access links, where a set of rate-balancing parameters in conjunction with duration of transmission governs the coupling across both links. Moreover, a novel algorithm is proposed based on grouping the cells to achieve rate balancing in different small cells. Next, in the second stage, the problem is generalized for multiaccess small cells. Here, each small cell is associated with multiple users to provide the service. The optimization is similarly decomposed into separate subchannel and power allocation by employing auxiliary slicing variables. It is shown that similar algorithms, as in the previous stage, are applicable by a slight change with the aid of slicing variables. Additionally, for the special case of line-of-sight BH links, simplified expressions for subchannel and power allocation are presented. The developed concepts are evaluated by extensive simulations in different case studies from full orthogonalization to dynamic clustering and full reuse in the downlink, and it is shown that the proposed framework provides significant improvement over the benchmark cases.**

***Index Terms*—Backhauling, joint optimization, rate balancing, small cells.**

I. Introduction

CAPACITY demand in cellular networks has been exponentially growing, and this trend is expected to continue [1]. The ultimate vision is to provide consistent and reliable communication to create the perception of "infinite capacity" for end users. So far, Long-Term Evolution of 3G (LTE) and LTE-Advanced [2], [3] are key intermediate steps in smooth migration toward this vision of future wireless networks. In particular, small cells and advanced relay cells have emerged as fundamental elements in these systems [4]–[6] to enhance the efficiency in both capacity and coverage. On the other hand, a higher degree of interworking has been envisioned between small cells and conventional macrocells in the future via the notion of "dual connectivity," which implies that the user can have simultaneous connections to both macro- and small-cell stations [7], [8]. This feature presents potential advantages such as separation between control and data channels, where a macrolayer provides control signaling to small cells to enhance mobility, overhead, and energy efficiency, whereas small cells focus on information delivery in data plane [8]. This is the subject of an ongoing work in LTE-Advanced and beyond from Release 12 [7].

To support networks comprising several small cells, wired fiber backhauling might be cost prohibitive or difficult to widely deploy in short to medium terms. Therefore, utilizing wireless backhauling seems a promising migration path. Emerging wireless systems have already taken initial steps in this direction, e.g., in 3GPP LTE-Advanced, layer-3 relaying strategies require self-backhauling on wireless LTE radio bearers to backhaul (BH) traffic between relay and donor eNodeBs [5]. Layer-3 relays mimic many of the functionalities in eNodeBs in smaller scale unlike conventional layer-1 or layer-2 relays.

Wireless backhauling for small cells and relay cells can introduce several challenges [6] when it comes to radio resource management (RRM) between the BH and the access links. Here, coupling between the two can limit the multiuser diversity gain of radio resource allocation in the access link. In the case of time-division duplexing (TDD) between the transmission links, each packet would be received after two consecutive transmissions, i.e., on the BH link to the small-cell station and from the small-cell station to the end user (access link). Therefore, the efficiency of the system depends upon the balance of resources between the two links for each small cell. This requires efficient resource partitioning on BH or access links and rate-balancing strategies between the two.

A. Related Works

There exists a very large literature in the area of RRM and resource allocation for radio access networks such as [9] and [10], where they fundamentally develop a framework for resource allocation in the context of orthogonal frequency-division multiple access (OFDMA) systems. Moreover, there are several research studies addressing resource allocation strategies in the context of relay-based and small-cell networks [11]–[18]. Here, [11] presents an integrated RRM solution for relay-based

Manuscript received December 9, 2013; revised June 3, 2014; accepted November 4, 2014. Date of publication December 9, 2014; date of current version November 10, 2015. This work was supported in part by the European Commission within the 7th Framework Program through the ICT Project iJOIN under Grant Agreement 317941 and in part by the University of Surrey, Guildford, U.K., through the 5GIC programme. The views and conclusions contained herein are those of the authors and should not be interpreted as necessarily representing the official policies or endorsements, either expressed or implied, of the iJOIN project or the European Commission. The review of this paper was coordinated by Prof. X. Wang.

The authors are with the Institute for Communication Systems, Department of Electronic Engineering, University of Surrey, Guildford GU2 7XH, U.K. (e-mail: m.shariat@surrey.ac.uk).

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TVT.2014.2379013

networks in the context of code-division multiple-access systems. In [12]–[14], heuristic resource allocation strategies are offered for in-band relaying in OFDMA networks. In particular, [14] focuses on uplink in-band relaying via coscheduling and load balancing between macro and relay cell users. In [15], Ng and Yu explore resource allocation and relay selection strategies for OFDMA systems when a hybrid of amplify-and-forward and decode-and-forward relays, i.e., layer-1 and layer-2 relays, is employed. In [16], the possibility of concurrent transmission between relays and macrocell networks is analyzed based on extreme value theory. Li *et al.* in [17] propose a cooperative resource allocation strategy between macro and relay stations in downlink transmission, and Madan *et al.* in [18] focus on cell-association strategies for small cells to achieve better load balancing between macro and small cells. A detailed survey on RRM for multicarrier cellular networks can be found in [19]. Additionally, [20] provides an overview of radio resource allocation schemes designed for relay-enhanced systems.

The aforementioned research studies are quite inspiring. However, some are mainly focused on the access side [9], [10], [18], particularly for conventional macro deployments [9], [10]. Others propose metrics that are not readily applicable to multicarrier systems nor apply constraints imposed by the BH limitation [11], [17]. On the other hand, some solutions are heuristic in nature without detailed mathematical insights [12]–[14]. More importantly, the majority of solutions assumes a fixed duration of transmission between the BH and access links and not utilizing an important degree of freedom that we intend to explore in this paper.

B. Contributions

In this paper, we systematically decouple the problem of wireless backhauling for small cells via decomposition theory [21], [22], taking into account the interaction of resource allocation parameters in different links and layers. Here, we extend our previous works in [23] and [24]. In [23], we proposed a low-complexity generic framework for resource allocation based on time sharing applicable to different scenarios in downlink and uplink. On the other hand, in [24], we proposed an efficient graph-based dynamic clustering framework to control the level of cotier interference between small cells. Combining the two solutions leads us to decouple the problem of resource allocation for small-cell networks across BH and access links by introducing a set of governing variables as a priority factor. The decoupled elements facilitate efficient RRM strategies that are flexible and applicable to a wide set of cases from full orthogonalization to full spectrum reuse between small cells. In particular, we present a novel framework to update the duration of transmissions in downlink between the BH and access links in TDD mode in conjunction with resource allocation in both links. Here, the time coupling element across the links is challenging yet presents another degree of freedom to jointly optimize the system performance. The mathematical derivations and the proposed solutions are novel and have not been explored in the literature at this level to the best of our knowledge. Such RRM strategies enable a flexible cloud-based radio access network (Cloud-RAN) that can be fully or partially driven by the Cloud

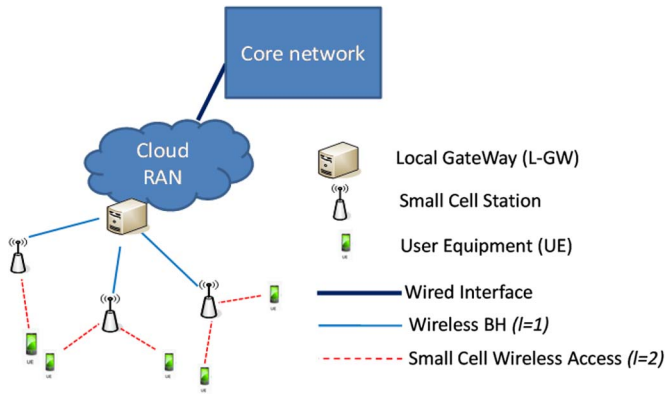

Fig. 1. Dense small-cell network.

[25], [26]. The outcome solutions are evaluated by extensive simulations in different case studies for downlink transmission.

II. System Model

Here, the system consists of a local Cloud-RAN Gateway (L-GW) and a set of small-cell stations $\boldsymbol{M}$, transmitting information to a set of users $\boldsymbol{N}$ over the set of subchannels $\boldsymbol{K}$. The users are grouped into disjoint subsets $\boldsymbol{N}_m$ such that $\boldsymbol{N} = \cup_m(\boldsymbol{N}_m)$ $(m \in 1, 2, \ldots, M)$. We also define $\boldsymbol{N}_0$ equivalent $\boldsymbol{M}$ to cover the BH transmission phase. We assume that $l = 1$ covers the BH transmission, whereas $l = 2$ refers to the access link.

Fig. 1 shows a typical example of our target network with $|\boldsymbol{M}| = M = 3$, $|\boldsymbol{N}| = N = 5$. Here, we assume that each subchannel is constrained to be exclusively used by a single link within small cells per time instance. However, across different small cells, different case studies are considered from orthogonal allocation to dynamic clustering and, finally, the full reuse of subchannels. Channel knowledge is considered to be available at L-GW entities connected to Cloud-RAN across the network. The Cloud-RAN processes the subchannel and power allocation and adjusts the transmission duration for BH and access links.

III. Problem Formulation

The resource (subchannel, power, and transmission duration) allocation problem can be mathematically formulated as an optimization problem with a certain objective function subject to involving constraints in the problem domain. Assuming weighted sum-rate (WSR) maximization as the optimization objective, a generic representation of the problem is as follows:

$$\max_{\boldsymbol{p},\boldsymbol{\tau},\alpha} \sum_{n\in\boldsymbol{N}} w_n R_n \text{ s.t.} \tag{1}$$

$$\sum_{n\in\boldsymbol{N}_{\tilde{m}}} \sum_{k\in\boldsymbol{K}} \tau_{n,k}^{(l)} p_{n,k}^{(l)} \le P_{\max,\tilde{m}}$$

$$\tilde{m} = \begin{cases} 0, & l = 1 \\ m \in \{1, 2, \ldots, M\}, & l = 2 \end{cases} \tag{2}$$

$$\sum_{n\in\boldsymbol{N}_{\tilde{m}}} \tau_{n,k}^{(l)} \le 1 \qquad \forall\, k \in \boldsymbol{K} \tag{3}$$

$$\tau_{n,k}^{(l)} \in \{0,1\} \qquad \forall\, n \in \boldsymbol{N}, \quad \forall\, k \in \boldsymbol{K} \tag{4}$$

$$R_n = R_{n(m)}\left(\alpha, \tau_{n,k}^{(l)}, p_{n,k}^{(l)}\right), \quad m \in \{1,2,\ldots,M\}. \tag{5}$$

Here, R_n represents the outcome average rate n of user over allocated subchannels across BH and access links, whereas $p_{n,k}^{(l)}$ is the instantaneous power that user n transmits over subchannel k on link (l). $\tau_{n,k}^{(l)}$ denotes the fraction of time that user n is allowed to exclusively use this subchannel on link (l) per cell. Obviously, (2) provides the power limitation constraint per node, i.e., $P_{\max,\tilde{m}}$, whereas (3) in conjunction with (4) imposes the intracell orthogonality of resource allocation, i.e., the exclusivity constraint. Equation (5) shows that the rate R_n per user n would be equivalent to the rate allocated to this user out of its serving small cell, i.e., m. In particular, it is worth noting that $R_{n(m)}$ depends on the allocated share of resources $\tau_{n,k}^{(l)}$, $p_{n,k}^{(l)}$ to a small cell on different links in BH $(l=1)$ and access sides $(l=2)$, the spectrum reuse strategy across different small cells, and, finally, the parameter α. Here, $\alpha \in (0,1)$ is the parameter to adjust the duration of transmission between BH and access links assuming TDD. We will detail this dependence later in this paper.

The grouping of (3) with (4) turns the optimization into a combinatorial problem that is intractable for large sets of subchannels and users. Furthermore, constraint (5) couples the resource allocation problem across both links, as well as between different small cells based on the spectrum reuse strategy.

In the following sections, the joint resource allocation procedure is addressed in detail for different scenarios.

A. Point-to-Point Small Cells (PPS)

To simplify the analysis, first, a point-to-point model is considered where each small cell is exclusively associated with a single user.

Problem A (PPS Without Time Sharing): Maximize the WSR of users by performing the resource allocation while satisfying the subchannel and power allocation constraints.

This problem is very similar to the original problem (1)–(5). Here, we try to expand the definition of (5) from the capacity point of view taking into account the PPS model, i.e.,

$$R_n = R_{n(m)}\left(\alpha, \tau_{n,k}^{(l)}, p_{n,k}^{(l)}\right) = \min\left(\alpha r_n^{(1)}, (1-\alpha) r_n^{(2)}\right) \tag{6}$$

$$r_n^{(l)} = \sum_{k \in \mathrm{K}} \tau_{n,k}^{(l)} \log_2\left(1 + \rho_{n,k}^{(l)} p_{n,k}^{(l)}\right). \tag{7}$$

$r_n^{(l)}$ is the achievable rate of user n at link (l), whereas $\rho_{n,k}^{(l)}$ represents the signal-to-noise ratio or signal-to-interference-plus-noise ratio density function of the constituent link. In particular, for the second phase, we have $\rho_{n,k}^{(2)} = g_{n(m),k}^{(2)} \Big/ \left(\sum_{m'} g_{n(m'),k}^{(2)} p_{n(m'),k}^{(2)} + \eta\right)$, $m' \neq m$.

Here, $g_{n(m),k}^{(2)}$ represents the channel gain of user n from serving cell m, whereas $g_{n(m'),k}^{(2)} p_{n(m'),k}^{(2)}$ models the resulting interference from reusing the same subchannel by small cell m' in nonorthogonal case studies. η models the noise power over the target subchannel. It is important to note that all the small cells are associated with L-GW of Cloud RAN for the BH link while each user is coupled to its corresponding small cell for the access link based on (7). As a result, we have $p_{n,k}^{(1)} = p_{m,k}^{(1)}$ and $\rho_{n,k}^{(1)} = \rho_{m,k}^{(1)}$, where m is the serving small cell for user n.

1) Relaxation: Problem A is still combinatorial due to the exclusivity constraint. Moreover, the minimization function as in (6) strongly couples the resource allocation across the links. To tackle these issues, the exclusivity constraint of (4) can be initially relaxed allowing for the time sharing of resources, i.e.,

$$\tau_{n,k}^{(l)} \in [0,1] \qquad \forall\, n \in \boldsymbol{N}, \quad \forall\, k \in \boldsymbol{K}. \tag{8}$$

The exclusivity constraint can be reimposed on the outcome solution, similar to the method we proposed in [23] with negligible performance loss. Additionally, the minimization problem can be interpreted as the following subproblems:

Subproblem A.1 (Bottleneck on the BH): In this regime, the achievable rate of access, from small cells to their associated users, is greater than the BH. As a result, this subproblem is formulated as follows subject to constraints (2), (3), and (8):

$$\max_{\boldsymbol{p},\boldsymbol{\tau},\alpha} \sum_{n \in \boldsymbol{N}} w_n \alpha r_n^{(1)} \text{ s.t.} \tag{9}$$

$$\alpha r_n^{(1)} \leq (1-\alpha) r_n^{(2)} \qquad \forall\, n \in \boldsymbol{N}. \tag{10}$$

As can be seen, constraint (10) is employed to ensure the operation within the assumed regime.

Subproblem A.2 (Bottleneck on the Access): Here, contrary to the previous case, the BH is dominant due to the superiority of its channel condition. Hence, this subproblem is formulated as follows subject to similar power and time-sharing constraints as A.1:

$$\max_{\boldsymbol{p},\boldsymbol{\tau},\alpha} \sum_{n \in \boldsymbol{N}} w_n (1-\alpha) r_n^{(2)} \text{ s.t.} \tag{11}$$

$$(1-\alpha) r_n^{(2)} \leq \alpha r_n^{(1)} \qquad \forall\, n \in \boldsymbol{N}. \tag{12}$$

Similar to subproblem A.1, constraint (12) monitors the operation within the boundaries of the assumed regime. It can be easily shown that any other combination of rates can be mapped into one of the aforementioned regimes through rearranging the resources among the users. Hence, by assuming the operation in either regime, the relaxed problem can be efficiently solved.

Remark 1: Subproblems A.1 and A.2 are not generally convex due to the presence of intercell interference on the access links in case of spectrum reuse. However, by applying the dynamic clustering framework proposed in [24], users of different cells can be effectively grouped into clusters with a low level of intracluster interference. Therefore, we decouple the interference coordination problem from resource allocation by mapping the problem to this graph-based solution. This leads to near convexity of the given subproblems, enabling us to effectively decompose them as will be outlined below.

To tackle the relaxed problems, we initially focus on A.1, and we form partial Lagrangian as follows subject to constraints (2),

(3), and (8):

$$L(\boldsymbol{p},\boldsymbol{\tau},\alpha,\boldsymbol{\mu}) = \sum_{n\in\boldsymbol{N}} w_n \alpha r_n^{(1)} - \sum_{n\in\boldsymbol{N}} \mu_n \left(\alpha r_n^{(1)} - (1-\alpha) r_n^{(2)}\right). \quad (13)$$

Here, the partial dual problem can be formed as $\xi(\boldsymbol{\mu}) \triangleq \sup_{\boldsymbol{p},\boldsymbol{\tau},\alpha} L(\boldsymbol{p},\boldsymbol{\tau},\alpha,\boldsymbol{\mu})$ subject to time-share and power constraints where μ_n (*or* $\boldsymbol{\mu}$) are dual prices introducing a set of balancing parameters between BH and access links. As it can be seen, subproblem A.1 can be easily decoupled into two WSR problems for both links based on (13) with weights $\phi_n^{(1)} = (w_n - \mu_n)$ and $\phi_n^{(2)} = \mu_n$, respectively. Similar decoupling can be achieved for A.2 by simply swapping the weights. This partial dual decomposition enables us to apply power and time-share allocation optimality conditions as derived in [23] with slight tuning of the weights for individual links.

2) Power Allocation:

Lemma 1: For a fixed time-sharing policy and transmission duration α, the optimal power allocation variables for subproblem A.1 or A.2 will be the solutions of the following set of optimization problem:

$$p_{n,k}^{(l)} = \arg\max_{\boldsymbol{p}} \left\{\phi_n^{(l)} \log_2\left(1+\rho_{n,k}^{(l)} p\right) - v_{m_n} p\right\} \quad \forall n \in \boldsymbol{N}, \quad \forall k \in \boldsymbol{K} \quad (14)$$

where v_{m_n} are the power prices to satisfy (2), and $\phi_n^{(l)}$ are the balancing parameters to govern the balance between both links according to either (10) or (12).

Proof: It is straightforward to show (14) by applying standard dual decomposition with dual variables v_{m_n} for (2) and considering μ_n as (13), where

$$\phi_n^{(l)} = \begin{cases} (w_n - \mu_n), & l = 1 \\ \mu_n, & l = 2. \end{cases} \quad (15)$$

Following a similar approach for A.2, the same optimality condition holds unless (15) that should be amended as follows:

$$\phi_n^{(l)} = \begin{cases} \mu_n, & l = 1 \\ (w_n - \mu_n), & l = 2. \end{cases} \quad (16)$$

□

Corollary 1: The solution to the power allocation problems (14) is similar to the standard multilevel waterfilling as in (17) $[x]^+ = \max(0, x)$, where

$$p_{n,k}^{(l)} = \left[\frac{\phi_n^{(l)}}{\ln(2) v_{m_n}} - \frac{1}{\rho_{n,k}^{(l)}}\right]^+. \quad (17)$$

Proof: A straightforward result of Lemma 1. □

3) Time-Share Allocation: Similar to the power allocation scenario, by fixing the average power $q_{n,k}^{(l)} = p_{n,k}^{(l)} \tau_{n,k}^{(l)}$ and α, the time-share variables can be retuned to increase the aggregate WSR across the network.

Lemma 2: For fixed allocated average power and transmission duration α, the optimal time-share variables for subproblem A.1 or A.2 will be the solutions of the following set of optimization problems:

$$\left(\tau_{1,k}^{(l)}, \ldots, \tau_{N,k}^{(l)}\right) = \arg\max_{\boldsymbol{\tau}} \left\{\sum_{n\in\boldsymbol{N}} \phi_n^{(l)} \tau_n \log_2\left(1+\rho_{n,k}^{(l)} q_{n,k}^{(l)} \big/ \tau_n\right)\right\} \quad \forall k \in \boldsymbol{K} \quad (18)$$

subject to constraints (3) and (8).

Proof: By fixing the average allocated power and the transmission duration, the optimization will be decomposed to the given equations per subchannel as the power constraints (2) are the only factors that couple the optimization problem across the subchannels [23]. □

Corollary 2: The time-share solutions of (18) can be found by solving the following set of equations:

$$U_n\left(y^{(l)}\right) = \phi_n^{(l)} \left[\log_2\left(1+y^{(l)}\right) - \frac{y^{(l)}}{\ln(2)\left(1+y^{(l)}\right)}\right] = \lambda_k^{(l)}, \quad y^{(l)} = \rho_{n,k}^{(l)} \frac{q_{n,k}^{(l)}}{\tau_{n,k}^{(l)}} \quad (19)$$

where $U_n(y^{(l)})$ is the marginal utility of user n at link (l), and $\lambda_k^{(l)}$ are the time-share prices to satisfy constraints (3) and (8).

Proof: This is a direct result of Lemma 2 [23]. □

4) Adjusting the Duration of Transmission:

Theorem 1: There is an optimum value of α where the solutions of both subproblems A.1 and A.2 converge, which is the optimal solution of problem A, where

$$\frac{r_n^{*(2)}}{r_n^{*(1)}} = \frac{\alpha^*}{(1-\alpha^*)} \qquad \forall n \in \boldsymbol{N}, \alpha \in (0,1). \quad (20)$$

Proof: Considering subproblem A.1 and (13), the tuning parameters can be iteratively updated based on the subgradient method as follows:

$$\boldsymbol{\mu}^{(t+1)} = \left[\boldsymbol{\mu}^{(t)} + \varepsilon(t)\left(\alpha \boldsymbol{r}^{(1)(t+1)} - (1-\alpha)\boldsymbol{r}^{(2)(t+1)}\right)\right]_{D_{\boldsymbol{\mu}}} \quad (21)$$

where $[\,]_{D_\mu}$ represents the projection of the dual variable on the corresponding domain ($\mu > 0$). $\varepsilon(t)$ is the coefficient that regulates the step size in the (sub)gradient direction and can be chosen constant but sufficiently small to guarantee the convergence [21].

In a similar manner for subproblem A.2, the tuning parameters can be updated as follows:

$$\boldsymbol{\mu}^{(t+1)} = \left[\boldsymbol{\mu}^{(t)} + \varepsilon(t)\left((1-\alpha)\boldsymbol{r}^{(2)(t+1)} - \alpha \boldsymbol{r}^{(1)(t+1)}\right)\right]_{D_{\boldsymbol{\mu}}}. \quad (22)$$

As the dual problem converges to the optimal dual prices ($t \to \infty : \boldsymbol{\mu}^{(t+1)} \approx \boldsymbol{\mu}^{(t)}$), we can clearly see that there is a unique

solution for α^* as follows for both subproblems:

$$\underset{t\to\infty}{\text{Limit}} \frac{r_n^{(2)(t)}}{r_n^{(1)(t)}+r_n^{(2)(t)}} = \alpha^* \quad \forall n \in \boldsymbol{N} \text{ where } \alpha \in (0,1). \tag{23}$$

Considering the PPS model, each user is exclusively attached to a single small cell; therefore, we can define $\alpha_m^{(t)}$ as follows for each small cell per iteration:

$$\alpha_m^{(t)} = \frac{r_n^{(2)(t)}}{r_n^{(1)(t)}+r_n^{(2)(t)}} \qquad \forall n \in \boldsymbol{N}, \quad \forall m \in \{1,2,\ldots,M\}. \tag{24}$$

Considering $\alpha_m^{(t)}$ as the optimal transmission duration for small cell m at iteration t, the transmission durations might independently evolve for different cells across different iterations of the optimization process. However, they will converge to the optimal α^* asymptotically based on Theorem 1. Defining average $\overline{\alpha}$ as (25), it is straightforward to show that this average transmission duration value also asymptotically converges to optimal α^* based on (23) and (24), i.e.,

$$\overline{\alpha}^{(t)} = \frac{1}{M}\sum_m \alpha_m^{(t)}. \tag{25}$$

This averaging mechanism provides a unique value per iteration across all cells to update dual prices based on (23) or (24) that it asymptotically reaches optimal solution.

To understand the optimality condition, it is instructive to examine the effect of adjusting the transmission duration on the joint capacity of BH and access links.

By definition, the instantaneous capacity region $\boldsymbol{C}_{\boldsymbol{\pi}}(\boldsymbol{\rho})$ is a set that consists of all the achievable rate vectors for the current channel state vector $\boldsymbol{\rho}$ under the constraints of a given resource allocation policy $\boldsymbol{\pi}$. For example, in problem A, (2) provides the constraints on the power allocation $\boldsymbol{p}$, whereas (3) and (4) impose the constraints on the time-sharing policy $\boldsymbol{\tau}$. Mathematically, the instantaneous capacity region can be considered as the union of all achievable rate vectors under the considered policy, i.e.,

$$\boldsymbol{C}_{\boldsymbol{\pi}}(\boldsymbol{\rho}) = \cup_i \boldsymbol{R}_{\boldsymbol{i}_{\boldsymbol{\pi}}}(\boldsymbol{\rho}) \text{ where } \boldsymbol{R}_{\boldsymbol{i}_{\boldsymbol{\pi}}}(\boldsymbol{\rho}) \text{ satisfies } \boldsymbol{\pi}. \tag{26}$$

In the case of joint BH and access links, each element of a typical achievable rate vector is the minimum of achievable rates (of a target user) on both links according to (6). Based on this assumption, for a fixed α, the total capacity region will be the intersection of corresponding capacity regions on different links, i.e.,

$$\boldsymbol{C}_{\boldsymbol{\pi}}(\alpha,\boldsymbol{\rho}) = \boldsymbol{C}_{\boldsymbol{\pi}}^{(1)}\left(\alpha,\boldsymbol{\rho}^{(1)}\right) \cap \boldsymbol{C}_{\boldsymbol{\pi}}^{(2)}\left((1-\alpha),\boldsymbol{\rho}^{(2)}\right). \tag{27}$$

Here, increasing the transmission duration expands the capacity region of the BH, whereas it will shrink the region on the access link, and *vice versa*. The capacity region for any value of α can be considered as a typical set of achievable rate vectors. As a result, the union of these typical sets can form the total capacity region based on (26), i.e.,

$$\boldsymbol{C}_{\boldsymbol{\pi}}(\boldsymbol{\rho}) = \cup_\alpha \boldsymbol{C}_{\boldsymbol{\pi}}(\alpha,\boldsymbol{\rho}). \tag{28}$$

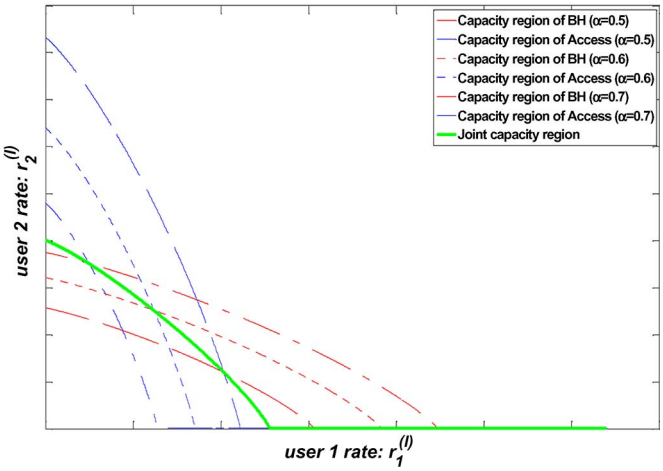

Fig. 2. Capacity boundaries in joint BH and access links.

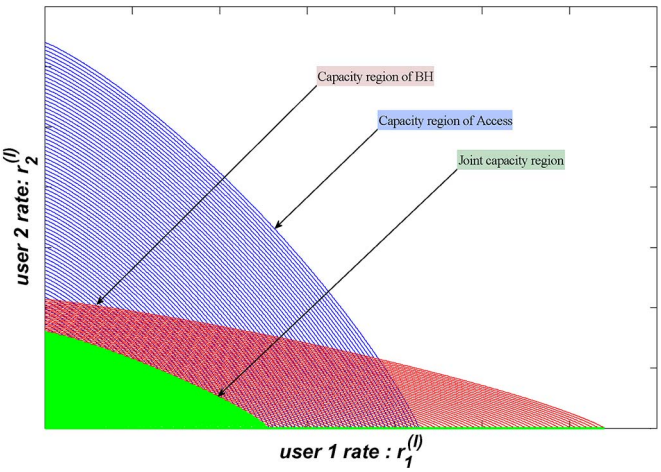

Fig. 3. Construction of joint capacity region for small cells.

In the joint optimization problem (in either links) for a given weight vector, the optimal solution lies on the boundary of the capacity region to guarantee optimality. According to Theorem 1, there should be an optimal α where the solutions of both subproblems converge. Intuitively, this unified solution should be located on the boundary of the capacity regions of both links. In other words, the boundaries of both links will intersect at the optimal α. Consequently, the outcome boundary of the joint capacity region can be achieved by tracing intersection points for different values of α. Fig. 2 shows the boundaries of the capacity region across both links for a two-user scenario. As it can be seen, the intersection points of the corresponding boundaries (with the same values of α) form the boundary of the joint capacity region in thick solid line. Fig. 3 shows the capacity regions of the same scenario for a uniform quantized set of α (with 100 levels). As shown, the joint capacity region can be constructed as the union of capacity regions for different values of α.

5) Subgradient Method With Small-Cell Grouping: As shown, the optimality conditions for both subproblems A.1 and A.2 follow a generic optimality conditions according to (17) and (19) per link where tuning parameter μ_n in conjunction

with duration of transmission balance the operation across both links. In other words, the newly emerged tuning parameters act as a priority factor in the power and time-share allocation similar to weighting parameters w_n. As a result, there is a direct interaction between the resource allocation and rate balancing between the BH and the access links. Therefore, each allocation policy introduces a new set of balancing parameters whereby each tuning set will directly affect the resource allocation.

Remark 2: In practice, as the resources are discrete (in the form of subchannels), the resource allocation is quite sensitive to the tuning of balancing parameters. Therefore, to come up with efficient and practical resource allocation strategies, some important observations should be taken into account.

- As the rate balancing governs the coupling between the links, it requires a slower time-scale compared with the power and time-share update to ensure stability and convergence.
- The resource allocation is sensitive to the relative variations of tuning parameters rather than their absolute values.

The given remark leads us toward a coherent solution to update the tuning variables:

Small-cell grouping (subalgorithm): Per iteration of optimization, small cells can be classified based on satisfying or violating the balancing constraint that is associated with the problem. Considering the PPS model, each user is exclusively attached to a single small cell. Therefore, the constraints are equivalent to per-user constraints. As an example, in subproblem A.1, users that satisfy inequality (10) per iteration can be grouped as (over)satisfied users, whereas others form the unsatisfied group. Intuitively, the tuning parameters of (over)satisfied users (or corresponding cells) can remain unchanged while the unsatisfied users are jointly updated with a similar increment. This method can avoid the unnecessary swapping of resources (subchannels) inside (over)satisfied or unsatisfied groups and gradually increases the priority factor of the unsatisfied users. As a result, the resources are more steadily transferred from the (over)satisfied group to the unsatisfied group until the total balance is achieved for all the users (cells).

Based on this subalgorithm, we can summarize the outcome joint resource allocation algorithm for problem A as follows.

Algorithm 1: Optimal Rate Balancing

1. **Initialization**: choose a sufficiently small uniform set for $\boldsymbol{\mu}^{(t=0)}$ and $\boldsymbol{\alpha}^{(t=0)}$.
2. **Rate update**: given $\boldsymbol{\mu}^{(t)}$, calculate $\boldsymbol{r}^{(1)(t+1)}$ and $\boldsymbol{r}^{(2)(t+1)}$ according to (17), (19), and (7).
3. **Transmission duration update**: update the transmission duration parameter according to (24) and (25).
4. **Small-cell grouping**: classify the small cells (users) according to either (10) or (12).
5. **Balancing update**: update the balancing parameters $\boldsymbol{\mu}^{(t+1)}$ according to small-cell grouping subalgorithm.
6. **Termination**: stop when the total balance is achieved between all BH and access links otherwise go to 2.

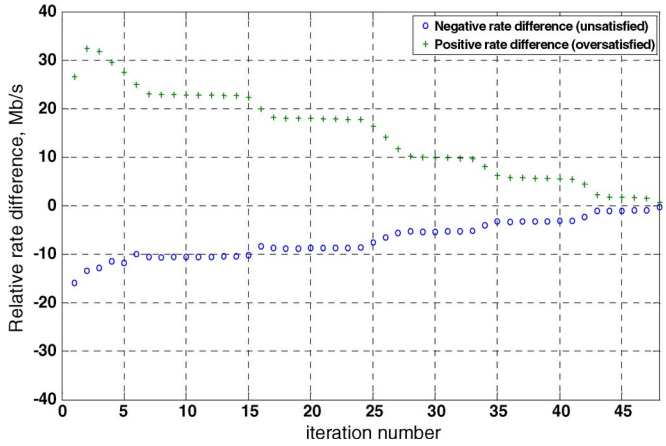

Fig. 4. Evolution of relative rate difference between access and BH links based on Algorithm 1.

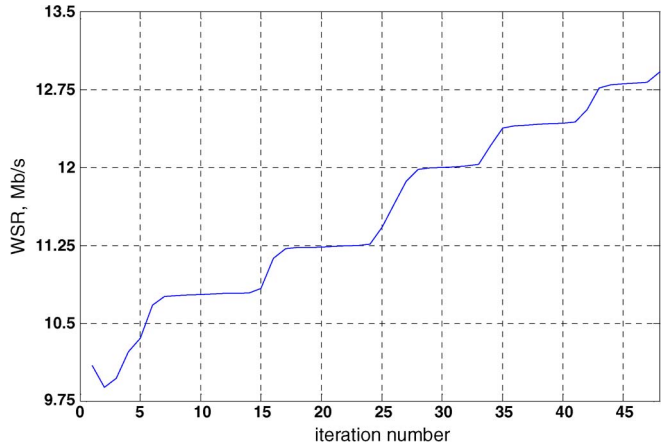

Fig. 5. Evolution of WSR in the network based on Algorithm 1.

Remark 3: In Algorithm 1, the rate vectors of both links are updated based on the balancing parameters of the previous iteration. In practice, if the variations of balancing parameters are small compared with the weighting factors of users w_n, it is possible to fix the priority factor on the bottleneck link according to either (15) or (16). This approximation, although suboptimal, simplifies the outcome algorithm by just tuning the rates on one link. This form is termed as *suboptimal rate balancing* in Section IV.

Figs. 4 and 5 show a typical evolution of relative rate difference (imbalance) and WSR according to Algorithm 1 in a two-user (two-cell) case. As shown, by each iteration, the subchannels are transferred from the (over)satisfied user with positive rate difference to the unsatisfied user with negative rate difference, and the transmission durations are adjusted accordingly. This procedure iteratively continues until the total balance is reached in both cells.

B. Multiaccess Small Cells (MAS)

Here, the developed model of PPS is further extended for the MAS where each cell is associated with multiple users.

1) Problem Formulation: In the case of MAS, the resource allocation in the BH link is not exclusive to one user per small cell. In other words, the achieved rate on the BH can be partitioned among all associated users of that small cell. To enable effective partitioning of resources in the BH, we define new auxiliary variables $\tilde{r}_n^{(1)}$. Unlike $r_n^{(1)}$ in the PPS case, $\tilde{r}_n^{(1)}$ are not directly defined based on the resource allocation policy as in (7), but they rather act as slicing variables to partition the resources allocated to a small-cell station among its associated users [21]. This leads us to a new problem definition for the MAS scenario.

Problem B (MAS Without Time Sharing): This problem is similar to problem A. However, (7) should be separately defined for the BH and access links as follows:

$$Q_m = \sum_{k\in\boldsymbol{K}} \tau_{m,k}^{(1)} \log_2\left(1+\rho_{m,k}^{(1)} p_{m,k}^{(1)}\right) \tag{29}$$

$$r_n^{(2)} = \sum_{k\in\boldsymbol{K}} \tau_{n,k}^{(2)} \log_2\left(1+\rho_{n,k}^{(2)} p_{n,k}^{(2)}\right). \tag{30}$$

Here, Q_m represents the rate associated to the small cell m in the BH.

2) Relaxation and Decomposition: Similar to problem A, the exclusivity constraint can be relaxed through time sharing as in (8). Moreover, the minimization problem is decomposed into the following subproblems:

Subproblem B.1 (Bottleneck on the BH): This is similar to subproblem A.1 where (29) and (30) replace (7).

Subproblem B.2 (Bottleneck on the Access): This is similar to subproblem A.2 where (29) and (30) replace (7).

Note that auxiliary variables $\tilde{r}_n^{(1)}$ substitute $r_n^{(1)}$ in both subproblems.

In subproblem B.1, by forming the partial Lagrangian subject to the power and time-share constraints, we have

$$L(\boldsymbol{p},\boldsymbol{\tau},\alpha,\boldsymbol{\gamma}) = \sum_{n\in\boldsymbol{N}} w_n \alpha \tilde{r}_n^{(1)} - \sum_{n\in\boldsymbol{N}} \gamma_n \left(\alpha\tilde{r}_n^{(1)} - (1-\alpha) r_n^{(2)}\right). \tag{31}$$

As mentioned, $\tilde{r}_n^{(1)}$ are not directly related to the resource allocation policy. As a result, Algorithm 1 cannot be immediately employed based on subproblem B.1.

3) Resource Slicing: However, the following relationship should be satisfied between Q_m and auxiliary variables $\tilde{r}_n^{(1)}$ as the feasibility constraint for the slicing variables:

$$\sum_{n\in\boldsymbol{N}_m} \tilde{r}_n^{(1)} \le Q_m. \tag{32}$$

This implies that at the optimal point of operation with full balance, equality is achieved, i.e., $\sum_{n\in\boldsymbol{N}_m} \tilde{r}_n^{(1)} = Q_m$. Therefore, a new partial Lagrangian can be formulated taking into account the above as an additional constraint, i.e.,

$$L(\boldsymbol{p},\boldsymbol{\tau},\alpha,\boldsymbol{\gamma},\boldsymbol{\beta}) = \sum_{n\in\boldsymbol{N}} w_n \alpha \tilde{r}_n^{(1)} - \sum_{n\in\boldsymbol{N}} \gamma_n \left(\alpha\tilde{r}_n^{(1)} - (1-\alpha) r_n^{(2)}\right) - \sum_{m\in\boldsymbol{M}} \beta_m \left(\sum_{n\in\boldsymbol{N}_m} \tilde{r}_n^{(1)} - Q_m\right). \tag{33}$$

As it can be seen, two sets of price variables emerge in this subproblem; γ_n are price parameters to balance the achievable rate per user on access link to the sliced quota of the BH, determined by $\tilde{r}_n^{(1)}$, whereas β_m are dual variables in charge of matching the BH capacity to the aggregate sliced quota per small cell.

The dual form of subproblem B.2 results in different price variables. In particular, this form can be more favorable due to the direct relation between the access rates $r_n^{(2)}$ and the allocation policy according to (30), which is similar to (7) in problem A. Here, the equality relationship between Q_m and the auxiliary variables $\tilde{r}_n^{(1)}$ at full balance can be utilized to better decouple the problem. Therefore, we can initially relax constraint (12) according to (32) to the following set of inequalities:

$$\sum_{n\in\boldsymbol{N}_m} (1-\alpha) r_n^{(2)} \le \alpha Q_m \qquad \forall\, m \in \boldsymbol{M}. \tag{34}$$

As a result, the problem maps to the following partial Lagrangian subject to the power and time-share constraints:

$$L(\boldsymbol{p},\boldsymbol{\tau},\alpha,\boldsymbol{\theta}) = \sum_{n\in\boldsymbol{N}} w_n (1-\alpha) r_n^{(2)} - \sum_{m\in\boldsymbol{M}} \theta_m \left(\sum_{n\in\boldsymbol{N}_m} (1-\alpha) r_n^{(2)} - \alpha Q_m\right). \tag{35}$$

As it can be seen, the given closed form decouples the problem in the BH and access links. Here, Corollaries 1 and 2 can be efficiently utilized for the power and time-share allocation with the following slight modification:

$$\phi_n^{(l)} = \begin{cases} \theta_m, & l=1 \\ (w_n - \theta_m), & l=2 \end{cases} \tag{36}$$

where θ_m is the tuning set similar to μ_n that monitors the rate balancing across the links. Unlike μ_n, the new tuning set θ_m recovers the allocated rates to the BH of cells Q_m rather than users on the first link. The rate-balancing algorithm (see Algorithm 1) can still be employed to calculate the optimal solution of problem B (via B.2) provided that $r_n^{(1)}$ and $r_n^{(2)}$ are substituted with Q_m and $\sum_{n\in\boldsymbol{N}_m} r_n^{(2)}$, respectively, in Steps 3–5. Note that auxiliary variables can be immediately recovered from (12) based on $r_n^{*(2)}$ and α^* assuming the equality holds. This would satisfy already relaxed constraint (12) and also (32).

4) Complexity Analysis: Having the generic picture of the proposed algorithm based on MAS, here, we explore the complexity order of the algorithm. As a recap of notations, N, K, M stand for the total number of users, subchannels, and small cells, respectively. We also assume that T iterations will be required until Algorithm 1 converges. Each iteration of algorithm involves time-share and power allocation updates according to (17) and (19). Assuming a single iteration for time-share update and waterfilling-based power allocation, the complexity of these subproblems will be MK and MK^2 for the BH and NK and NK^2 for access links. The duration of transmission update requires M executions. Similarly, small-cell grouping requires M executions. There will be one iteration update for joint tuning of parameters θ_m according to the proposed algorithm. Therefore, summing up different steps, the whole complexity of the algorithm will be on the order of $[(M+N)K^2 + (M+N)K + 2M + 1]T$. We will provide more details on the convergence of algorithm and the required iterations T for convergence in Section IV.

C. Special Case

As a special case, a high-capacity BH can be considered where a line-of-sight (LoS) symmetric connection is present between the L-GW of Cloud-RAN and its associated small cells. This assumption can simplify the rate-balancing procedure as we will subsequently illustrate in problem C.

Problem C (MAS With LoS BH): Maximize the WSR on the access link subject to resource allocation constraints, i.e.,

$$\max_{\mathbf{p},\boldsymbol{\tau},\alpha} \sum_{n\in\boldsymbol{N}} w_n(1-\alpha)r_n^{(2)} \text{ s.t.} \tag{37}$$

$$\sum_{n\in\boldsymbol{N}} (1-\alpha)r_n^{(2)} \le \alpha Q_{\max}. \tag{38}$$

Here, $Q_{\max}$ shows the total achievable rate of the BH link that should be sliced among different small cells based on their access requirements.

Considering partial Lagrangian (subject to the resource allocation constraints), we have

$$L(\boldsymbol{p},\boldsymbol{\tau},\alpha,\psi) = \sum_{n\in\boldsymbol{N}} w_n(1-\alpha)r_n^{(2)} - \psi\left(\sum_{n\in\boldsymbol{N}} (1-\alpha)r_n^{(2)} - \alpha Q_{\max}\right). \tag{39}$$

As it can be seen, here, a single tuning parameter ψ controls the balancing of rates among the small cells. Therefore, to meet the balance, the transmission duration parameter can be directly tuned as follows:

$$\alpha^* = \frac{\sum_{n\in\boldsymbol{N}} r_n^{(2)}}{Q_{\max} + \sum_{n\in\boldsymbol{N}} r_n^{(2)}} \tag{40}$$

where $r_n^{(2)}$ can be directly calculated based on their respective weights of $w_n \cong (w_n - \psi)$ assuming $w_n \gg \psi$.

TABLE I
SIMULATION PARAMETERS FOR CASE STUDIES IN A [27]

Small cell	BH	Access
System bandwidth	5 MHz	
Carrier frequency	2.0 GHz	
Total TX power	37 dBm	
Distance-dependent path loss	Hotzone Model 1 (Outdoor Macro)	Hotzone Model 1 (Outdoor Pico)
Shadowing	Lognormal, zero mean, 8 dB standard deviation	
Fast fading channel	Rayleigh block fading	
UE / small cell dropping	Variable based on scenario	

IV. SIMULATION STUDIES

Here, the developed concepts and algorithms are evaluated by means of extensive simulations.

A. Case Studies With Orthogonal Allocation Between Small Cells

Here, we consider topologies with orthogonal allocation between small cells. This enables us to focus on the efficiency proposed algorithms in different cases. We run Monte Carlo simulations based on the parameters consistent with 3GPP specifications [27] to examine the effect of bottleneck links on the considered algorithms. Furthermore, we assume normalized weights across the case studies here. Some common simulation parameters of case studies in A are presented in Table I.

1) PPS Scenario: Initially, we assume that the PPS model is employed where each small cell is associated to a single user in its vicinity. The users are assumed to be almost stationary so the path loss and shadowing values are fixed during the simulation duration. The samples are averaged over 2000 independent snapshots. Furthermore, $\boldsymbol{\alpha}^{(t=0)}$ and $\boldsymbol{\mu}^{(t=0)}$ are set to 0.5 and 0.001, respectively. For the symmetric (Sym) cases, the small cells are located at the perimeter of a circle with 200-m distance from the L-GW; users are located at equal distance from their respective small cells (500 m). In asymmetric (Asym) scenarios, one small cell is shifted to the distance of 900 m; however, the associated user keeps the same distance as that before it.

At first, a two-user case is considered under symmetric and asymmetric bottleneck conditions. Fig. 6 shows the simulation result for this case. As it can be seen, rate balancing provides considerable improvement in the efficiency of the resource allocation compared with the cases with independent resource allocation across BH and access links. In the case of the symmetric bottleneck condition, the suboptimal algorithm provides near-optimal performance, whereas in the asymmetric bottleneck condition, the gap to the optimal solution increases. This is consistent with our theoretical inspection in previous sections.

Fig. 7 shows the simulation result for a three-user case. Here, a similar trend is observable in both bottleneck conditions.

2) MAS Scenario: In the second scenario, the MAS model is utilized where each small cell is associated with multiple users. Here, the users are assumed to be mobile. As a result, the shadowing and path-loss values change across different

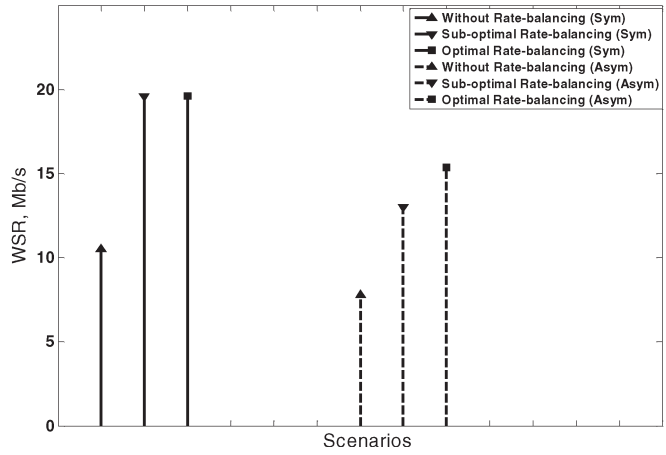

Fig. 6. Evaluating algorithmic efficiency of rate balancing for A-1 ($M = 2$).

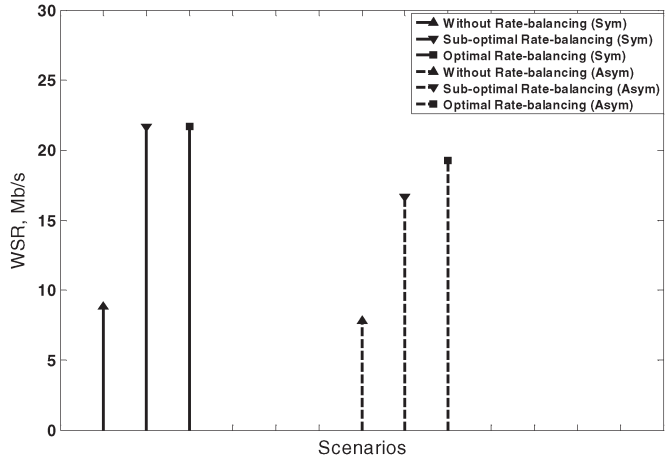

Fig. 7. Evaluating algorithmic efficiency of rate balancing for A-1 ($M = 3$).

snapshots. The samples are averaged over 2000 independent snapshots. $\boldsymbol{\alpha}^{(t=0)}$ and $\boldsymbol{\mu}^{(t=0)}$ are set as the PPS case. The small cells are located at the perimeter of a circle with 500-m distance from the L-GW, and the users are randomly but uniformly distributed in the vicinity of each small cell with a similar maximum distance from their respective small-cell station.

Fig. 8 shows the simulation result for different numbers of users in this scenario ($M = 2$). As expected, increasing the number of users enhances the WSR in all cases due to better multiuser diversity. Rate balancing provides significant improvement in the efficiency of resource allocation compared with the benchmark algorithms. This scenario indicates that in a realistic case where a mixture of symmetric and asymmetric bottleneck conditions exists, the performance of the suboptimal algorithm is quite close to the optimal solution. As a result, the suboptimal algorithm can be considered as an alternative to the optimal algorithm.

B. Case Studies With Spectrum Reuse Between Small Cells

In these case studies, we evaluate the performance of the proposed algorithm in a more detailed system-level topology that is again consistent with 3GPP specifications [27]. In particular,

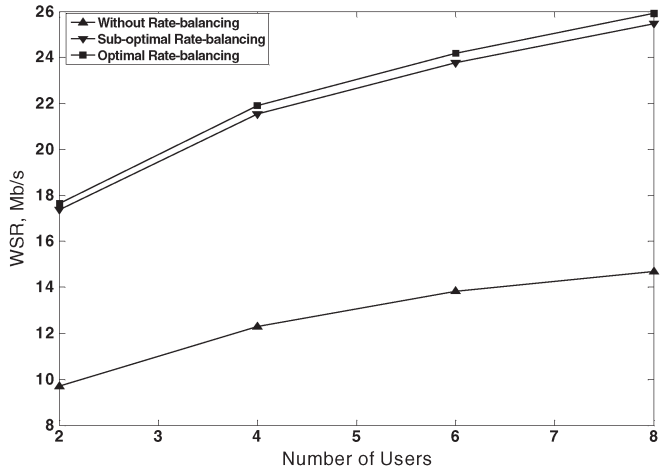

Fig. 8. Evaluating algorithmic efficiency of rate balancing for A-2 ($M = 2$).

TABLE II
SIMULATION PARAMETERS FOR CASE STUDIES IN B [27]

Small cell	BH	Access
System bandwidth	10 MHz	
Carrier frequency	2.0 GHz	
Total TX power	24 dBm	
Distance-dependent path loss	Macro-to-Relay (Outdoor)	Relay-to-UE (Outdoor)
Shadowing	Lognormal, zero mean, 10 dB standard deviation	
Antenna gain / connector loss	5 dBi (omni) / 0 dBi	
Fast fading channel	Rayleigh block fading	
UE dropping	8 users per cell (32 in total)	
Radius for small cell dropping in a cluster	90 m	
Minimum small cell station to UE distance	5 m	
Minimum small cell station to base station distance	20 m	

we examine the impact of different interference coordination strategies from dynamic clustering to full reuse. The common simulation parameters of case studies in B are presented in Table II. Here, we have an outdoor random cell and user deployment in a cluster in line with scenario 2 in [28]. In particular, four small-cell stations are randomly dropped in a cluster area (ring) with minimum distances, as outlined in Table II. Concerning intracell scheduling, proportional fair scheduling is used for a multichannel system per small cell to provide a fair allocation of resources between multiple users. Therefore, user weights are tuned based on this algorithm and are normalized to total weights per cell for a fair comparison. The samples are averaged over 5000 independent snapshots. Furthermore, $\boldsymbol{\alpha}^{(t=0)}$ and $\boldsymbol{\mu}^{(t=0)}$ are set as case studies in A.

1) Dynamic Clustering: The proposed solution is initially studied under dynamic clustering as interference coordination [24]. Here, we compare three RRM schemes as follows:

- without rate balancing, where independent resource scheduling happens on the BH and access links;

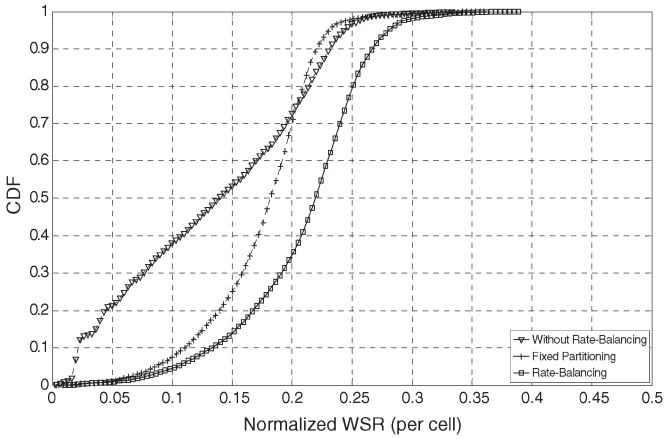

Fig. 9. CDF of normalized WSR for case study B-1.

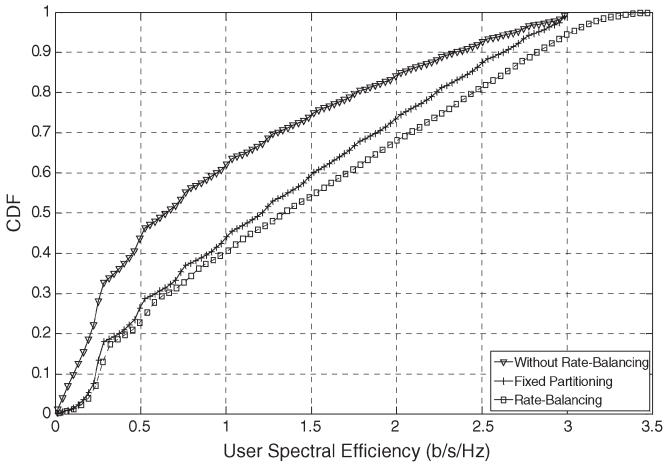

Fig. 10. CDF of user spectral efficiency for B-1.

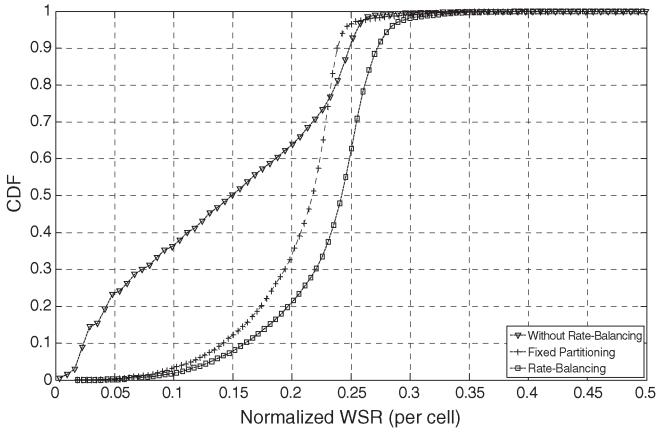

Fig. 11. CDF of normalized WSR for case study B-2.

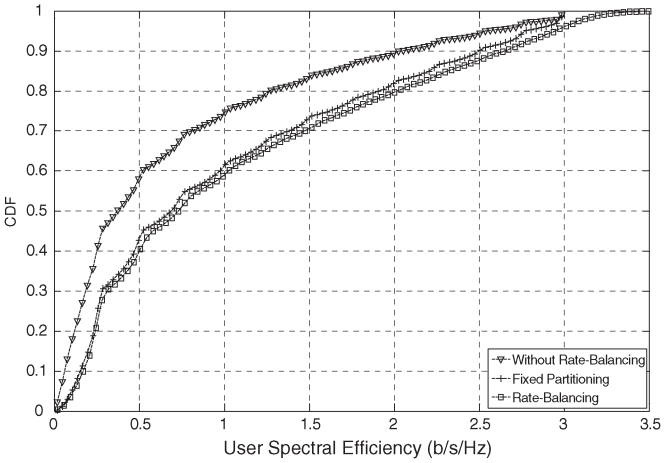

Fig. 12. CDF of user spectral efficiency for B-2.

- fixed partitioning, where the BH resource is equally partitioned between the small cells in a static manner;
- rate balancing, where we employ the proposed scheme combined with dynamic clustering.

We consider two performance indicators, cumulative distribution function (CDF) of normalized WSR (per cell) and cdf of user spectral efficiency in bits per second per Hertz.

As it can be seen in Figs. 9 and 10, the proposed rate-balancing scheme can provide significant improvement in both user spectral efficiency and WSR (per cell) in this case study for both performance indicators. It is also interesting to note that fixed partitioning on the BH can be generally a better strategy compared with independent but dynamic RRM in BH and access links considering the results without rate balancing. However, fixed partitioning can be an expensive solution for small-cell operation as it yields low BH utilization when it comes to small cells in low-load scenarios.

2) Reuse One: In this case study, we apply the rate-balancing mechanism in reuse one without any active interference coordination. The results are compared for similar RRM strategies and performance indicators as B-1.

Figs. 11 and 12 show simulations results of this case study. As can be seen, a similar trend as B-1 is observable for the performance indicators, in particular, for cdf of WSR (per cell). It is worth noting that the in reuse one case, the optimization problem and related capacity regions on BH and access links are not generally convex. As a result, the duality gap of proposed solution will be nonzero. Nevertheless, the system performance can still benefit from rate balancing and joint optimization, as is evident from the results. However, the major gain of the algorithm is achieved when rate-balancing strategies are employed in conjunction with interference coordination algorithms such as dynamic clustering.

3) Analysis on Evolution of α and Convergence Rate: Here, we provide a more detailed analysis on the evolution of α as the adjusting parameter for the duration of phases in the dynamic clustering case. Furthermore, we examine the convergence rate of the algorithm across different snapshots for the same case.

Fig. 13 shows the histogram of the optimal value of α across different snapshots. As it can be seen, the value is dynamically adjusted based on the instantaneous channel conditions in both links across different snapshots.

Fig. 14 shows the histogram of the required number of iterations (for convergence) across different snapshots. As shown, the algorithm can converge in all snapshots with fewer than 250 iterations for T. Specifically, the majority of snapshots will converge in around 100 iterations.

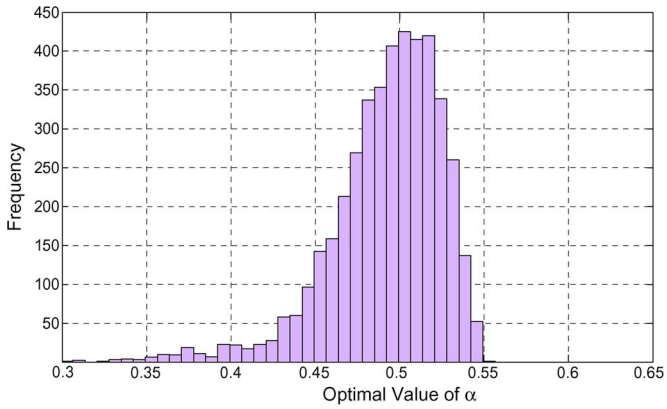

Fig. 13. Histogram of optimal value of α.

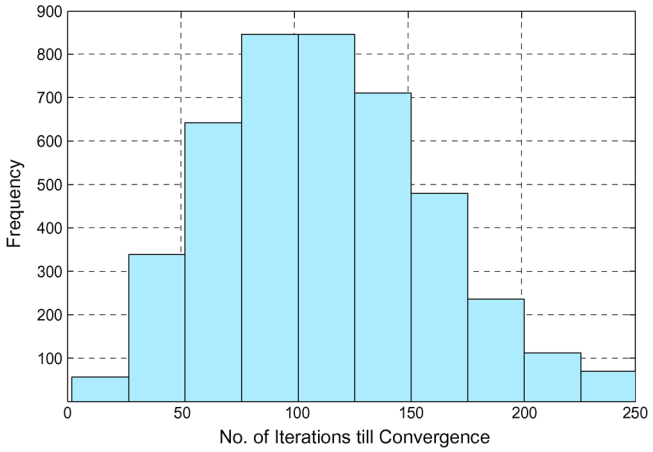

Fig. 14. Histogram of required number of iterations.

V. Conclusion

In this paper, the problem of joint resource allocation between BH and access links in TDD has been addressed for dense small-cell networks. The problem was mathematically decomposed into per-link subchannel and power allocation where a set of rate-balancing parameters combined with a phase duration parameter governed the coupling among the links. Moreover, novel algorithms were derived for rate balancing by employing the concepts of small-cell grouping and resource slicing. Finally, the efficiencies of proposed concepts and algorithms were evaluated by system-level simulations. As shown, joint optimization with rate balancing could provide significant improvement over independent resource allocation across BH and access links, as well as fixed BH partitioning strategies. In particular, the proposed algorithm could deliver significant performance improvement in conjunction with efficient interference coordination strategies such as dynamic clustering.

Acknowledgment

The authors would like to thank Dr. R. Hoshyar for the contributions in line with the earlier work [23].

References

[1] Requirements, Evaluation Criteria, and Submission Templates for the Development of IMT-Advanced, ITU-R, Accessed May 2014. [Online]. Available: http://www.itu.int/md/R07-SG05-C-0068/

[2] *Evolved Universal Terrestrial Radio Access (E-UTRA) and Evolved Universal Terrestrial Radio Access Network (E-UTRAN), Overall Description*, Third-Generation Partnership Project TS 36.300, Mar. 2014.

[3] *Further Advancements for E-UTRA Physical Layer Aspects*, Third-Generation Partnership Project TR 36.814, Mar. 2010.

[4] A. Damnjanovic *et al.*, "A survey on 3GPP heterogeneous networks," *IEEE Wireless Commun.*, vol. 18, no. 3, pp. 10–21, Jun. 2011.

[5] H. Raza, "A brief survey of radio access network backhaul evolution: Part II," *IEEE Commun. Mag.*, vol. 51, no. 5, pp. 170–177, May 2013.

[6] C. Hoymann, *et al.*, "Relaying operation in 3GPP LTE: Challenges and solutions," *IEEE Commun. Mag.*, vol. 50, no. 2, pp. 156–162, Feb. 2012.

[7] C. Hoymann, D. Larsson, H. Koorapaty, and J.-F. Cheng, "A lean carrier for LTE," *IEEE Commun. Mag.*, vol. 51, no. 2, pp. 74–80, Feb. 2013.

[8] H. Ishii, Y. Kishiyama, and H. Takahashi, "A novel architecture for LTE-B: C-plane/U-plane split and Phantom Cell concept," in *Proc. IEEE Globecom Workshops*, 2012, pp. 624–630.

[9] G. Song and Y. Li, "Cross-layer optimization for OFDM wireless networks—Part I: Theoretical framework," *IEEE Trans. Wireless Commun.*, vol. 4, no. 2, pp. 614–624, Mar. 2005.

[10] G. Song and Y. Li, "Cross-layer optimization for OFDM wireless networks—Part II: Algorithm development," *IEEE Trans. Wireless Commun.*, vol. 4, no. 2, pp. 625–634, Mar. 2005.

[11] 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.

[12] L. Huang, M. Rong, L. Wang, Y. Xue, and E. Schulz, "Resource allocation for OFDMA based relay enhanced cellular networks," in *Proc. IEEE VTC*, 2007, pp. 3160–3164.

[13] Z. Zhao, J. Wang, S. Redana, and B. Raaf, "Downlink resource allocation for LTE-Advanced networks with Type1 relay nodes," in *Proc. IEEE VTC*, 2012, pp. 1–5.

[14] D. W. Kifle, O. Bulakci, A. B. Saleh, S. Redana, and F. Granelli, "Joint backhaul co-scheduling and relay cell extension in LTE-advanced networks uplink performance evaluation," in *Proc. Eur. Wireless*, 2012, pp. 1–8.

[15] 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, Feb. 2007.

[16] O. Oyman, "Opportunistic scheduling and spectrum reuse in relay-based cellular networks," *IEEE Trans. Wireless Commun.*, vol. 9, no. 3, pp. 1074–1085, Mar. 2010.

[17] Q. Li, R. Q. Hu, Y. Qian, and G. Wu, "Intracell cooperation and resource allocation in a heterogeneous network with relays," *IEEE Trans. Veh. Technol.*, vol. 62, no. 4, pp. 1770–1784, May 2013.

[18] R. Madan, *et al.*, "Cell association and interference coordination in heterogeneous LTE-A cellular networks," *IEEE J. Sel. Areas Commun.*, vol. 28, no. 9, pp. 1479–1489, Dec. 2010.

[19] S. Sadr, A. Anpalagan, and K. Raahemifar, "Radio resource allocation algorithms for the downlink of multiuser OFDM communication systems," *IEEE Commun. Surveys Tuts.*, vol. 11, no. 3, pp. 92–106, 3rd Quart. 2009.

[20] M. Salem *et al.*, "An overview of radio resource management in relay-enhanced OFDMA-based networks," *IEEE Commun. Surveys Tuts.*, vol. 12, no. 3, pp. 422–438, 3rd Quart. 2010.

[21] D. P. Palomar and M. Chiang, "A tutorial on decomposition methods for network utility maximization," *IEEE J. Sel. Areas Commun.*, vol. 24, no. 8, pp. 1439–1451, Aug. 2006.

[22] D. P. Palomar and M. Chiang, "Alternative distributed algorithms for network utility maximization: Framework and applications," *IEEE Trans. Autom. Control*, vol. 52, no. 12, pp. 2254–2269, Dec. 2007.

[23] R. Hoshyar, M. Shariat, and R. Tafazolli, "Subcarrier and power allocation with multiple power constraints in OFDMA systems," *IEEE Commun. Lett.*, vol. 14, no. 7, pp. 644–646, Jul. 2010.

[24] E. Pateromichelakis, M. Shariat, A. Quddus, M. Dianati, and R. Tafazolli, "Dynamic clustering framework for multi-cell scheduling in dense small cell networks," *IEEE Commun. Lett.*, vol. 17, no. 9, pp. 1802–1805, Sep. 2013.

[25] EU FP7 project iJOIN, "iJOIN | Interworking and JOINt Design of an Open Access and Backhaul Network Architecture for Small Cells Based on Cloud Networks," Accessed May 2014. [Online]. Available: http://www.ict-ijoin.eu

[26] P. Rost *et al.*, "Cloud technologies for flexible 5G radio access networks," *IEEE Commun. Mag.*, vol. 52, no. 5, pp. 68–72, May 2014.
[27] *Evolved Universal Terrestrial Radio Access (E-UTRA): Further Advancements for E-UTRA Physical Layer Aspects*, Third-Generation Partnership Project TR 36.814, Mar. 2010.
[28] *Small Cell Enhancements for E-UTRA and E-UTRAN-Physical Layer Aspects*, Third-Generation Partnership Project TR 36. 872, Dec. 2013.

Mehrdad Shariat received the B.Sc. degree in telecommunications engineering from Iran University of Science and Technology, Tehran, Iran, in 2005 and the Ph.D. degree in mobile communications from the University of Surrey, Guildford, U.K., in 2010.

Since 2010, he has been a Research Fellow with the Institute for Communication Systems, home of the 5G Innovation Centre, University of Surrey. He has been involved in several U.K. and EU cofunded projects, including Mobile VCE (Core 4), BeFEMTO, iJOIN, and MiWaveS, as well as collaborative industry projects on small-cell backhauling. His research interests include cross-layer optimization, radio resource management, and packet scheduling for backhaul and access in cellular and mesh networks.

Emmanouil Pateromichelakis received the Diploma degree in information and communication systems engineering from the University of the Aegean, Samos, Greece, in 2008 and the M.Sc. and Ph.D. degrees in mobile communications from the University of Surrey, Guildford, U.K., in 2009 and 2013, respectively.

Since then, he has been a Research Fellow with the Institute for Communication Systems, home of the 5G Innovation Centre, University of Surrey. His research interests include radio resource management, multicell cooperation, and scheduling for orthogonal frequency-division multiple-access cellular/multihop networks.

Atta ul Quddus received the M.Sc. degree in satellite communications and the Ph.D. degree in mobile cellular communications from the University of Surrey, Guildford, U.K., in 2000 and 2005, respectively.

He is currently a Lecturer in wireless communications with the Department of Electronic Engineering, University of Surrey. During his research career, he has led several national and international research projects that also contributed toward Third-Generation Partnership Project standardization. His current research interests include machine-type communication, cloud radio access networks, and device-to-device communication.

Dr. Quddus won the Centre for Communications Systems Research Excellence Prize sponsored by Vodafone for his research on adaptive filtering algorithms in 2004.

Rahim Tafazolli received the Ph.D. degree in electrical engineering from the University of Surrey, Guildford, U.K.

He is currently the Director of the Institute for Communication Systems and the 5G Innovation Center, Faculty of Engineering and Physical Sciences, University of Surrey, Guildford, U.K. He is also the Chair of the European Union NetWorks Technology Platform Expert Group and a Board Member of the UK Future Internet Strategy Group. He is the author of more than 500 research papers in refereed journals and international conferences and has been an Invited Speaker. He is also the Editor of two books on Technologies for Wireless Future published by Wiley (vol. 1 in 2004 and vol. 2 in 2006). He was appointed as a Fellow of the Wireless World Research Forum in April 2011 in recognition of his personal contribution to the wireless world and for heading one of Europe's leading research groups.