Joint Resource Allocation and User Association for Heterogeneous Cloud Radio Access Networks

Ying Loong Lee*, Li-Chun Wang†, Teong Chee Chuah*, Jonathan Loo‡
*Faculty of Engineering, Multimedia University, 63100 Cyberjaya, Selangor, Malaysia
Email: lee.ying.loong12@student.mmu.edu.my, tcchuah@mmu.edu.my
†Department of Electrical and Computer Engineering, National Chiao Tung University, Hsinchu 300, Taiwan R.O.C.
Email: lichun@g2.nctu.edu.tw
‡School of Science and Technology, Middlesex University, London NW4 4BT, United Kingdom
Email: J.Loo@mdx.ac.uk

***Abstract*—Cloud radio access networks (C-RANs) have been regarded as a promising architecture for energy-efficient fifth generation systems. In this paper, a new joint remote radio head (RRH) activation, user-RRH pairing and resource allocation strategy is proposed for heterogeneous C-RANs (H-CRANs). We first formulate an optimization problem to maximize the energy efficiency of H-CRANs. Then, a low-complexity suboptimal solution is developed. Our proposed mechanism consists of three key procedures: 1) RRH activation is performed based on greedy RRH selection; 2) user-RRH pairing is performed based on the channel quality; 3) the resource allocation problem is solved by dual decomposition. Simulation results show that the proposed strategy can improve energy efficiency significantly.**

***Keywords*-H-CRAN; RRH activation; user association; resource allocation; energy efficiency**

I. Introduction

The fifth generation (5G) mobile cellular systems are expected to provide gigabit data rates to mobile users for broadband applications, which requires much higher capacity compared to the fourth generation (4G) systems. However, achieving such a high capacity requires high power consumption by both base stations and user equipment for the downlink and uplink, respectively. Thus, energy efficiency becomes the key design goal for 5G systems.

To meet the energy efficiency requirement, a new architecture known as cloud radio access network (C-RAN) has been proposed [1]. A C-RAN consists of a set of remote radio heads (RRHs) and a centralized baseband unit (BBU) pool. These RRHs are connected to the BBU pool via the *fronthaul* links. The RRHs generally serves as radio frequency (RF) transmitters and receivers, which only performs basic RF functionalities, whereas the BBU pool performs baseband signal processing and upper-layer functionalities [2], [3]. When the C-RAN coexists with the macrocell base station (MBS), it is called heterogeneous C-RAN (H-CRAN) as shown in Fig. 1. The main feature of the H-CRAN is that control signals are transmitted by the MBS to the mobile users, which can facilitate the mobility management of small cell networks.

In the literature, a number of related studies for C-RANs [4]–[11] have been carried out. In [4], the power consumption of C-RANs was minimized via RRH activation and beamforming. A power allocation scheme for multiple input multiple output (MIMO)-based C-RANs was proposed in [5] to maximize energy efficiency. In [6], power allocation and beamforming are jointly investigated for C-RANs. Subchannel allocation is further considered in [7], together with power allocation and beamforming were jointly studied to improve the energy efficiency of C-RANs. In [8], the authors proposed a cross-layer resource allocation, RRH activation and beamforming scheme for C-RANs to minimize power consumption. Optimal resource allocation for C-RANs was studied in [9] under a coordinated multipoint transmission (CoMP) framework with power consumption and fronthaul capacity constraints. RRH activation and user association for improving energy efficiency of C-RANs was investigated in [10]. In [11], a subchannel and power allocation scheme was proposed for maximizing the energy efficiency of H-CRANs.

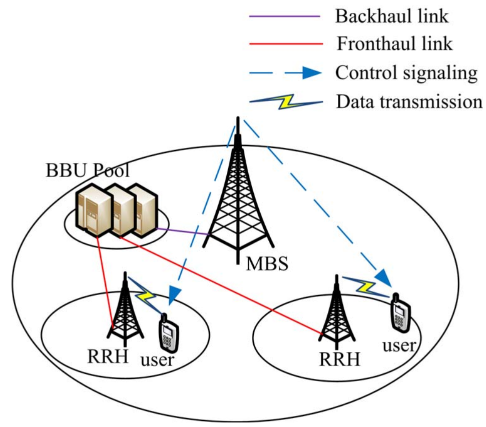

Figure 1. Architecture of H-CRAN.

To the best of our knowledge, joint consideration of RRH activation, user association, and subchannel and power allocation for improving energy efficiency of H-CRANs is not available in the literature. Most of the related studies [4]–[8], [10], [11] have not considered the constraints of the limited capacity of fronthaul links. In fact, high-capacity fronthaul links such as optical fiber may incur high deployment costs especially in ultra-dense small cells. Also, energy efficiency is not investigated in [4], [6], [8] and [9]. In the literature, load-dependent fronthaul power consumption models have not been considered in investigating the energy efficiency performance of H-CRANs. Therefore, we are motivated to study a joint RRH activation, user-RRH pairing and resource allocation to improve the energy efficiency of H-CRANs under constrained fronthaul capacity based on a load-dependent fronthaul power consumption model.

In this paper, we develop a comprehensive mechanism for RRH on/off mechanism, user selection[1] and radio resource allocation for H-CRANs. We focus on the downlink of an H-CRAN whereby the power is mainly consumed by the network infrastructure (e.g. RRHs, BBU pools and fronthaul links). The proposed strategy is formulated as an optimization problem that maximizes energy efficiency of an H-CRAN subject to the limited fronthaul capacity constraint. The problem is then transformed and solved using an iterative algorithm in which RRH activation is performed based on greedy selection and user-RRH pairing is performed based on channel quality. The resource allocation problem is solved by dual decomposition. The proposed strategy is evaluated and compared with several baseline schemes in terms of energy efficiency.

The remainder of this paper is organized as follows. Section II introduces the system model and presents the problem formulation. The solution algorithm is proposed in Section III. In Section V, performance evaluation of the proposed strategy is presented. Finally, Section VI concludes this paper.

II. System Model and Problem Formulation

Fig. 1 shows the system model of an H-CRAN [3] consisting of a BBU pool connected to an MBS and a set of RRHs (small cells). In the considered architecture, an MBS transmits control signaling to CRAN, whereas data transmission is performed by the RRHs. Denote $\mathcal{S}$, $\mathcal{U}$ and $\mathcal{K}$ as the sets of RRHs, mobile users and subchannels, respectively. Further, we define a_s as the activation indicator of RRH s, whereby $a_s = 1$ indicates that RRH s is activated and $a_s = 0$ when it is deactivated.

Also, b_{su} is defined as the pairing indicator of RRH s with user u whereby $b_{su} = 1$ if user u is paired with RRH s; otherwise $b_{su} = 0$. For resource allocation among users associated with each RRH, let ω_{ku} be the assignment indicator of subchannel k to user u whereby $\omega_{ku} = 1$ if subchannel k is allocated to user u; otherwise $\omega_{ku} = 0$. Also, p_{sk} is defined as the nonnegative transmission power of RRH s on subchannel k. To ease analysis, the following assumptions are made: 1) The network is perfectly synchronized; 2) All RRHs share the entire channel bandwidth available; 3) Each subchannel experiences flat and slow fading.

The power consumption model of an H-CRAN basically consists of the power consumed by the RRHs, the fronthaul links, the backhaul link and the BBU pool[2]. The power consumption of each RRH is given as [12]:

$$P_{\text{RRH},s} = a_s \left(P_{0,s} + \eta_s \sum_{u\in\mathcal{U}} b_{su} \sum_{k\in\mathcal{K}} \omega_{ku} p_{sk} \right) + (1 - a_s) P_{\text{sleep},s}, \tag{1}$$

where $P_{0,s}$ is the static power consumption of RRH s when it is activated, $P_{\text{sleep},s}$ is the total power consumed by RRH s when it is deactivated and η_s is the slope of the load-dependent power consumption of RRH s.

We adopt a power consumption model in [13] whereby the total power consumed by fronthaul links are proportional to the network traffic carried to their associated RRHs. The power consumption model of a fronthaul link can be expressed as:

$$P_{\text{fh},s} = a_s \left(P_{\text{c},s} + \beta_s R_s \right). \tag{2}$$

In (2), $P_{\text{c},s}$ is the constant power consumption of a fronthaul link which is given as $P_{\text{c},s} = P_{\text{ft},s} + \frac{\tau_s P_{\text{sw},s}}{n_{\text{port},s}}$ [14]. For RRH s, $P_{\text{ft},s}$, τ_s, $P_{\text{sw},s}$ and $n_{\text{port},s}$ are the power consumed by the fronthaul transceiver, the percentage of the load-independent power consumption of the fronthaul aggregation switch, the maximum power consumption of the switch and the number of ports of the switch, respectively. R_s is the total network traffic of the fronthaul link associated with RRH s and β_s is the power consumed per bit/s by a fronthaul link which is written as $\beta_s = \frac{(1-\tau_s)P_{\text{sw},s}}{n_{\text{port},s} R_{\text{fh},s}}$ [14], where $R_{\text{fh},s}$ is the maximum traffic load that can be carried by the switch in the fronthaul link associated with RRH s, that is, the fronthaul capacity of RRH s.

The power consumption model in [13] can also be adopted for the backhaul link between the BBU pool and the MBS (cf. Fig. 1). However, since the backhaul link only carries control signaling between the BBU pool and the MBS, we assume that the control traffic constantly consumes a fixed amount of the backhaul bandwidth. Thus, the power consumed by the backhaul link can be assumed to be constant.

We express the total power consumption of the H-CRAN as:

$$P = \sum_{s\in\mathcal{S}} \left(P_{\text{RRH},s} + P_{\text{fh},s} \right) + P_{\text{bh}}. \tag{3}$$

The rate utility function of the H-CRAN is written as:

$$R = \sum_{s\in\mathcal{S}} a_s \sum_{u\in\mathcal{U}} b_{su} w_u \sum_{k\in\mathcal{K}} \omega_{ku} R_{sku}, \tag{4}$$

where w_u is the weighting coefficient corresponding to the data rate achievable by user u which can be adjusted to achieve different notions of fairness [15]. R_{sku}, the

[1]In this paper, the terms 'user selection', 'user association' and 'user-RRH pairing' are interchangeable.

[2]We exclude the power consumption of the MBS since the MBS does not take part in data transmission under our system model. Nonetheless, it can easily be included into our power consumption model as the MBS is always activated for control signaling, thereby incurring static power consumption.

data rate achievable by user u associated with RRH s on subchannel k, is expressed as:

$$R_{sku} = B \log_2 \left(1 + \frac{p_{sk} g_{sku}}{\sum_{i \in \mathcal{S} \backslash \{s\}} a_i p_{ik} g_{iku} + N_0} \right), \quad (5)$$

where B is the bandwidth of a subchannel, g_{sku} is the downlink channel gain between RRH s and user u on subchannel k, and N_0 is the additive white Gaussian noise (AWGN) power.

The utility function that corresponds to the energy efficiency of the H-CRAN is defined as:

$$U_{\text{EE}} = \frac{R}{P}. \quad (6)$$

It is noteworthy that energy efficiency is defined as a ratio of the total transmission rate to total power consumption. Thus, by setting $w_u = 1$ for all $u \in \mathcal{U}$, (4) becomes the total transmission rate of the H-CRAN and (6) is equivalent to the energy efficiency of the H-CRAN.

The main objective of this paper is to maximize (6). Thus, the joint RRH activation, user-RRH pairing and resource allocation problem for an H-CRAN can be formulated as follows:

$$\max_{\mathbf{a},\mathbf{b},\boldsymbol{\omega},\mathbf{p}} U_{\text{EE}}(\mathbf{a},\mathbf{b},\boldsymbol{\omega},\mathbf{p}) = \frac{R(\mathbf{a},\mathbf{b},\boldsymbol{\omega},\mathbf{p})}{P(\mathbf{a},\mathbf{b},\boldsymbol{\omega},\mathbf{p})}, \quad (7)$$

subject to:

$$a_s \sum_{u \in \mathcal{U}} b_{su} \sum_{k \in \mathcal{K}} \omega_{ku} p_{sk} \leq P_{\text{max},s} \quad \forall s \in \mathcal{S} \quad (7a)$$

$$a_s \sum_{u \in \mathcal{U}} b_{su} \sum_{k \in \mathcal{K}} \omega_{ku} R_{sku} \leq R_{\text{fh},s} \quad \forall s \in \mathcal{S} \quad (7b)$$

$$a_s b_{su} \sum_{k \in \mathcal{K}} \omega_{ku} R_{sku} \geq R_{\text{min},u} \quad \forall s \in \mathcal{S}, u \in \mathcal{U} \quad (7c)$$

$$\sum_{s \in \mathcal{S}} a_s b_{su} = 1 \quad \forall u \in \mathcal{U} \quad (7d)$$

$$a_s \sum_{u \in \mathcal{U}} b_{su} \omega_{ku} \leq 1 \quad \forall s \in \mathcal{S}, k \in \mathcal{K} \quad (7e)$$

$$p_{sk} \geq 0 \quad \forall s \in \mathcal{S}, k \in \mathcal{K}, \quad (7f)$$

where $\mathbf{a} = [a_1, \ldots, a_{|\mathcal{S}|}]$, $\mathbf{b} = [b_{11}, \ldots, b_{|\mathcal{S}||\mathcal{U}|}]$, $\boldsymbol{\omega} = [\omega_{11}, \ldots, \omega_{|\mathcal{K}||\mathcal{U}|}]$ and $\mathbf{p} = [p_{11}, \ldots, p_{|\mathcal{S}||\mathcal{K}|}]$. Constraint (7a) ensures that the total transmission power of each RRH s does not exceed the maximum allowable transmission power, $P_{\text{max},s}$. Constraint (7b) is the fronthaul capacity constraint whereby the total transmission rate of each RRH s must not exceed the fronthaul capacity, $R_{\text{fh},s}$. Each user u is ensured in constraint (7c) that its minimum bit rate, $R_{\text{min},u}$ is achieved. In constraint (7d), each user is ensured to only associate with one RRH. Constraint (7e) ensures that no two or more users associated with the same RRH will receive the same subchannels. Constraint (7f) is a nonnegative transmission power constraint. In fact, (7) is a nonconvex mixed-integer programming problem, which is generally NP-hard. To solve (7), we show in the next section that it can be solved efficiently using an iterative greedy algorithm.

III. Solution Algorithm

In this paper, we propose a low-complexity suboptimal iterative greedy algorithm, as depicted in Algorithm 1, to efficiently solve (7). In this algorithm, RRH activation is performed based on the greedy approach similar to that in [10] which deactivates the RRH that has the least contribution to the total energy efficiency iteratively; user-RRH pairing is performed based on signal-to-interference-plus-noise ratios (SINRs); subchannel and power allocation is performed based on dual decomposition.

Algorithm 1 Greedy RRH activation and SINR-based user-RRH pairing

1: Set $a_s = 1$ for all $s \in \mathcal{S}$, $U_{\text{old}} = 0$ and $\mathcal{S}_{\text{a}} = \mathcal{S}$.
2: Set $b_{su} = 1$ for all $s \in \mathcal{S}_{\text{a}}$ and $u \in \mathcal{U}$ such that the received wideband SINR from RRH s to user u is the highest.
3: Solve (7) for $\boldsymbol{\omega}$ and $\mathbf{p}$ (See Algorithm 2)
4: Calculate P, R and U_{EE} using (3), (4) and (6), respectively.
5: **while** $U_{\text{EE}} > U_{\text{old}}$ **do**
6: Set $U_{\text{old}} = U_{\text{EE}}$.
7: Evaluate the throughput of RRH s, i.e., R_s and calculate $\frac{R_s}{P}$ for all $s \in \mathcal{S}_{\text{a}}$.
8: Deactivate RRH s such that its corresponding $\frac{R_s}{P}$ is the smallest among all RRHs. Then, set $a_s = 0$ and $\mathcal{S}_{\text{a}} = \mathcal{S}_{\text{a}} \backslash \{s\}$.
9: Set $b_{su} = 1$ for all $s \in \mathcal{S}_{\text{a}}$ and $u \in \mathcal{U}$ such that the received wideband SINR from RRH s to user u is the highest.
10: Solve (7) for $\boldsymbol{\omega}$ and $\mathbf{p}$ (See Algorithm 2).
11: Calculate P, R and U_{EE} using (3), (4) and (6), respectively.
12: **end while**

In Steps 1-4 of Algorithm 1, all RRHs are assumed to be activated, i.e., $a_s = 1$ for all $s \in \mathcal{S}$ and the set of active RRHs, $\mathcal{S}_{\text{a}}$ includes all RRHs. Each user u is paired with RRH s with the largest received wideband SINR. The wideband SINR received by each user from each RRH can be estimated as:

$$\Gamma_{su} = \frac{p_s \bar{g}_{su}}{\sum_{i \in \mathcal{S} \backslash \{s\}} p_i \bar{g}_{iu} + N_0} \quad \forall s \in \mathcal{S}, u \in \mathcal{U}, \quad (8)$$

where p_s is the total power transmitted by RRH s and $\bar{g}_{su}$ is the average channel gain received by user u from RRH s on the entire channel bandwidth. Thus, $b_{su} = 1$ if RRH s provides the largest wideband SINR to user u. This SINR-based user association ensures that the channel quality experienced by the users is higher. Thus, the users can achieve a higher throughput, hence possibly a higher energy efficiency performance. Then, we can proceed to solve (7) for $\boldsymbol{\omega}$ and $\mathbf{p}$. After that, we can calculate the total power consumption P, the weighted sum rate R and the energy efficiency U_{EE} using (3), (4) and (6), respectively.

Next, in Steps 5-11 of Algorithm 1, we first set $U_{\text{old}} = U_{\text{EE}}$. Then, we evaluate the throughput of each RRH and evaluate $\frac{R_s}{P}$ for all $s \in \mathcal{S}_{\text{a}}$ where $R_s = a_s \sum_{u \in \mathcal{U}} b_{su} \sum_{k \in \mathcal{K}} R_{sku}$. Then, RRH s is deactivated such that it corresponds to the smallest $\frac{R_s}{P}$ among all RRHs in $\mathcal{S}_{\text{a}}$. The key idea of this method is to deactivate the RRHs that contribute the least to the energy efficiency, because their achievable throughput is very low and unlikely to improve the energy efficiency if other RRHs are deactivated instead. Again, the SINR-based user-RRH pairing is performed and the corresponding P, R and U_{EE} are calculated. If the new U_{EE} is larger than U_{old}, Steps 5-11 are repeated until this condition does not hold. In this way, the RRHs that contribute the least to the overall energy efficiency will be deactivated if only the energy efficiency is improved.

To solve (7) in Steps 3 and 10 of Algorithm 1, we first rewrite (7) assuming that RRH activation and user-RRH pairing have been performed, as follows:

$$\max_{\boldsymbol{\omega},\mathbf{p}} U_{\text{EE}}(\boldsymbol{\omega},\mathbf{p}) = \frac{R(\boldsymbol{\omega},\mathbf{p})}{P(\boldsymbol{\omega},\mathbf{p})}, \tag{9}$$

subject to:

$$\sum_{u\in\mathcal{U}_s}\sum_{k\in\mathcal{K}} \omega_{ku} p_{sk} \le P_{\text{max},s} \quad \forall s \in \mathcal{S}_{\text{a}} \tag{9a}$$

$$\sum_{u\in\mathcal{U}_s}\sum_{k\in\mathcal{K}} \omega_{ku} R_{sku} \le R_{\text{fh},s} \quad \forall s \in \mathcal{S}_{\text{a}} \tag{9b}$$

$$\sum_{k\in\mathcal{K}} \omega_{ku} R_{sku} \ge R_{\text{min},u} \quad \forall s \in \mathcal{S}_{\text{a}}, u \in \mathcal{U}_s \tag{9c}$$

$$\sum_{u\in\mathcal{U}_s} \omega_{ku} \le 1 \quad \forall s \in \mathcal{S}_{\text{a}}, k \in \mathcal{K} \tag{9d}$$

$$p_{sk} \ge 0 \quad \forall s \in \mathcal{S}_{\text{a}}, k \in \mathcal{K}, \tag{9e}$$

where $\mathcal{S}_{\text{a}}$ is the set of activated RRHs, U_s is the set of users associated with RRH s. It is noted in (12) that $R_{sku} = B\log_2\left(1 + \frac{p_{sk}g_{sku}}{\sum_{i\in\mathcal{S}_{\text{a}}\setminus\{s\}} p_{ik}g_{iku} + N_0}\right)$ which is equivalent to (5) and $R_s = \sum_{u\in\mathcal{U}_s}\sum_{k\in\mathcal{K}} R_{sku}$.

To solve (9), the nonlinear fractional programming approach in [16] is used. Without loss of generality, we let $\psi = \frac{R(\boldsymbol{\omega},\mathbf{p})}{P(\boldsymbol{\omega},\mathbf{p})}$ and $\psi^* = \max_{\boldsymbol{\omega},\mathbf{p}} \frac{R(\boldsymbol{\omega},\mathbf{p})}{P(\boldsymbol{\omega},\mathbf{p})} = \frac{R(\boldsymbol{\omega}^*,\mathbf{p}^*)}{P(\boldsymbol{\omega}^*,\mathbf{p}^*)}$ where $\{\boldsymbol{\omega}^*,\mathbf{p}^*\}$ is the optimal solution to (9). Then, we obtain the following theorem.

Theorem 1. *ψ^* is achieved if and only if*

$$\begin{aligned}&\max_{\boldsymbol{\omega},\mathbf{p}} R(\boldsymbol{\omega},\mathbf{p}) - \psi^* P(\boldsymbol{\omega},\mathbf{p}) \\ &= R(\boldsymbol{\omega}^*,\mathbf{p}^*) - \psi^* P(\boldsymbol{\omega}^*,\mathbf{p}^*) = 0,\end{aligned} \tag{10}$$

where $R(\boldsymbol{\omega},\mathbf{p}) \ge 0$ and $P(\boldsymbol{\omega},\mathbf{p}) > 0$.

Proof: Refer to [16] for a similar proof. ■

By Theorem 1, (9) can be equivalently expressed as:

$$\max_{\boldsymbol{\omega},\mathbf{p}} f(\boldsymbol{\omega},\mathbf{p}) = R(\boldsymbol{\omega},\mathbf{p}) - \psi^* P(\boldsymbol{\omega},\mathbf{p}), \tag{11}$$

subject to (9a)-(9e). However, ψ^* has to be found for (11). As such, we employ Dinkelbach's method in [16] and design an iterative algorithm to solve (11). The iterative algorithm is summarized in Algorithm 2.

In Algorithm 1, the outer loop updates ψ in each iteration $\{\boldsymbol{\omega},\mathbf{p}\}$ obtained from the previous iteration until the convergence is achieved, i.e., $R(\boldsymbol{\omega},\mathbf{p}) - \psi P(\boldsymbol{\omega},\mathbf{p}) < \epsilon$ where ϵ is a very small positive value. In the inner loop, the following problem is solved:

$$\max_{\boldsymbol{\omega},\mathbf{p}} f(\boldsymbol{\omega},\mathbf{p}) = R(\boldsymbol{\omega},\mathbf{p}) - \psi P(\boldsymbol{\omega},\mathbf{p}), \tag{12}$$

Algorithm 2 Resource allocation algorithm

```
Set the maximum number of iterations I_max, convergence
  condition ε and ψ^(1) = 0.
Set iteration index i = 1.
for 1 ≤ i ≤ I_max (Outer Loop) do
  Solve the problem in (11) (Inner Loop).
  if R(ω, p) − ψ^(i) P(ω, p) < ε then
    break
  else
    Set ψ* = ψ^(i).
    Set ψ^(i+1) = R(ω^(i), p^(i)) / P(ω^(i), p^(i)) and i = i + 1.
  end if
end for
```

subject to (9a)-(9e). In the following theorem, we show that Algorithm 2 always converges to the optimal solution to (12).

Theorem 2. *The solution obtained from Algorithm 2 always converges to the global optimal solution to* (12).

Proof: Refer to [16] for a similar proof. ■

In the inner loop of Algorithm 2, the problem in (12) can be solved by dual decomposition. Firstly, the Lagrangian function of (12) can be written as follows:

$$\begin{aligned}\mathcal{L}(\boldsymbol{\omega},\mathbf{p},\boldsymbol{\lambda},\boldsymbol{\phi},\boldsymbol{\alpha}) &= \sum_{s\in\mathcal{S}_{\text{a}}}\sum_{u\in\mathcal{U}_s} w_u \sum_{k\in\mathcal{K}} \omega_{ku} R_{sku} \\ &- \psi \sum_{s\in\mathcal{S}_{\text{a}}}\left(P_{0,s} + \sum_{u\in\mathcal{U}_s}\sum_{k\in\mathcal{K}} \omega_{ku}\left(\eta_s p_{sk} + \beta_s R_{sku}\right)\right) \\ &- \psi\left(\sum_{s\in\mathcal{S}_{\text{a}}} P_{\text{c},s} + \sum_{s\in\mathcal{S}\setminus\mathcal{S}_{\text{a}}} P_{\text{sleep},s} + P_{\text{bh}}\right) \\ &+ \sum_{s\in\mathcal{S}_{\text{a}}} \lambda_s \left(P_{\text{max},s} - \sum_{u\in\mathcal{U}_s}\sum_{k\in\mathcal{K}} \omega_{ku} p_{sk}\right) \\ &+ \sum_{s\in\mathcal{S}_{\text{a}}} \phi_s \left(R_{\text{fh},s} - \sum_{u\in\mathcal{U}_s}\sum_{k\in\mathcal{K}} \omega_{ku} R_{sku}\right) \\ &+ \sum_{s\in\mathcal{S}_{\text{a}}}\sum_{u\in\mathcal{U}_s} \alpha_{su}\left(\sum_{k\in\mathcal{K}} \omega_{ku} R_{sku} - R_{\text{min},u}\right),\end{aligned} \tag{13}$$

where $\boldsymbol{\lambda} = \{\lambda_1, \ldots, \lambda_{|\mathcal{S}_{\text{a}}|}\}$, $\boldsymbol{\phi} = \{\phi_1, \ldots, \phi_{|\mathcal{S}_{\text{a}}|}\}$ and $\boldsymbol{\alpha} = \{\alpha_{11}, \ldots, \alpha_{|\mathcal{S}_{\text{a}}||\mathcal{U}_s|}\}$ are vectors of nonnegative Lagrange multipliers corresponding to constraints (9a), (9b) and (9c) respectively.

The Lagrangian dual function of (13) can be expressed as $D(\boldsymbol{\lambda},\boldsymbol{\phi},\boldsymbol{\alpha}) = \max_{\boldsymbol{\omega},\mathbf{p}} \mathcal{L}(\boldsymbol{\omega},\mathbf{p},\boldsymbol{\lambda},\boldsymbol{\phi},\boldsymbol{\alpha})$. The dual optimization problem can thus be formulated as follows:

$$\min_{\boldsymbol{\lambda},\boldsymbol{\phi},\boldsymbol{\alpha}} D(\boldsymbol{\lambda},\boldsymbol{\phi},\boldsymbol{\alpha}), \tag{14}$$

subject to $\lambda_s \ge 0$ and $\phi_s \ge 0$ for all $s \in \mathcal{S}_{\text{a}}$, and $\alpha_{su} \ge 0$ for all $s \in \mathcal{S}_{\text{a}}$ and $u \in \mathcal{U}_s$.

In this way, the convex problem in (14) can be solved using convex optimization techniques. However, the solution to a dual problem only gives the upper bound of the primal problem if the latter is not convex. Therefore, there may exist a nonzero duality gap between (14) and

(12), i.e., $\min_{\boldsymbol{\lambda},\boldsymbol{\phi},\boldsymbol{\alpha}} D(\boldsymbol{\lambda},\boldsymbol{\phi},\boldsymbol{\alpha}) - \max_{\boldsymbol{\omega},\mathbf{p}} f(\boldsymbol{\omega},\mathbf{p}) \neq 0$. However, the duality gap can be proven to approximate zero if the number of subchannels is sufficiently large in the following theorem [17]:

Theorem 3. *The duality gap between* (14) *and* (12) *approaches zero if* $|\mathcal{K}|$ *is sufficiently large.*

Proof: The proof is similar to those in [11], [17]. ∎

By Theorem 3, the solution to (14) will approximate that to (12) if the number of subchannels is sufficiently large. As such, we can solve (14) by dual decomposition. The solution approach is similar to that in [18]. Firstly, assuming that the equal transmission power has been allocated by each RRH s on each subchannel k, then optimal subchannel allocation can be performed as follows:

$$\omega_{ku^*} = \begin{cases} 1 & u^* = \arg\max_{u\in\mathcal{U}_s} m_{sku} \\ 0 & \text{otherwise} \end{cases} \quad \forall s \in \mathcal{S}_{\text{a}}, k \in \mathcal{K}, \tag{15}$$

where $m_{sku} = (w_u - \psi\beta_s - \phi_s + \alpha_{su})R_{sku} - (\psi\eta + \lambda_s)p_{sk}$.

Proof: See Appendix A for the derivation of (15). ∎

Let $q(s,k)$ indicates the user associated with RRH s that is allocated subchannel k, i.e., $q(s,k) = u \in \mathcal{U}_s$ whereby $\omega_{ku} = 1$, notation ω_{ku} can be removed and (13) can be rewritten as:

$$\begin{aligned} \mathcal{L}(\mathbf{p},\boldsymbol{\lambda},\boldsymbol{\phi},\boldsymbol{\alpha}) &= \sum_{s\in\mathcal{S}_{\text{a}}}\sum_{k\in\mathcal{K}} w_{q(s,k)}R_{skq(s,k)} \\ &- \psi\sum_{s\in\mathcal{S}_{\text{a}}}\left(P_{0,s} + \sum_{k\in\mathcal{K}}\left(\eta_s p_{sk} + \beta_s R_{skq(s,k)}\right)\right) \\ &- \psi\left(\sum_{s\in\mathcal{S}\setminus\mathcal{S}_{\text{a}}} P_{\text{sleep}} + P_{\text{bh}}\right) \\ &+ \sum_{s\in\mathcal{S}_{\text{a}}}\lambda_s\left(P_{\max,s} - \sum_{k\in\mathcal{K}} p_{sk}\right) \\ &+ \sum_{s\in\mathcal{S}_{\text{a}}}\phi_s\left(R_{\text{fh},s} - \sum_{k\in\mathcal{K}} R_{skq(s,k)}\right) \\ &+ \sum_{s\in\mathcal{S}_{\text{a}}}\sum_{k\in\mathcal{K}}\alpha_{sq(s,k)}R_{skq(s,k)} - \sum_{s\in\mathcal{S}_{\text{a}}}\sum_{u\in\mathcal{U}}\alpha_{su}R_{\min,u}. \end{aligned} \tag{16}$$

Using the Karush-Kuhn-Tucker (KKT) conditions [19], optimal power allocation can be derived (See (17) at the top of the next page). Note that $[x]^+$ is equivalent to $\max(0,x)$. With (15) and (17), the dual problem in (14) can be solved iteratively using a subgradient method [20] where the Lagrange multipliers are iteratively updated as follows:

$$\lambda_s^{(t+1)} = \lambda_s^{(t)} - \delta_1\left(P_{\max,s} - \sum_{u\in\mathcal{U}_s}\sum_{k\in\mathcal{K}}\omega_{ku}p_{sk}\right) \ \forall s \in \mathcal{S}_{\text{a}}, \tag{18}$$

$$\phi_s^{(t+1)} = \phi_s^{(t)} - \delta_2\left(R_{\text{fh},s} - \sum_{u\in\mathcal{U}_s}\sum_{k\in\mathcal{K}}\omega_{ku}R_{sku}\right) \forall s \in \mathcal{S}_{\text{a}}, \tag{19}$$

$$\alpha_{su}^{(t+1)} = \alpha_{su}^{(t)} - \delta_3\left(\sum_{k\in\mathcal{K}}\omega_{ku}R_{sku} - R_{\min,u}\right) \quad \forall s \in \mathcal{S}_{\text{a}}, u \in \mathcal{U}_s, \tag{20}$$

where δ_1, δ_2 and δ_3 are positive step sizes corresponding to (18), (19) and (20), respectively, which satisfy infinite travel conditions [20]; $\lambda_s^{(t)}$, $\phi_s^{(t)}$ and $\alpha_{su}^{(t)}$ are the respective λ_s, ϕ_s and α_{su} at the t-th iteration. For each update of the Lagrange multipliers, the subchannel and power allocation are recomputed again using (15) and (17). The process is repeated in the inner loop of Algorithm 2 until convergence or the predetermined maximum number of iterations, $T_{\max}$ is reached.

IV. Complexity Analysis

In Algorithm 2, the maximum number of iterations required to solve (9) is $I_{\max}T_{\max}$ where $I_{\max}$ is the maximum number of iterations for the outer loop and $T_{\max}$ is the maximum number of iterations for the inner loop. Therefore, the maximum number of iterations for Algorithm 1 can be estimated as $|\mathcal{S}|I_{\max}T_{\max}$. Hence, the asymptotic complexity of our proposed algorithm is of $\mathcal{O}(|\mathcal{S}|I_{\max}T_{\max})$. The proposed scheme is intended to be executed by the BBU pool periodically in order to keep up with the channel variations.

V. Results and Discussion

A single-cell H-CRAN network, which consists of an MBS with a macrocell radius of 500 m and six picocell RRHs, is considered. The RRHs are randomly distributed within the macrocell since small cells are deployed at random locations. We set the number of subchannels to 100 with each having a bandwidth of 180 kHz, following the 3GPP specifications [21]. Here, we assume that all fronthaul links are identical, therefore $R_{\text{fh},s} = R_{\text{fh}}$, $P_{\text{ft},s} = P_{\text{ft}}$, $P_{\text{sw},s} = P_{\text{sw}}$, $\tau_s = \tau$ and $n_{\text{port},s} = n_{\text{port}}$ for all $s \in \mathcal{S}$. The power parameters of the RRHs are set according to the power consumption of the picocells as follows: $P_{\max,s} = 30$ dBm, $P_{0,s} = 6.8$ W, $P_{\text{sleep},s} = 4.3$ W and $\eta_s = 4.0$ [12]. For the fronthaul links, we set the following parameters as in [14]: $P_{\text{ft}} = 3$ W, $P_{\text{sw}} = 300$ W, $\tau = 0.8$ and $n_{\text{port}} = 24$. The power consumed by the backhaul link is assumed to be 13.25 W[3]. For channel modeling, we consider Rayleigh fading, which are independently and identically distributed (i.i.d.) with zero mean and unit variance. We also follow the 3GPP specifications [22] by considering log-normal shadowing which is also i.i.d. with zero mean and a standard deviation of 10 dB, and the small cell path loss model: $140.7 + 36.7\log d$ where d is the distance between the RRH and the user in km. The noise power spectral density and noise figure are set to -174 dBm/Hz and 9 dB [22] respectively. All the users are uniformly distributed within the network, which is a practical user distribution. For the proposed scheme, δ_1,

[3]This value is calculated using the same power consumption model as the fronthaul link. The relevant parameters are set similar to those of the fronthaul link, except that the backhaul capacity and the control traffic carried in the backhaul link are assumed to be 100 Mb/s and 10 Mb/s, respectively.

$$p_{sk} = \left[\frac{B(w_{q(s,k)} - \psi\beta_s - \phi_s + \alpha_{sq(s,k)})}{(\lambda_s + \psi\eta_s)\ln 2} - \frac{\sum_{i\in\mathcal{S}_\mathrm{a}\setminus\{s\}} p_{ik} g_{ikq(s,k)} + N_0}{g_{skq(s,k)}} \right]^+ \quad \forall s \in \mathcal{S}_\mathrm{a}, k \in \mathcal{K}. \tag{17}$$

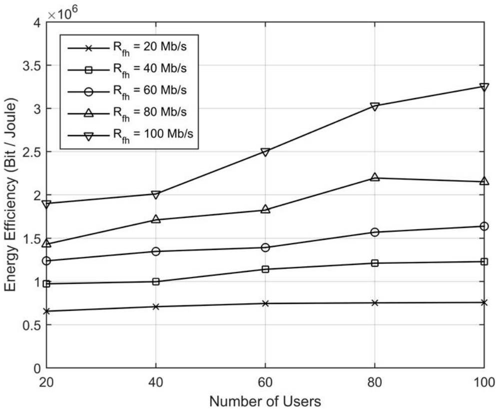

Figure 2. Energy efficiency of the proposed scheme for the H-CRAN with different fronthaul capacities.

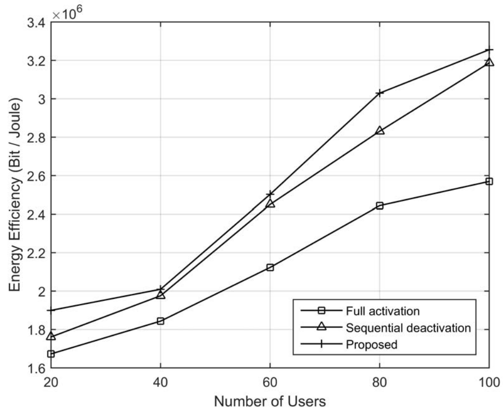

Figure 3. Comparison between the proposed scheme and the baseline schemes for the H-CRAN with 100 Mbps fronthaul capacity.

δ_2 and δ_3 are set following the square summable but not summable rule [20] with $\delta_1 = \delta_2 = \delta_3 = 0.001$ at the first iteration, $I_{\max} = 100$, $T_{\max} = 100$, $\epsilon = 0.01$, and $w_u = 1$ for all $u \in \mathcal{U}$. All results are averaged over 100 simulation runs.

Fig. 2 shows the energy efficiency achieved by the proposed scheme for the H-CRAN with different fronthaul capacities. It is observed that the energy efficiency improves with the fronthaul capacity as the latter approaches the capacity that can be supported by an RRH on the wireless channel.

Next, we compare the proposed scheme with the following several baseline schemes.

- Full activation: All RRHs are activated and energy efficiency of the H-CRAN is optimized via resource allocation, which similar to that in [11] except that fronthaul capacity constraints are considered.
- Sequential activation: RRHs are deactivated one by one following an ascending index order. Deactivation halts when the energy efficiency cannot be further improved.

Fig. 3 shows the energy efficiency performance of the H-CRAN with 100 Mb/s fronthaul capacity. The proposed scheme outperforms the baseline schemes with average energy efficiency gains of 4.16% and 18.2% over the sequential deactivation and full activation schemes, respectively. The full activation and sequential deactivation schemes do not find out which RRHs are the most suitable to be deactivated, unlike our proposed scheme, thus resulting in inferior energy efficiency performance.

VI. Conclusion

In this paper, we have proposed a joint RRH activation, user-RRH pairing and resource allocation scheme for maximizing the energy efficiency of H-CRANs. We have formulated an optimization problem that maximizes energy efficiency of an H-CRAN subjecting to the limited fronthaul capacity. We have designed an iterative algorithm that performs greedy RRH activation, SINR-based user-RRH pairing and solved the resource allocation problem by dual decomposition. Simulation results have demonstrated that the proposed scheme provides an average energy efficiency gain of 4.16%-18.2% compared to the baseline schemes. In future, we will consider a more comprehensive resource allocation model that involves macrocell and RRH users.

Appendix A. Derivation of (15)

Let (13) be rewritten as follows:

$$\begin{aligned}
&\mathcal{L}(\mathbf{p}, \boldsymbol{\omega}, \boldsymbol{\lambda}, \boldsymbol{\phi}, \boldsymbol{\alpha}) \\
&= \sum_{s\in\mathcal{S}_\mathrm{a}} \sum_{k\in\mathcal{K}} \sum_{u\in\mathcal{U}_s} L_{sku}(p_{sk}, \omega_{ku}, \lambda_s, \phi_s, \alpha_{su}) \\
&- \psi \left(\sum_{s\in\mathcal{S}_\mathrm{a}} (P_{0,s} + P_{\mathrm{c},s}) + \sum_{s\in\mathcal{S}\setminus\mathcal{S}_\mathrm{a}} P_{\mathrm{sleep},s} + P_{\mathrm{bh}} \right) \\
&+ \sum_{s\in\mathcal{S}_\mathrm{a}} \lambda_s P_{\max,s} + \sum_{s\in\mathcal{S}_\mathrm{a}} \phi_s R_{\mathrm{fh},s} - \sum_{s\in\mathcal{S}_\mathrm{a}} \sum_{u\in\mathcal{U}_s} \alpha_{su} R_{\min,u}
\end{aligned}$$

where

$$\begin{aligned}
&L_{sku}(\omega_{ku}, p_{sk}, \lambda_s, \phi_s, \alpha_{su}) \\
&= \omega_{ku}\left((w_u - \psi\beta_s - \phi_s + \alpha_{su})R_{sku} - (\psi\eta + \lambda_s)p_{sk}\right)
\end{aligned}$$

Given the values of the Lagrange multipliers and that power allocation has been performed, then, for each RRH s and each subchannel k, the user associated with RRH s that gives the largest value of $L_{sku}(\omega_{ku} =$

$1, p_{sk}, \lambda_s, \phi_s, \alpha_{su})$, i.e., user u^* will be allocated subchannel k:

$$\begin{aligned} u^* &= \arg\max_{u \in \mathcal{U}_s} L_{sku}(\omega_{ku} = 1, p_{sk}, \lambda_s, \phi_s) \\ &= \arg\max_{u \in \mathcal{U}_s} \Big((w_u - \psi\beta_s - \phi_s + \alpha_{su}) R_{sku} \\ &\quad - (\psi\eta + \lambda_s) p_{sk} \Big) \\ &\qquad\qquad\qquad\qquad \forall s \in \mathcal{S}_{\mathrm{a}}, k \in \mathcal{K} \end{aligned}$$

ACKNOWLEDGMENT

The authors would like to thank the anonymous reviewers for the constructive comments and suggestions to further improve the quality of the paper.

REFERENCES

[1] C.-H. I et al., "Toward green and soft: a 5G perspective," *IEEE Commun. Mag.*, vol. 52, no. 2, pp. 66–73, Feb. 2014.

[2] A. Checko et al., "Cloud RAN for mobile networks - a technology overview," *IEEE Commun. Surveys Tuts.*, vol. 17, no. 1, pp. 405–426, Mar. 2015.

[3] M. Peng, Y. Li, J. Jiang, J. Li, and C. Wang, "Heterogeneous cloud radio access networks: a new perspective for enhancing spectral and energy efficiency," *IEEE Wireless Commun. Mag.*, vol. 21, no. 6, pp. 126–135, Dec. 2014.

[4] P.-R. Li, T.-S. Chang, and K.-T. Feng, "Energy-efficient power allocation for distributed large-scale MIMO cloud radio access networks," in *Proc. WCNC*, Istanbul, Turkey, Apr. 2014, pp. 1856–1861.

[5] X. Huang, G. Xue, R. Yu, and S. Leng, "Joint scheduling and beamforming coordination in cloud radio access networks with QoS guarantees," *IEEE Trans. Veh. Technol.*, pp. 1–12, Aug. 2015, in press.

[6] C. Yoon and D.-H. Cho, "Energy efficient beamforming and power allocation in dynamic TDD based CRAN system," *IEEE Commun. Lett.*, pp. 1–4, Aug. 2015.

[7] Y. Shi, J. Zhang, and K. B. Letaief, "Group sparse beamforming for green cloud-RAN," *IEEE Trans. Wireless Commun.*, vol. 13, no. 5, pp. 2809–2823, May 2014.

[8] J. Tang, W. P. Tay, and T. Q. S. Quek, "Cross-layer resource allocation with elastic service scaling in cloud radio access networks," *IEEE Trans. Wireless Commun.*, vol. 14, no. 9, pp. 5068–5081, Sep. 2015.

[9] J. Li, M. Peng, A. Cheng, Y. Yu, and C. Wang, "Resource allocation optimization for delay-sensitive traffic in fronthaul constrained cloud radio access networks," *IEEE Syst. J.*, pp. 1–12, Nov. 2014, in press.

[10] G. K. Tran et al., "Dynamic cell activation and user association for green 5G heterogeneous cellular networks," in *Proc. 26th Int. Symp. PIMRC*, Hong Kong, China, Aug.-Sep. 2015, pp. 2364–2368.

[11] M. Peng, K. Zhang, J. Jiang, H. Wang, and W. Wang, "Energy-efficient resource assignment and power allocation in heterogeneous cloud radio access networks," *IEEE Trans. Veh. Technol.*, vol. 64, no. 11, pp. 5275–5287, Nov. 2015.

[12] G. Auer et al., "How much energy is needed to run a wireless network," *IEEE Wireless Commun. Mag.*, vol. 18, no. 5, pp. 40–49, Oct. 2011.

[13] S. Tombaz, A. Vastberg, and J. Zander, "Energy- and cost-efficient ultra-high-capacity wireless access," *IEEE Wireless Commun. Mag.*, vol. 18, no. 5, pp. 18–24, Oct. 2011.

[14] S. Tombaz, K. W. Sung, and J. Zander, "Impact of densification on energy efficiency in wireless access networks," in *Proc. GC Wkshps*, Anaheim, California, USA, Dec. 2012, pp. 57–62.

[15] I. C. Wong and B. L. Evans, "Optimal downlink OFDMA resource allocation with linear complexity to maximize ergodic rates," *IEEE Trans. Wireless Commun.*, vol. 7, no. 3, pp. 962–971, Mar. 2008.

[16] W. Dinkelbach, "On nonlinear fractional programming," *Manag. Sci.*, vol. 13, no. 7, pp. 492–498, Mar. 1967.

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

[18] K. Lee, O. Jo, and D.-H. Cho, "Cooperative Resource Allocation for Guaranteeing Intercell Fairness in Femtocell Networks," *IEEE Commun. Lett.*, vol. 15, no. 2, pp. 214–216, Feb. 2011.

[19] S. Boyd and L. Vandenberghe, *Convex optimization*. Cambridge University Press, 2004.

[20] S. Boyd, L. Xiao, and A. Mutapcic, "Subgradient methods," Notes for EE364b, Stanford University, 2006-07.

[21] "Technical specification group radio access network; evolved universal terrestrial radio access (E-UTRA); physical channels and modulation," TS 36.211, 3rd Generation Partnership Project (3GPP), Sophia-Antipolis, France, Sep. 2015. [Online]. Available: http://www.3gpp.org/DynaReport/36.211.htm

[22] "3rd generation partnership project; technical specification group tadio access network; small cell enhancements for e-utra and e-utran - physical layer aspects (release 12)," TR 36.872, 3rd Generation Partnership Project (3GPP), Sophia-Antipolis, France, Sep. 2013. [Online]. Available: http://www.3gpp.org/dynareport/36872.htm