Green Heterogeneous Cloud Radio Access Networks: Potential Techniques, Performance Trade-offs, and Challenges

Yuzhou Li, Tao Jiang, Kai Luo, and Shiwen Mao

ABSTRACT

As a flexible and scalable architecture, heterogeneous cloud radio access networks (H-CRANs) inject strong vigor into the green evolution of current wireless networks. But the brutal truth is that EE improves at the cost of other indices such as SE, fairness, and delay. It is thus important to investigate performance trade-offs for striking flexible balances between energy-efficient transmission and excellent QoS guarantees under this new architecture. In this article, we first propose some potential techniques to energy-efficiently operate H-CRANs by exploiting their features. We then elaborate the initial ideas of modeling three fundamental trade-offs, namely EE-SE, EE-fairness, and EE-delay trade-offs, when applying these green techniques, and present open issues and challenges for future investigation. These related results are expected to shed light on green operation of H-CRANs from adaptive resource allocation, intelligent network control, and scalable network planning.

INTRODUCTION

BACKGROUND AND MOTIVATION

The dramatic increase in the number of smartphones and tablets with ubiquitous broadband connectivity has triggered an explosive growth in mobile data traffic [1]. Cisco forecasts that the amount of global mobile data traffic will increase 7-fold from 2016 to 2021, the majority of which are generated by energy-hungry applications such as mobile video [1]. This is also referred to as the well-known 1000× data challenge in cellular networks. Meanwhile, the number of devices connected to the global mobile communication networks will reach 100 billion in the future, and that of mobile terminals will surpass 10 billion by 2020 [2].

Although unprecedented opportunities for the development of wireless networks are created by the massive traffic amount and connected devices, a concomitant crux is that this growth skyrockets the energy consumption (EC) and greenhouse gas emissions in the meantime. From statistical data, the information and communication technology (ICT) industry is responsible for 2 percent of worldwide CO_2 emissions and 2–10 percent of global EC, of which more than 60 percent is directly attributed to radio access networks (RANs) [3]. In this regard, 5G wireless communication networks are anticipated to provide spectral efficiency (SE) and energy efficiency (EE) growth by a factor of at least 10 and 10 times longer battery life of connected devices [2].

CONCEPT OF H-CRANS

To meet the 1000× data challenge, heterogeneous networks (HetNets), composed of a diverse set of small cells (e.g., microcells, picocells, and femtocells) overlaying the conventional macrocells, have been introduced as one of the most promising solutions [2]. However, the ubiquitous deployment of HetNets is accompanied by the following shackles:

- Severe interference: The spectrum reuse among cells incurs severe mutual interference, which may significantly reduce the expected system SE and also decrease the network EE.
- Unsatisfactory EE: The densely deployed small cells lead to escalated EC and thus reduced EE, and also increases capital expenditures (CAPEX) and operational expenditures (OPEX).
- No computing-enhanced coordination centers: There are no centralized units with strong computing abilities to globally coordinate multi-tier interference and execute cross-RAN optimization, which dramatically limits cooperative gains among cells.
- Inflexibility and unscalability: Fragmented base stations (BSs) result in inflexible and unscalable network control and operations, thus leading to redundant network planning and inconvenient network upgrade.

To overcome these challenges faced by HetNets, cloud RANs (C-RANs), new centralized cellular architectures armed with powerful cloud computing and virtualization techniques, have been put forward in parallel to coordinate interference and manage resources across cells and RANs [4]. In C-RANs, a large number of low-cost low-power remote radio heads (RRHs), connecting to the baseband unit (BBU) pool through the fronthaul links, are randomly deployed to enhance the wireless capacity in hotspots. Consequently, the combination of HetNets and C-RANs, known

The authors first propose some potential techniques to energy-efficiently operate H-CRANs by exploiting their features. They then elaborate the initial ideas of modeling three fundamental trade-offs, namely EE-SE, EE-fairness, and EE-delay trade-offs, when applying these green techniques, and present open issues and challenges for future investigation.

Yuzhou Li, Tao Jiang, and Kai Luo are with Huazhong University of Science and Technology; Shiwen Mao is with Auburn University.

Digital Object Identifier:
10.1109/MCOM.2017.1600807

0163-6804/17/$25.00 © 2017 IEEE

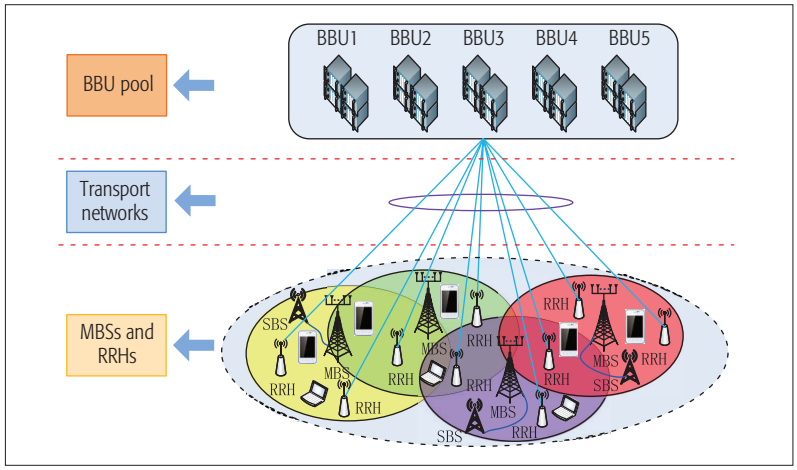

Figure 1. The architecture of H-CRANs.

as heterogeneous C-RANs (H-CRANs), becomes a potential solution to support both spectral- and energy-efficient transmission.

Green H-CRANs

As mentioned above, one of the main missions of H-CRANs from their birth is to construct eco-friendly and cost-efficient wireless communication systems. Benefiting from H-CRANs' global coordination ability, many promising techniques, such as joint processing/allocation, traffic load offloading, energy balance, self-organization, and adaptive network deployment, can be applied in these scenarios for energy-efficient transmissions. Unfortunately, the network EE improves usually at the cost of the performance of other technique metrics, such as SE, fairness, and delay, all of which, however, are equally important as EE to guarantee users' quality of service (QoS). That is, there are EE-SE, EE-fairness, and EE-delay trade-offs. It is thus interesting to investigate these performance trade-offs in H-CRANs for establishing rules to flexibly balance the network EE and users' QoS demands when greening H-CRANs.

Compared to existing works (e.g., [5]) on the system architecture or radio resource management (RRM), mainly in terms of EE and SE, this article focuses on the green evolution of H-CRANs, and particularly investigates it from the perspective of EE-SE, EE-fairness, and EE-delay trade-offs instead of the indices themselves. To reach our targets, we organize the remainder of this article as follows. In the following section, we first simply review the architecture of H-CRANs and then exploit their features to propose three potential techniques for green H-CRANs. Then we introduce the possible methods to depict these trade-offs and also provide corresponding challenges and open problems when applying these proposed techniques. We conclude the article in the final section.

Architecture of H-CRANs and Potential Green Techniques

In C-RANs, the idea of dividing conventional cellular BSs into two parts, BBUs and RRHs, is introduced. BBUs are then integrated into centralized BBU pools, where cloud computing and virtualization techniques are implemented to enhance computational ability and to virtualize network function. BBUs are responsible for resource control and signal processing, and RRHs for information radiation and reception, with their interconnection via dedicated transport networks. Thus, the cloud-computing-enhanced centralized BBU pools facilitate cross-cell and cross-RAN information sharing, which paves the path for global resource optimization adapting to network conditions (e.g., channel conditions, interference strength, traffic loads). H-CRANs absorb this architecture in C-RANs and maintain macro BSs (MBSs) and small cell BSs (SBSs) in HetNets to support both global control and seamless communications.

Architecture of H-CRANs

As shown in Fig. 1, H-CRANs are composed of three functional modules.

Real-Time Virtualized and Cloud-Enhanced BBU Pool: Equipped with powerful virtualization techniques and strong real-time cloud computing ability, BBU pools integrate independent BBUs scattered in cells.

High-Reliability Transport Networks: RRHs are connected to BBUs in the BBU pool via high-bandwidth low-latency fronthaul links such as optical transport networks. The data and control interfaces between the BBU pool and MBSs are S1 and X2, respectively [6].

MBSs, SBSs, and RRHs: In H-CRANs, multiple access points (APs), for example, MBSs, SBSs, and RRHs, coexist. MBSs are deployed mainly for network control and mobility performance improvement, for example, decreasing handover times to avoid ping-pong effects for high-mobility users. SBSs and RRHs are geographically distributed within cells close to users to increase capacity and decrease transmit power in the meantime.

In H-CRANs, the function separation between BBUs and RRHs, the decoupling between control and data planes, and the cloud-computing-enhanced centralized integration of BBUs facilitate efficient management of densely deployed mobile networks. For example, the operators only need to install new RRHs and connect them to the BBU pool to expand network coverage and improve network capacity. Moreover, flexible software solutions can easily be implemented under this architecture. For instance, operators can upgrade RANs and support multi-standard operations only through software update by deploying software defined radio (SDR).

Potential Techniques for Green H-CRANs

The four revolutionary changes, that is, function separation, control-data decoupling, centralized architecture, and cloud-computing-enhanced processing, make H-CRANs significantly different from existing second generation (2G), 3G, and 4G wireless networks. By exploiting these features, it is possible to construct H-CRANs that are flexible in network management, adaptive in network control, and scalable in network planning. As a result, energy-efficient operation of H-CRANs without significant loss in other indices such as SE, fairness, and delay can be achieved.

Joint Resource Optimization across RRHs and RANs: In H-CRANs, each BBU first collects its individual network conditions and then shares this information within the BBU pool. As a result, this distributed-collection centralized-control archi-

tecture, further enhanced by virtualization techniques and cloud computing, enables efficient transmission/reception cooperation across RRHs and convenient global control across RANs. Consequently, the existing cooperative techniques, such as coordinated multipoint (CoMP) transmission, enhanced inter-cell interference coordination (eICIC), and interference alignment (IA), can readily be implemented in H-CRANs. All these techniques are self-contained in theory but have rarely been applied to conventional cellular networks because of difficulties in sharing and handling global network information.

As introduced above, multi-RANs and multi-APs with different coverage and functions are deployed in H-CRANs. As a result, unlike traditional single-mode terminals communicating only through a RAN's AP, multi-mode terminals could send and receive data concurrently through multiples of them. This indicates H-CRANs with a new characteristic of network diversity, which can be exploited to design user association strategies. By this, traffic load distributions among RANs and APs can be well balanced, which in turn affects the working states of RANs and resource optimization, and thus affects network interference and EE.

Moreover, under this new centralized architecture, the network EE can be further improved by incorporating more resource allocation dimensions (e.g., power allocation, subcarrier assignment, user association, and RRH operation) into the formulations. Figure 2 shows that joint optimization of RRH operation and power allocation improves EE by up to 84 percent compared to the power-allocation-only algorithm in downlink H-CRANs. Thus, through the aforementioned joint resource optimization and network-diversity-aware user association, significant improvement in EE and reduction in EC can be achieved.

Large-Scale MBS and SBS Deployment: Compared to the transmit power, the overall static power consumption by MBSs and SBSs, composed of cooling and circuit power, are usually much larger [7]. For example, a typical Universal Mobile Telecommunications System (UMTS) BS consumes 800–1500 W with RF output power of 20–40 W. As a result, under the constraints of basic coverage requirements, the deployment of MBSs and SBSs, characterized by the distance between two MBS sites and the number of SBSs per site, affects the area power consumption (APC) and the area SE (ASE) significantly in H-CRANs. The general purpose of large-scale MBS and SBS deployment is to macroscopically plan an appropriate number of BSs to support users' demands for energy saving by avoiding the static power consumption.

Intuitively, the APC will sharply decrease if we reduce the number of MBSs (i.e., increase the inter-site distance). Meanwhile, the ASE will also decrease, because the increased inter-site distance reduces the spectrum reuse. Similarly, the number of SBSs deployed in each site will also affect the APC and the ASE. As an example, Fig. 3 clearly shows the significant impacts of the configuration of MBSs and SBSs on the APC and ASE under practical parameter settings. Therefore, we need careful network planning from a large-scale perspective to flexibly balance these two metrics and to conveniently upgrade the system.

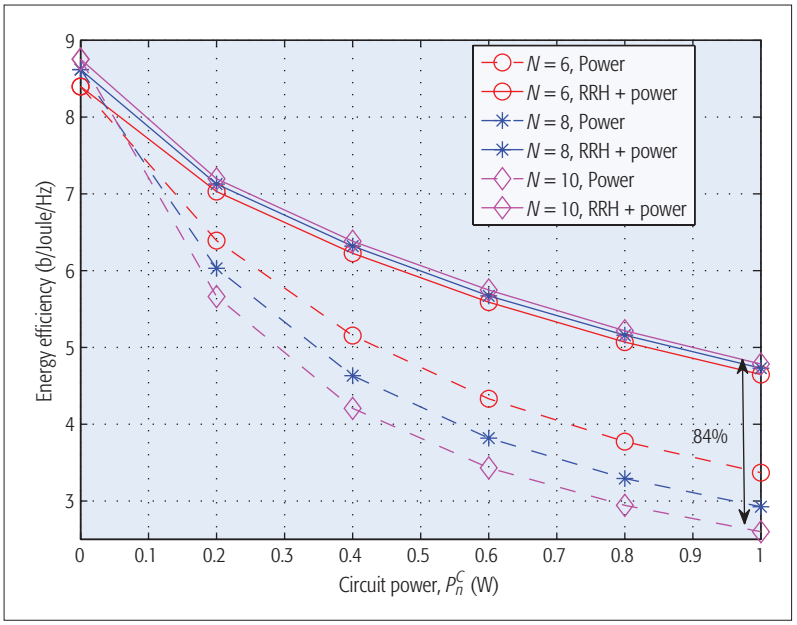

Figure 2. An example: EE variation with the circuit power of each RRH, denoted by P_n^c, in downlink H-CRANs, where an MBS, N RRHs, and 16 users are included. In this example, we maximize the network EE by optimizing RRH operation and power allocation subject to constraints of users' minimum rate requirements of R^{req} = 2 b/Hz.

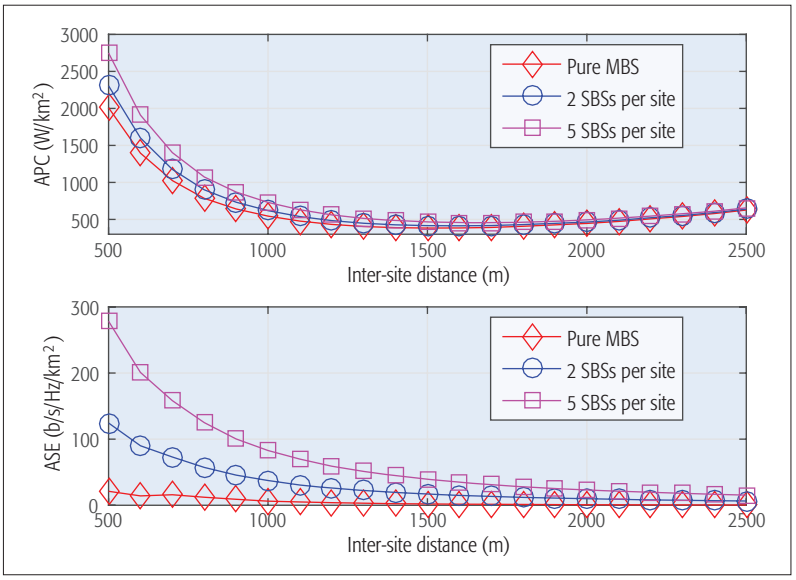

Figure 3. An example: The APC and ASE vs. the inter-site distance subject to a 95 percent coverage constraint. In the figure, we adopt the practical models for the BS power consumption given by $P^{tot} = aP^{tx} + b$, where a_{MBS} = 22.6, b_{MBS} = 412.4 W, a_{SBS} = 5.5, and b_{SBS} = 32 W (note that SBSs refer to micro BSs in the figure) [15].

Load-Aware RRH Operations: The so-called worst case network planning philosophy has been widely adopted to guarantee users' QoS even during peak traffic periods in conventional cellular networks. However, mobile traffic loads usually vary in both spatial and temporal domains, which is referred to as the tidal phenomenon. Specifically, the fraction of time when the traffic is below 10 percent of the peak during a day is about 30 percent on weekdays and 45 percent on week-

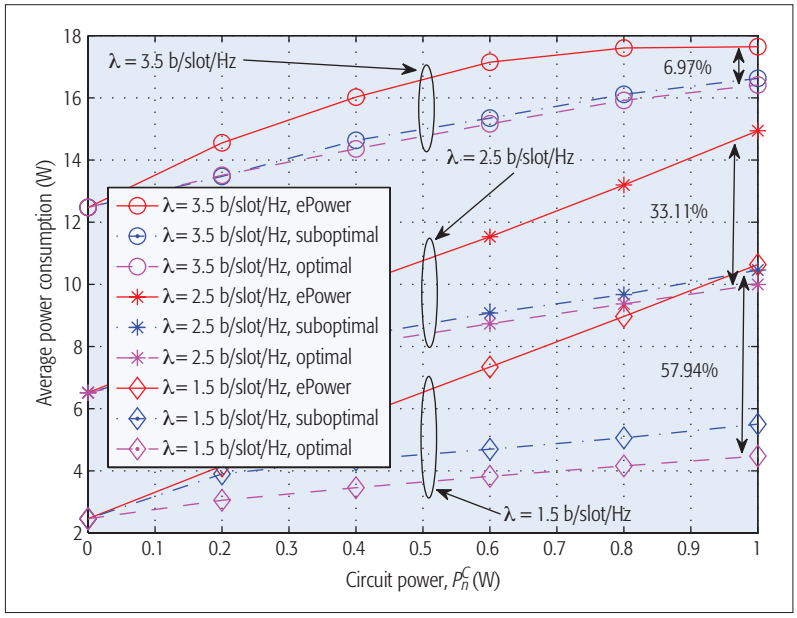

Figure 4. An example: average power consumption with the circuit power of each RRH, denoted by P_n^C, under different traffic arrival rates λ in downlink H-CRANs, where a MBS, 8 RRHs, and 12 users are included. In this example, we formulate a network EE maximization problem that enfolds stochas tic and time-varying traffic arrivals to jointly optimize RRH operation and power allocation.

ends [8]. As a result, a large number of RRHs are extremely underutilized in cases of dense deployment in H-CRANs during off-peak periods, but RRHs still consume circuit power even with little or no activity. Consequently, a significant waste of EC and a sharp decrease in EE will be caused if RRHs are underutilized but still activated. Thus, apart from the aforementioned spatial deployment, energy conservation can also be achieved by exploiting temporal traffic variations. For the fixed deployment, we can adopt load-aware network control in H-CRANs to perform on/off operations of RRHs adapting to spatial and temporal traffic amounts to improve EE.

As an example, we consider a downlink H-CRAN to show the impacts of load-aware RRH on/off operations on energy expenditure. Specifically, we jointly optimize RRH operation and power allocation to maximize the network EE with stochastic and time-varying traffic arrivals taken into account. Two algorithms, optimal and suboptimal, are developed to solve the problem. Figure 4 shows that the proposed algorithms can dramatically reduce the energy consumption compared to the algorithm without RRH operation (i.e., only optimizing power allocation), denoted by ePower, especially in light and middle traffic states (up to a 58 percent gain in light traffic states when the traffic arrival rate λ = 1.5 b/slot/Hz).

Performance Trade-offs and Challenges for Green H-CRANs

Leveraging the proposed potential green techniques in H-CRANs, it is then of importance to explore the key theories that support ubiquitously energy-efficient transmission and meanwhile provide satisfactory QoS for users. Among them, performance trade-offs deserve significant consideration [9].

Apart from the widely studied deployment efficiency-EE, EE-SE, bandwidth-power, and delay-power trade-offs [9], there are two additional fundamental trade-offs: EE-fairness and EE-delay trade-offs. This section elaborates the ideas of modeling these two trade-offs, analyzes challenges and open problems, and provides some possible solutions. Since H-CRANs originally are designed to enhance the network SE and thus the wireless capacity as well, we also review the key concepts and present challenges associated with the EE-SE trade-off under this new architecture.

EE-SE Trade-off

Vast existing research falls into this area due to the following reasons. The traditional indices EC and SE measure how small the amount of energy is needed to satisfy users' QoS and how efficiently limited spectrum is utilized, respectively. However, both of them fail to quantify how efficiently the energy is consumed (i.e.,, EE). Moreover, the optimality of EE and EC and that of EE and SE are not always achieved simultaneously and may even conflict with each other [9]. As a consequence, the existing results from the EC minimization or the SE maximization usually can hardly provide insights into EE-SE trade-off problems.

The general idea of modeling the EE-SE trade-off is that the system maximizes the network EE [10] or a weighted EE-SE trade-off index [11] under the constraints of users' QoS and resource allocation (e.g., power allocation and RRH operation). As a common feature, these works usually assume infinite backlog, that is, there is always data for transmission in the buffer. Under this view, formulations are presented and algorithms are developed only based on the observation time, where the network EE is defined as the ratio of the instantaneous achievable sum rate R_{tot} to the corresponding total power consumption P_{tot} [10, Eq. 5]. Note that P_{tot} is usually modeled to include both transmit and circuit energy consumption, which is affected by the power amplifier inefficiency, transmit power, and circuit power. In this article, we call these formulations short-term (i.e., snapshot-based) models, since only short-term system performance is considered. Accordingly, we denote the network EE of this kind of definition by $EE_{short\text{-}term}$ for simplicity.

Although there have been a large number of works addressing the EE-SE trade-off based on the short-term models, lots of problems remain open in complex H-CRANs. First, jointly considering multi-dimensional resource optimization and multi-available signal processing techniques, it is challenging to formulate EE-SE trade-off problems with network conditions and users' requirements both taken into account in H-CRANs. Furthermore, due to the nonconvexity of $EE_{short\text{-}term}$ [10, Eq. 5; 11, Eq. 26], EE-SE trade-off problems are usually difficult to solve even if we only optimize power allocation in spectrum-sharing H-CRANs. As a result, these problems become much more complicated once we extend from one-dimensional to multi-dimensional resource optimization. Thus, how to develop joint resource allocation algorithms that reach the theoretical limits of the network EE and thus serve as benchmarks to evaluate performance of other heuristic algorithms is

another challenge. Moreover, it is also necessary to develop cost-efficient and easy-to-implement algorithms with acceptable performance levels to solve these problems for practical applications.

EE-Fairness Trade-Off

The widely studied EE-optimal problems (NEPs) in H-CRANs emphasize the network EE maximization without considering EE fairness (i.e., ignoring the EE of individual links). By purely benefiting the links in good network conditions (e.g., excellent wireless channel, little interference, low traffic loads, or all), the NEPs improve the network EE at the cost of the EE of the links in poor conditions. As a result, the NEPs would inevitably lead to severe unfairness among links in terms of EE. However, as traditional concerns on individual links' SE or EC, it is also important to guarantee the EE of each link in the users' perception. It is therefore of interest to investigate the EE-fairness trade-off in H-CRANs, but to the best of our knowledge, studies on this issue have so far been very scarce.

To intuitively show the EE-fairness trade-off, we take the max-min EE fairness in an uplink orthogonal frequency-division multiple access (OFDMA)-based cellular network (it can be seen as a special case of single-cell H-CRANs) as an example. Specifically, we maximize the EE of the worst case link subject to subcarrier assignment and power allocation constraints to ensure the max-min EE fairness among links, which is referred to as the max-min EE-optimal problem (MEP). In Fig. 5, we compare the statistical performance between the NEP and the MEP from three aspects: the EE of the network, the best link, and the worst link. Observe that the EE of the best and worst links in the NEP differs significantly, while the EE, whether of the network, the best link, or the worst link in the MEP, is well balanced. This is because the NEP maximizes the network EE at the cost of the EE fairness among links, but reversely, the MEP sacrifices the network EE to guarantee the max-min EE fairness.

Figure 5 exhibits the phenomenon of the EE-fairness trade-off, but we are still at a very primary stage of revealing and tuning this trade-off, limited by the following two challenges:

- Unified frameworks to quantify and formulate the EE-fairness trade-off are currently not available.
- General techniques or analytical methods to tackle the EE-fairness trade-off problems are still open.

It should be pointed out that the utility theory, originally used to investigate the rate-fairness trade-off [12], is a possible method to demystify the quantitative EE-fairness trade-off.

EE-Delay Trade-off

As far as we know, the concept of the EE-delay trade-off was first proposed by H. V. Poor *et al.* in 2009 [13], where the authors showed that the delay constraints would lead to a loss in EE at equilibrium by a game-theoretical approach. However, to date, how to quantify and control the EE-delay trade-off is still unresolved.

In our view, one possible reason that prevents the existing works including [13] from obtaining a quantitative trade-off is the choice of adopting short-term models with the full buffer assumption, where $EE_{short\text{-}term}$ is used to characterize the network EE. However, different from the full buffer assumption, practical H-CRANs operate in the presence of time-varying wireless channels and stochastic traffic arrivals, both of which significantly affect the EE and delay, and thus the EE-delay trade-off. Hence, short-term formulations in general cannot reflect the delay due to their independence of time and without considering traffic arrivals. As a result, it is unlikely for such models to show the explicit EE-delay relationships.

We further illustrate the principles behind the EE-delay trade-off with two extreme cases. Regarding stochastic traffic arrivals, in the case of aggressive emphasis on the EE, transmission decisions should be triggered only when network conditions are good enough, by which the delay performance degrades inevitably. Alternatively, to ensure small delay, the network has to transmit data at the cost of energy expenditure even when network conditions are very poor, which undoubtedly decreases the EE. Thus, to model the EE-delay trade-off, the following two issues need to be considered:

- How to decide whether to transmit data or defer a transmission in each slot in terms of the EE and delay and how to optimize resource allocation such as power allocation, subcarrier assignment, and RRH operation if transmission is chosen
- How to ensure that deferring transmissions to anticipate more advantageous network conditions becoming available in the future would not result in an uncontrollable delay because of time-variant, stochastic, and unpredicted network conditions

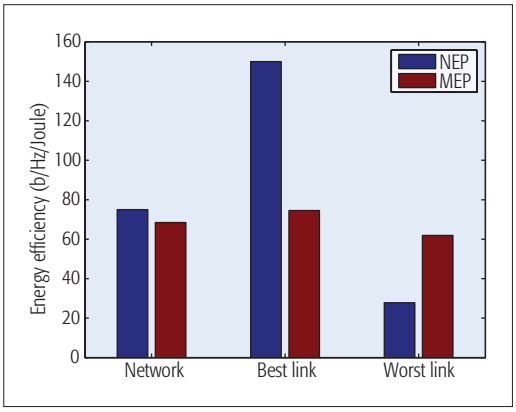

Figure 5. Illustration of the EE-fairness trade-off. In this example, we consider an uplink OFDMA-based cellular network and formulate an optimization problem that maximizes the EE of its worst case link subject to subcarrier assignment and power allocation constraints. In the figure, the number of users $K = 16$, number of subcarriers $N = 128$, power amplifier inefficiency factor $\xi_k = 18$, terminal's circuit power $P_k^C = 0.4$ W, user's rate requirement $R_k^{req} = 15$ b/s/Hz, and maximum transmit power $P_k^{max} = 0.2$ W for all k. Note that the EE of the best/worst link is obtained by saving the EE of the link who has the highest/lowest EE in each sample and then taking an average on 5000 of them.

How to develop joint resource allocation algorithms that reach the theoretical limits of the network EE and thus serve as benchmarks to evaluate performance of other heuristic algorithms is another challenge. Moreover, it is also necessary to develop cost-efficient and easy-to-implement algorithms with acceptable performance levels to solve these problems for practical applications.

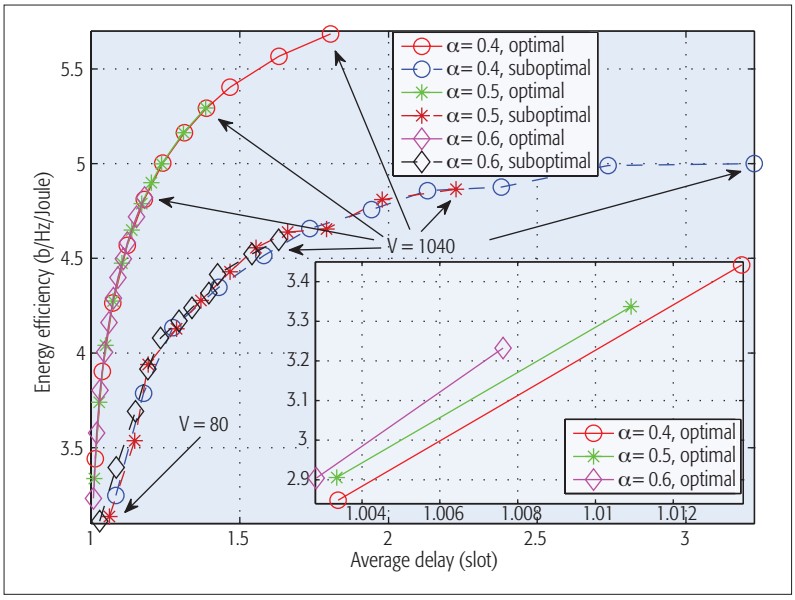

Figure 6. Illustration of the EE-delay trade-off. In this example, we consider a downlink single-MBS H-CRAN and maximize its network $EE_{long\text{-}term}$ subject to a queue length control constraint by jointly optimizing RRH operation and power allocation. In the figure, the traffic arrival rate $\lambda = 2.5$ b/slot/Hz, RRH's circuit power $P_n^c = 0.4$ W, number of RRHs $N = 8$, and number of users $M = 12$. In particular, $V \geq 0$ and $\alpha \in [0,1]$ are two control parameters introduced to adjust the EE-delay trade-off.

In what follows, we present a possible method to model and reveal the quantitative EE-delay trade-off.

To formulate EE and delay in a framework, we first need to shift from previously short-term to long-term models. In long-term formulations, random traffic arrivals can be enfolded to obtain a dynamic arrival-departure queue for each user, given as $Q_i(t+1) = \max[Q_i(t) - R_i(t), 0] + A_i(t)$, $\forall i$ [14]. Here, $A_i(t)$ and $Q_i(t)$ denote the amount of newly arrived data and queue length of user i at slot t, respectively. Note that the average delay can be characterized by queue length, as it is proportional to the queue length for a given traffic arrival rate from Little's Theorem.

Furthermore, it is also necessary to inject the concept of time into the EE definition $EE_{short\text{-}term}$ in order to bridge the EE and delay. One possible way to achieve this is to define the EE from a long-term average perspective, given by the ratio of the long-term aggregate data delivered to the corresponding long-term total power consumption in [14, Eq. 10]. For simplicity, we denote this kind of network EE definition by $EE_{long\text{-}term}$. From [10, 14], we know that $EE_{long\text{-}term}$ can also be seen as an extension of $EE_{short\text{-}term}$, because it degenerates to $EE_{short\text{-}term}$ if there are no time averages and expectations in $EE_{long\text{-}term}$. Then, by integrating the queue length control (i.e., delay control) and EE maximization into a framework, we can depict the EE and average delay simultaneously.

We utilize the above ideas to display the EE-delay trade-off in H-CRANs by formulating a stochastic optimization problem that maximizes the network $EE_{long\text{-}term}$ subject to a queue length control constraint through joint optimization of RRH operation and power allocation. Two algorithms, referred to as optimal and suboptimal, are developed to solve this problem. Figure 6 intuitively shows the EE-delay trade-off, where $V \geq 0$ and $\alpha \in [0,1]$ are two control parameters introduced in the model to adjust the EE-delay trade-off. Specifically, from Fig. 6, for the same V, the smaller α is, the better the EE, and the larger the average delay. In addition, for the same α, the bigger V is, the better the EE, and the larger the average delay. These observations together exhibit the EE-delay trade-off, which can be explicitly balanced by V and α. Hence, the long-term model can be used to tune the EE-delay trade-off via adjusting V and α. More clearly, α is used to confine the trade-off range between the EE and average delay (a small α gives a large range and vice versa) and V to tune the trade-off point between the EE and average delay (a small V yields a small delay but low EE and vice versa).

Although [13] found the EE-delay trade-off and [14] obtained an EE-delay trade-off of $[O(1/V), O(V)]$, the optimal EE-delay trade-off, that is, the optimal order for the average delay in V when the EE increases to the optimal by the law of $O(1/V)$, is still unknown. Moreover, [13, 14] focused on the average delay, and thus the obtained results therein are valid only for non-real-time traffic such as web browsing and file transfers. However, there are some other real-time applications, for example, voice and mobile video, in H-CRANs that impose hard-deadline (or maximum delay) constraints. It is thus worthwhile to study how to provision deterministic delay guarantees and improve the EE in the meantime. Moreover, in more realistic H-CRANs with both non-real-time and real-time traffic, it is also well worth investigating how to flexibly balance the EE-delay performance for each kind of traffic from a perspective of systematic design and further devise control algorithms. Potential techniques that can be used to settle these unresolved issues are stochastic optimization, dynamic programming, Markov decision process, queue theory, and stochastic analysis.

Conclusions

Under the triple drives of capacity enhancement, EE improvement, and communication ubiquity, H-CRANs have emerged as a promising architecture for future wireless network design. In this article, we have first exploited the features of H-CRANs to propose three green techniques and then particularly have focused on three fundamental trade-offs, namely EE-SE, EE-fairness, and EE-delay trade-offs. We have introduced the methods to model and analyze these trade-offs, presented open issues and challenges, and also provided some potential solutions. However, we are still at a very primary stage in these studies, and thus further investigations on exploitation of the high-dimension, flexible, and scalable architecture of H-CRANs are eagerly needed for a green future.

Acknowledgment

This work was supported in part by the National Science Foundation of China under Grants 61601192, 61601193, 61631015, and 61471163; the U.S. NSF under Grant CNS-1320664; the Major Program of the National Natural Science Foundation of Hubei in China under Grant 2016CFA009; and the Fundamental Research Funds for the Central Universities under Grant 2016YXMS298.

References

[1] Cisco Visual Networking Index: Global Mobile Data Traffic Forecast Update, 2016–2021, Cisco, Feb. 2017.
[2] IMT-2020 (5G) Promotion Group, "5G Vision and Requirements," white paper, May 2014.
[3] A. Fehske *et al.*, "The Global Footprint of Mobile Communications: The Ecological and Economic Perspective," *IEEE Commun. Mag.*, vol. 49, no. 8, Aug. 2011, pp. 55–62.
[4] China Mobile Research Institute, "C-RAN: The Road Towards Green RAN," white paper, Oct. 2011.
[5] M. Peng *et al.*, "Recent Advances in Cloud Radio Access Networks: System Architectures, Key Techniques, and Open Issues," *IEEE Commun. Surveys & Tutorials*, vol. 18, no. 3, 3rd qtr. 2016, pp. 2282–2308.
[6] M. Peng *et al.*, "Heterogeneous Cloud Radio Access Networks: A New Perspective for Enhancing Spectral and Energy Efficiencies," *IEEE Wireless Commun.*, vol. 21, no. 6, Dec. 2014, pp. 126–35.
[7] J. Wu, "Green Wireless Communications: From Concept to Reality," *IEEE Wireless Commun.*, vol. 19, no. 4, Aug. 2012, pp. 4–5.
[8] E. Oh *et al.*, "Toward Dynamic Energy-Efficient Operation of Cellular Network Infrastructure," *IEEE Commun. Mag.*, vol. 49, no. 6, June 2011, pp. 56–61.
[9] Y. Chen *et al.*, "Fundamental Trade-offs on Green Wireless Networks," *IEEE Commun. Mag.*, vol. 49, no. 6, June 2011, pp. 30–37.
[10] C. Xiong et al., "Energy- and Spectral-Efficiency Tradeoff in Downlink OFDMA Networks," *IEEE Trans. Wireless Commun.*, vol. 10, no. 11, Nov. 2011, pp. 3874–86.
[11] J. Tang *et al.*, "Resource Efficiency: A New Paradigm on Energy Efficiency and Spectral Efficiency Tradeoff," *IEEE Trans. Wireless Commun.*, vol. 13, no. 8, Aug. 2014, pp. 4656–69.
[12] C. Joe-Wong et al., "Multiresource Allocation: Fairness-Efficiency Tradeoffs in a Unifying Framework," *IEEE/ACM Trans. Net.*, vol. 21, no. 6, Dec. 2013, pp. 1785–98.
[13] F. Meshkati, H. Poor, and S. Schwartz, "Energy Efficiency-Delay Tradeoffs in CDMA Networks: A Game-Theoretic Approach," *IEEE Trans. Info. Theory*, vol. 55, no. 7, July 2009, pp. 3220–28.
[14] M. Sheng *et al.*, "Energy Efficiency and Delay Tradeoff in Device-to-Device Communications Underlaying Cellular Networks," *IEEE JSAC*, vol. 34, no. 1, Jan. 2016, pp. 92–106.
[15] O. Arnold *et al.*, "Power Consumption Modeling of Different Base Station Types in Heterogeneous Cellular Networks," *Future Networks and Mobile Summit*, June 2010, pp. 1–8.

Biographies

Yuzhou Li [M'14] (yuzhouli@hust.edu.cn) received his Ph.D. degree in communications and information systems from the School of Telecommunications Engineering, Xidian University, Xi'an, China, in December 2015. Since then, he has been with the School of Electronic Information and Communications (EIC), Huazhong University of Science and Technology (HUST), Wuhan, China, where he is currently an assistant professor. His research interests include 5G wireless networks, marine object detection and recognition, and undersea localization.

Tao Jiang [M'06, SM'10] (taojiang@hust.edu.cn) is currently a Distinguished Professor with the School of Electronic Information and Communications, HUST. He has authored or co-authored over 300 technical papers and five books in the areas of wireless communications and networks. He is the Associate Editor-in-Chief of *China Communications* and on the Editorial Boards of *IEEE Transactions on Signal Processing* and *IEEE Transactions on Vehicular Technology*, among others.

Kai Luo (kluo@hust.edu.cn) received his B.Eng. degree from the School of EIC, HUST, in 2006. Then he received his Ph.D. degree in electrical engineering from Imperial College London in 2013. In 2013, he joined the Institute of Electronics, Chinese Academy of Sciences. Since 2014, he has been an assistant professor with the School of EIC, HUST. His research interests are signal processing, MIMO communications, and heterogeneous networks.

Shiwen Mao [S'99, M'04, SM'09] (shiwen.mao@gmail.com) received his Ph.D. in electrical and computer engineering from Polytechnic University, Brooklyn, New York, in 2004. He is the Samuel Ginn Distinguished Professor and director of the Wireless Engineering Research and Education Center at Auburn University, Alabama. His research interests include wireless networks and multimedia communications. He is a Distinguished Lecturer of the IEEE Vehicular Technology Society and on the Editorial Boards of *IEEE Transactions on Multimedia* and *IEEE Multimedia*, among others.

In more realistic H-CRANs with both non-real-time and real-time traffic, it is also well worth investigating how to flexibly balance the EE-delay performance for each kind of traffic from a perspective of systematic design and further devise control algorithms.