Energy-Efficient NOMA Heterogeneous Cloud Radio Access Networks: Enabling Techniques and Challenges

Fuhui Zhou, *Member, IEEE*, Yongpeng Wu, *Member, IEEE*,
Rose Qingyang Hu, *Senior Member, IEEE*, and Yuhao Wang, *Senior Member, IEEE*

Abstract

Heterogeneous cloud radio access networks (H-CRANs) are envisioned to be a promising technology for the fifth generation (5G) wireless networks. H-CRANs enable users to enjoy diverse services with high energy efficiency, high spectrum efficiency, and low-cost operation, which is achieved by using cloud computing techniques. However, H-CRANs face many technical challenges due to massive user connectivity, increasingly severe spectrum scarcity and energy-constrained devices. These challenges may significantly decrease the user quality of experience if not properly tackled. Non-orthogonal multiple access (NOMA) schemes exploit non-orthogonal resources to provide services for multiple users and are receiving increasing attention for their potential of improving spectrum and energy efficiency in 5G networks. In this article a framework for energy-efficient NOMA H-CRANs is established. The enabling technologies for NOMA H-CRANs are surveyed, and the challenges to implement these technologies are discussed. This article also presents a performance evaluation related to the energy efficiency of the proposed H-CRANs.

Index Terms

Fuhui Zhou and Yuhao Wang are with the School of Information Engineering, Nanchang University, P. R. China, 330031 (e-mail: zhoufuhui@ncu.edu.cn, wangyuhao@ncu.edu.cn).

Yongpeng Wu is with Institute for Digital Communications, Universitüt Erlangen-Nürnberg, Erlangen D-91058, Germany (e-mail: yongpeng.wu@fau.de).

Rose Qingyang Hu is with Electrical and Computer Engineering Department, Utah State University, USA (e-mail: rose.hu@usu.edu).

The research reported in this article was supported by the Natural Science Foundation of China the National Natural Science Foundation of China (61261010, 61561034, 61661028, 61601347 and 61501356).

Heterogeneous cloud radio access networks, 5G, energy efficiency, non-orthogonal multiple access, mmWave, CR, massive MIMO.

I. INTRODUCTION

THE explosive increase of smart devices, such as smart phones, smart tablets, as well as the emerging broadband and high-rate services, such as augmented reality (AR), virtual reality (VR), and the massive number of devices constructing the Internet of Things (IoT), are in need of designing an energy-efficient communication system in order to achieve environmental friendly, green economic, and sustainable operations. Compared with the fourth generation (4G) systems, the fifth generation (5G) systems can provide 1000 times system capacity, 10 times spectral efficiency (SE), at least 100 times energy efficiency (EE), 1 ms latency and 100 times higher connectivity density [1]. As a promising new technology and architecture, heterogeneous cloud radio access networks (H-CRANs) have drawn great attention in both industry and academia. H-CRANS aim to achieve high flexibility, tremendous capacity, super EE, wide coverage and cost-effective operation [2]-[6] mainly by incorporating powerful cloud computing and visualization techniques into heterogeneous networks (HetNets).

Different from HetNets, in H-CRANs, base stations from different tiers (e.g., macro cells, micro cells, pico cells, femto cells) are decoupled into baseband units (BBUs) and remote radio heads (RRHs). All the BBUs construct a BBU pool in a cloud center, and RRHs are deployed close to user equipments (UEs). The BBU pool efficiently performs baseband signal processing (e.g., modulation, coding, radio resource allocation, media access control, etc.) through cloud computing and visualization techniques, while RRHs remotely conduct radio transmission/receiving processing and convert the radio signals to digital base band signals. H-CRANs can greatly increase the flexibility of the network architecture, improve the system spectrum efficiency, significantly reduce energy consumption and operational expenditures. Moreover, the user experience can be remarkably improved due to the reduced distance between RRHs and UEs and the RRHs association from different tiers.

It is envisioned that numerous multiple access (MA) technologies will be exploited in future H-CRANs in order to mitigate the inter and intra cell interference and to improve the spectrum efficiency and user experience. As a new MA technique, non-orthogonal multiple access (NOMA) has been identified as a promising candidate for significantly improving SE and EE of 5G mobile communication networks [7]. Unlike the conventional orthogonal MA (OMA) schemes, the NOMA can allocate nonorthogonal (e.g., power-domain NOMA, code-domain NOMA, sparse code multiple access, etc) resources to different users. The nonorthogonal feature enables NOMA techniques to have apparent advantages in supporting

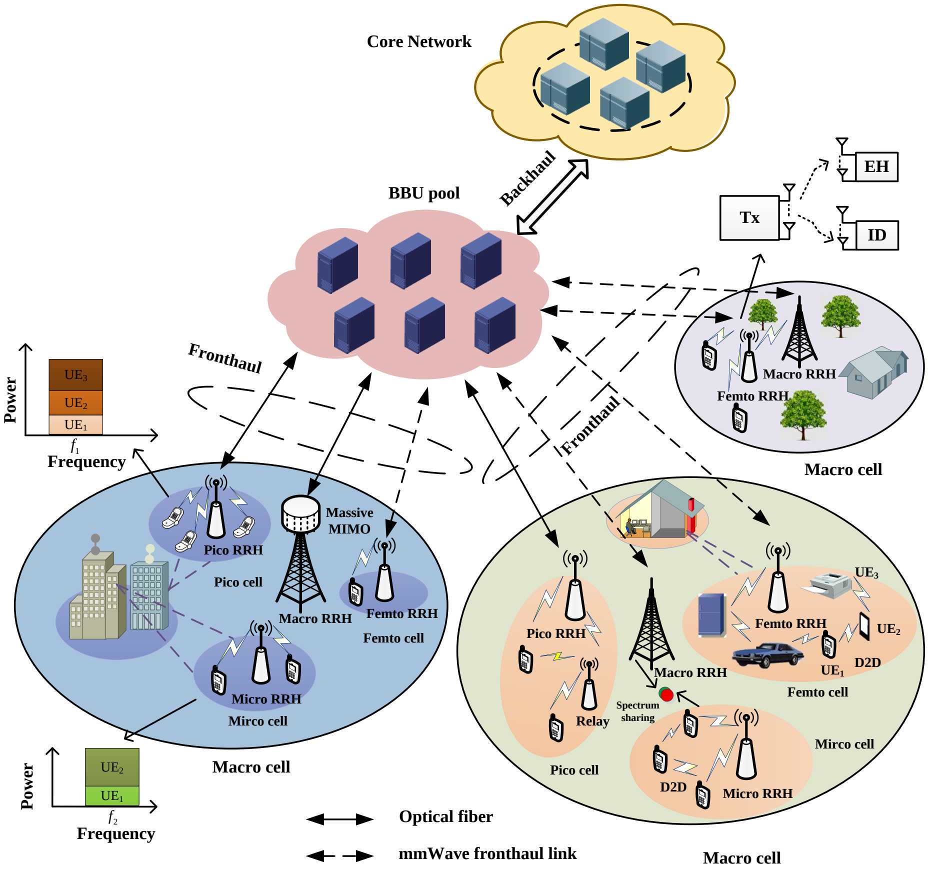


Fig. 1. The system architecture of NOMA H-CRANs.

high SE and EE, massive connectivity, and low transmission latency at the cost of introducing NOMA interference [8]-[10].

While there have been extensive research done for H-CRANs, no much study has been conducted for NOMA H-CRANs. In [2], an energy-efficient H-CRANs framework was established. It is shown that both EE and SE can be significantly improved by using H-CRANs. The authors in [3] discussed five key techniques based on cloud computing for H-CRANs. In [4], challenges for three promising resource allocation schemes were investigated in H-CRANs. Resource sharing in H-CRANs from three levels (spectrum, infrastructure, and network) was analyzed in [5]. None of these existing work on H-CRANs investigate the beneath MA techniques, which actually are of significant importance in H-CRANs for interference mitigation, EE and SE improvement and low-cast operation. Thus different from the previous work, this article discusses the enabling techniques and challenges for NOMA supported H-CRANs.

The rest of the article is organized in the following way. It first presents the H-CRAN architecture and highlights the NOMA implementation. Then it discusses the promising enabling techniques and challenges, particularly related to applying NOMA to H-CRANs. These techniques and challenges include massive multiple-input multiple-output (MIMO), cognitive radio (CR), wireless charging, millimeter-wave (mmW) communications, cooperative transmission and device-to-device (D2D) communications. Based on the presented NOMA H-CRAN framework, quality of expectation (QoE) is evaluated. The article concludes with Section V.

II. NOMA H-CRANS ARCHITECTURE

The system architecture of NOMA H-CRANs is shown in Fig.1. It consists of macro cells, micro cells, pico cells and femto cells. Macro cells or micro cells with high power RRHs provide large coverage areas but relatively low data rates (e.g., urban, suburban, or rural areas), while pico cells or femto cells with low power RRHs are deployed in small coverage areas requiring high data rate transmission (e.g., airports, stadiums, homes). Unlike the conventional HetNets, the radio resource allocation and media access control are performed in the BBU pool for H-CRANs. The links between RRHs to the BBU pool can be wired (e.g., optical fiber) or wireless (e.g., millimeter-wave). The selection is up to the operational cost and the wireless channel conditions since the cost of deploying optical fiber links could be high and the impact of the wireless channel conditions on the achievable capacity is critical. It is envisioned that both wire and wireless links will be exploited in the future H-CRANs.

In order to satisfy diverse service requirements, different applications and cells with different scales, and also to improve the operational efficiency, OMA and NOMA can coexist in the H-CRANs. Specifically, OMA is appropriate for femto cells, where a small number of UEs exist and real-time services with high data rates are required. Furthermore, in a small area like femto cell, there may not exist too much channel diversity, making OMA more suitable. On the other hand, NOMA can be a desirable candidate for cells that require a high connectivity density and frequent small data transmissions, e.g., in a large-scale shopping center. The combination of OMA and NOMA is appropriate to the scenarios where a massive UEs connectivity and high rates are both required, e.g., in an urban area. For example, the combination of orthogonal frequency division multiple access (OFDMA) with NOMA is adapted to multiple femto cells (e.g., multiple neighborhoods), which is achieved by using numbers of orthogonal subbands and each subband exploits NOMA to provide services for one femto cell. Moreover, different NOMA schemes and OMA schemes have their own appropriate application scenarios, which depend on the tradeoff between the achievable performance and implementation complexity. As an example, the power-domain NOMA scheme exploits the power domain to allow multiple users access simultaneously and requires successive interference cancellation (SIC) in the receiver. It can be used if the channel conditions from different

RRHs to UEs have sufficient diversity on pathloss and fading. The authors in [8] have comprehensively discussed the applied situations of different NOMA schemes.

III. ENABLING TECHNIQUES AND CHALLENGES

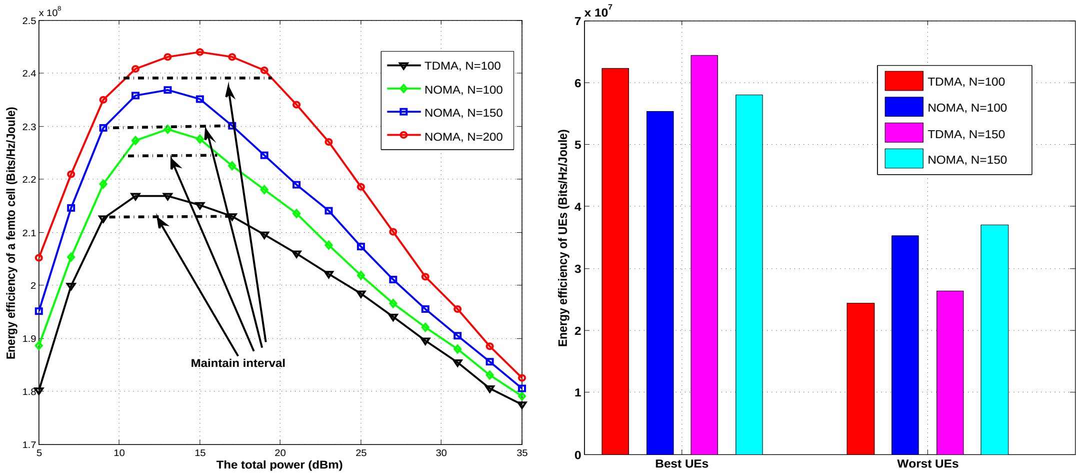
In this section, we discuss in details the enabling techniques for NOMA H-CRANs and the corresponding challenges. Then, in the following section, we present the performance evaluation on the UEs' QoE achieved by NOMA H-CRANs.

A. Combination Massive MIMO with NOMA

Applying NOMA techniques into H-CRANs causes extra interference among different cells and UEs due to the usage of non-orthogonal resources. Although the interference can be partially addressed by conducting baseband signal processing in the BBU pool using cloud computing techniques, the required communication overhead between diverse RRHs and the BBU pool are extremely high. One promising scheme is to use massive MIMO techniques, which equip RRHs with orders of magnitude more antennas (e.g., 100 or more). The combination of massive MIMO and NOMA can significantly improve the EE and SE of NOMA H-CRANs where only simple linear signal processing approaches are required [11]. Moreover, unlike in the conventional NOMA HetNets, multi-user detection (e.g., SIC for power domain NOMA, a message passing algorithm for code domain NOMA) can be moved into the BBU pool, which greatly simplifies the structure of RRHs and makes it possible for RRHs to be cost-efficiently deployed in a large-scale area. Furthermore, the combination of massive MIMO and NOMA can provide a large number of degrees of freedom and suppress the inter-cells interference, which are achieved by performing precoding for UEs' signals from different cells.

However, the application of massive MIMO into NOMA H-CRANs has several challenges to be addressed. The pilot contamination problem in massive MIMO may be severe due to the usage of non-orthogonal resources. Moreover, the design of an optimal precoding for suppressing inter-cell interference needs a large number of channel state information (CSI), which is extremely difficult due to the traffic load and delay constraints. The problem is even more severe when RRHs equipped with massive MIMO form cooperative communications in a hotspot area (e. g., lager emporiums, stadium). Furthermore, the hardware impairment of massive MIMO affects the performance of NOMA H-CRANs to a large extent.

Fig. 2 presents the comparison between EE achieved by using the power-domain NOMA scheme and EE obtained by using the time-division multiple access (TDMA) scheme in a downlink femto cell. The number of UEs is 5 and the number of downlink RRHs is 3. The constant power consumption of RRHs is 10 dBm. The number of antennas of each RRH is set as 100, 150 or 200. The distance between each



(a) (b)

Fig. 2. (a) The total EE of a femto cell with five downlink UEs versus the total transmit power of downlink RRHs under different numbers of antennas, $N = 100, 150, 200$; (b) The fairness comparison among five UEs under different numbers of antennas and different multiple access strategies, $N = 100$ and $N = 150$.

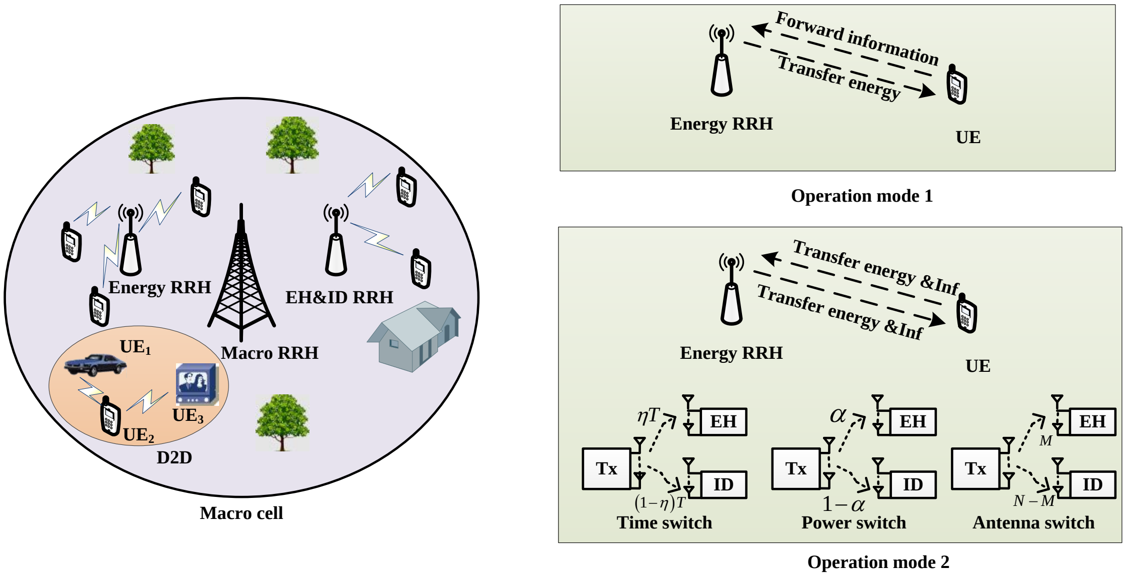
RRH is 100 m, and UEs are uniformly distributed. From Fig. 2(a) we can see that the power-domain NOMA scheme can achieve much better EE gains than the TDMA scheme. In practice, the maximum EE may not able to attain due to several factors such as channel fading, the power allocation strategy, interference uncertainty, etc.. The real achievable EE can surpass the prescribe threshold (such as the minimum EE requirement) with a good margin. This margin is defined as the “maintain margin”. It is found the value of maintain margin depends on the MA scheme and the number of antennas. The comparison of the fairness among five UEs between the power-domain NOMA scheme and the TDMA scheme is shown in Fig. 2(b). The total transmit power of RRHs is set to 14 dBm. It is observed that the power-domain NOMA scheme can provide a better fairness compared with the TDMA scheme. This is due to the fact that in the power-domain NOMA a larger power is allocated to the UEs with a worse CSI, which elevates the fairness among UEs.

B. Combination Cognitive Radio with NOMA

Cognitive radio (CR) is envisaged to overcome the increasingly severe spectrum scarcity problem due to fixed spectrum allocation strategy, the ubiquitous wireless devices access, and the ever increasing capacity requirements. CR enables unlicensed UEs to use the spectrum resources of licensed UEs in an opportunistic mode, spectrum-sharing mode, or in a sensing-sharing mode. SE can be greatly enhanced by using CR techniques. There are several standards proposed for CR, such as IEEE 802.22, IEEE 802.11 TGaf, IEEE 802.16h and IEEE 802.19. Recently, the exploitation of NOMA in CR has been considered in 3GPP-LTE. It has been proved that both SE and EE can be improved by using NOMA in CR networks compared with these achieved by using the conventional OMA [12]. Moreover, the combination of NOMA with CR can be facilitated in H-CRANs, where spectrum sensing, spectrum assignment, power allocation and interference control can be efficiently performed in the BBU pool by using the powerful cloud computing techniques. An additional advantage of applying NOMA with CR in H-CRANs is that multiple unlicensed UEs can be served simultaneously, which is unrealistic for OMA unless multiple spectrum bands can be provided.

Although the combination of NOMA with CR in H-CRANs is attractive, there are impediments for realizing it. The most challenging problem is how to control the interference caused by the simultaneous access of multiple unlicensed UEs in order to protect the quality of service (QoS) of licensed UEs. Specifically, when an opportunistic access mode is used, multiple unlicensed UEs can simultaneously access the spectrum bands of licensed UEs using NOMA techniques only when licensed UEs are detected to be inactive. In this case, spectrum sensing algorithms with high performance are required in the BBU pool, and the sensing duration is of significant importance to obtain a good tradeoff between the protection of the QoS of licensed UEs and the total throughput of unlicensed UEs. When a spectrum-sharing mode is applied, the power allocation strategy is crucial since the unlicensed UEs can coexist with licensed UEs as long as the interference to the licensed UEs is tolerable. Finally, if the sensing-sharing mode is used, which is a hybrid mode of the opportunistic access and spectrum sharing, the joint design of spectrum sensing algorithms, sensing duration and the power allocation strategy is vital since the interference comes from two aspects, one from the missing detection of licensed UEs and the other from the coexistence of licensed UEs and unlicensed UEs when licensed UEs are detected to be active.

The selection of the CR operation mode in NOMA with CR in H-CRANs depends on the tradeoff between the performance and the implementation complexity of the BBU pool. Particularly, the sensing-sharing mode can be selected if a high implementation complexity can be affordable in the BBU pool. The reason is that it can provide a good performance due to the flexible power allocation strategy based on the spectrum sensing result, that is, a high transmit power level of dowlink RRHs or uplink UEs can be used when the licensed UEs are detected to be inactive, whereas a low transmit power level is chosen when the licensed UEs are detected to be active. If the BBU pool prefers a low implementation complexity, the combination of NOMA with CR in H-CRANs is available when the spectrum sharing mode is selected due to its facilitated implementation. The opportunistic access mode can be selected in the case that the interference caused to the licensed UEs should be strictly controlled (e.g., IEEE 802.22 for the TV bands).



Macro cell

Operation mode 1

Operation mode 2

Fig. 3. Operation modes and wireless charging receiver structure. T denotes the frame duration; η represents the time switching factor; α denotes the power splitting factor; N denotes the total number of antennas of the receiver and M represents the number of energy harvesting antennas.

C. *Wireless Charging with NOMA*

In H-CRANs, a major limitation of the network performance is the presence of battery driven energy-constrained UEs and RRHs (e.g., energy-limited wireless sensor, mobile phone, electrical vehicles). Wireless charging techniques that enable energy-constrained devices to replenish energy from the surrounding electromagnetic radiations are deemed promising solutions to conquer this limitation [13]. An advantage of exploring wireless charging techniques in H-CRANs is the high power transfer efficiency due to the short distance among RRHs and UEs. Moreover, when wireless charging and NOMA techniques are applied in H-CRANs, NOMA UEs (say group 1)that are close to RRHs can also be energy sources for those NOMA UEs (say group 2) far away from RRHs but close to group 1 UEs.

As shown in Fig. 3, there are two operation modes for RRHs when RRHs are identified as the energy sources for NOMA UEs, namely, wireless charging mode, and simultaneous wireless charging and information transmission mode. In the first mode, RRHs only provide energy supply for NOMA UEs in the downlink and forward information from NOMA UEs to the BBU pool in the uplink. In the second mode, RRHs can transfer energy and transmit information to NOMA UEs simultaneously in the downlink, and harvest energy from NOMA UEs and forward information from NOMA UEs to the BBU pool simultaneously in the uplink. The structure of RRHs and UEs is simple in the first mode whereas a higher hardware implementation at the RRHs and UEs is required in the second mode. For the second mode, in order to practically realize simultaneous wireless charging and transmission, a received signals is split into two parts, e.g., one for energy harvesting (EH), and one for information decoding (ID). Based on the splitting domain (time, power, antenna), there are different protocols to achieve signal splitting. For the time-domain protocol, the RRH receivers or UEs receivers switch in time between EH and ID. For the power-domain protocol, the received signals are split into two signal streams with different power levels by equipping a power splitting component in the receivers, e.g., one stream for EH and one for ID. Finally, the antenna domain protocol requires RRH receivers or UEs receivers to equip antenna array, and the antenna array is divided into two groups, one group for realizing EH and the other group for achieving ID. The selection of the signal splitting protocol depends on the affordable hardware implementation ability and the scale requirement in the receiver. For example, if the receiver has a low hardware implementation ability and requires a small size (e.g, wireless sensor), the time-domain protocol is more appropriate.

Due to the inherent characteristics of wireless charging techniques, H-CRANs with wireless charging components are susceptible to eavesdroppers, which may intercept confidential transmitted information. The security issue is very challenging for wireless charging in H-CRANs, especially when wireless charging techniques is combined with NOMA to provide service for multiple UEs. The reason is that eavesdroppers may disguise themselves as NOMA UEs and enjoy the service provided by RRHs. The other challenge for the combination of wireless charging techniques with NOMA is the hardware implementation, which needs a high complex structure in the receivers.

D. Millimeter-Wave Communications with NOMA

The availability of 28, 38, 60 and 73 GHz for communications has attracted investigations on mmW communications, which operate on mmWave bands between 30 and 300 GHz. mmWave communications has been considered as an important technique for 5G due to its potential for providing ultra-wide bands services, enabling large numbers of miniaturized antennas (≥ 32) to be deployed in a small dimension, and allowing massive connectivity of different devices with diverse service requirements [14]. It can be foreseen that the combination of mmWave communications with NOMA in H-CRANs can further improve SE and accommodate more device connectivity. This combination is appropriate for the urban outdoor environment due to the high UEs density, small cell radii (about 100-200 m), and lower mobility. mmWave is also a desirable technology for the fronthaul links between RRHs and the BBU pool, which need wide-band services and tremendous capacity.

Although the combination of mmWave communications with NOMA in H-CRANs has many exciting advantages, there are several obstacles required to be overcome. Most of the existing investigations have studied the application of mmWave to point-to-point communications (e.g., the fronthaul link between one RRH and the BBU pool). However, since NOMA is proposed for multiple UEs, novel mechanisms are required to achieve this combination in H-CRANs. Besides, in order to deal with the vulnerability of shadowing and the intermittent connectivity, which are caused by the ultra-high frequency of mmW signals, beamforming technology in mmWave with NOMA is critical yet very challenging. The reason is that the performance gains achieved by using mmW communications highly depend on the directional transmissions and the transmission directions as multiple NOMA users have to be simultaneously considered.

E. Cooperative Transmission with NOMA

Cooperative relay transmission has been recognized as an effective scheme to enlarge the service coverage, improve the system capacity and EE by exploiting spatial diversity, especially when the transmit power and size of devices are limited. The idea of the combining cooperative relay transmission with NOMA in H-CRANs is to allow UEs with weak channel condition consider UEs with strong channel condition as relays, which help transmit information to RRHs. This combination is more promising in H-CRANs since the inter-cell and intra-cell interference can be effectively controlled even when there are massive connectivity requirements. Moreover, the linear signal processing and relay selection can be performed in the BBU pool, which can simplify the structure of RRHs and improve EE due to the decrease of the power consumption of RRHs. Additionally, this combination may play a very prominent role in H-CRANs where mmW communications are exploited since the vulnerability of shadowing can be overcome by decreasing the transmission distance.

The challenges of the combining cooperative relay transmission with NOMA in H-CRANs come from the implementation complexity and the CSI overhead. Moreover, the implementation complexity and the overhead increase with the number of NOMA UEs and the relay RRHs. It is impractical to achieve this combination when there are massive NOMA UEs and relay RRHs. One promising solution is to select and pair nearby NOMA UEs and RRHs to realize cooperative transmission. The selection of the pairing scheme depends on the channel conditions, the relay transmission scheme (e.g., amplify and forward, decode and forward) and the transmit power limit of UEs and RRHs. The pairing procedure can be performed in either a centralized or a distributed manner. For the centralized pairing manner, the BBU pool collects all the required CSIs and location information from all UEs and RRHs, and then pairs them into clusters. For the distributed pairing manner, RRHs group nearby UEs as a cluster based on the CSIs and location information from UEs. Generally speaking, a higher performance gains can be obtained by using the centralized pairing manner at the cost of high overhead, whereas the distributed pairing manner has advantages in the flexibility and mobility but a lower performance gain.

F. D2D Communications with NOMA

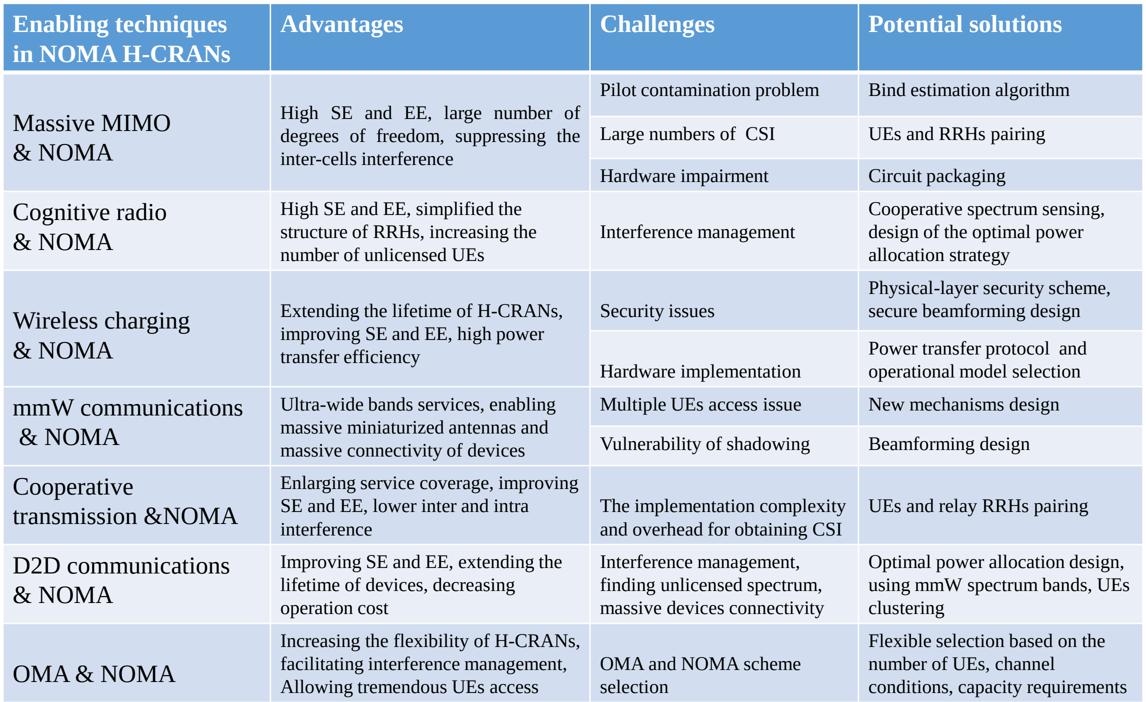
With the advance of smart grid networks, intelligent homes and integrated transportation systems, future 5G systems need to face with enormous communications among devices. Different from the conventional cellular network where the communication among devices must go through a base station even when the distances among devices are very short, D2D communications enable devices to directly connect with other devices. It can extensively improve SE, EE and extend the lifetime of devices. In H-CRANs, the combination of NOMA with D2D communications is an attractive scheme that can achieve the direct communication among multiple UEs. For example, as shown in Fig. 1, UE_1 can use the combination of NOMA with D2D communications to connect with UE_2 and UE_3. The direct connection between UE_1 and UE_3 can not achieve due to the poor channel condition, but UE_1 can connect with UE_3 with the help of UE_2. In this case, UE_2 with strong channel conditions is assumed to have the ability of full-duplex D2D communications.

There are two ways to realize the combination of NOMA with D2D communications in H-CRANs, namely, the in-band mode and the out-band mode. The in-band mode enables D2D UEs to share the same spectrum with H-CRANs links, whereas D2D UEs exploit the unlicensed spectrum for the out-band mode. The motivation of using the in-band mode is that the interference of the unlicensed spectrum is uncontrollable and the interference caused to H-CRANs links is tolerable. For the out-band mode, mmWave spectrum bands are appropriate candidates. These two modes have their respective pros and cons, and neither is deemed absolutely better than the other one. The challenge of achieving the combination in the in-band mode is the interference management, and thus an adaptive power allocation strategy is of utmost importance. For the out-band mode, the impediment of the combination may be the severe shadowing, which is quite normal in mmWave spectrum bands. Another challenge for the two modes is the requirement of a massive devices access due to the unaffordable overhead.

Finally, Fig. 4 is given to comprehensively summarize the advantages, challenges and potential solutions of each enabling technique.

IV. PERFORAMNCE EVALUATION

In future H-CRANs, different service requirements from diverse devices need to be satisfied, such as throughput, EE, network delay, etc.. In this case, the optimal design of H-CRANs is extremely difficult



Enabling techniques in NOMA H-CRANs	Advantages	Challenges	Potential solutions
Massive MIMO & NOMA	High SE and EE, large number of degrees of freedom, suppressing the inter-cells interference	Pilot contamination problem	Bind estimation algorithm
		Large numbers of CSI	UEs and RRHs pairing
		Hardware impairment	Circuit packaging
Cognitive radio & NOMA	High SE and EE, simplified the structure of RRHs, increasing the number of unlicensed UEs	Interference management	Cooperative spectrum sensing, design of the optimal power allocation strategy
Wireless charging & NOMA	Extending the lifetime of H-CRANs, improving SE and EE, high power transfer efficiency	Security issues	Physical-layer security scheme, secure beamforming design
		Hardware implementation	Power transfer protocol and operational model selection
mmW communications & NOMA	Ultra-wide bands services, enabling massive miniaturized antennas and massive connectivity of devices	Multiple UEs access issue	New mechanisms design
		Vulnerability of shadowing	Beamforming design
Cooperative transmission &NOMA	Enlarging service coverage, improving SE and EE, lower inter and intra interference	The implementation complexity and overhead for obtaining CSI	UEs and relay RRHs pairing
D2D communications & NOMA	Improving SE and EE, extending the lifetime of devices, decreasing operation cost	Interference management, finding unlicensed spectrum, massive devices connectivity	Optimal power allocation design, using mmW spectrum bands, UEs clustering
OMA & NOMA	Increasing the flexibility of H-CRANs, facilitating interference management, Allowing tremendous UEs access	OMA and NOMA scheme selection	Flexible selection based on the number of UEs, channel conditions, capacity requirements

Fig. 4. Comprehensive survey of enabling techniques for NOMA H-CRANs

to achieve since the optimal values for the performance metrics (e.g., throughput, EE, network delay) change from service to service. To this end, service providers increasingly focus on UEs' quality of experience (QoE) instead of network QoS. According to the definition given by the ITU, QoE is the overall acceptability of an application or service from the perspective of UEs. It is generally evaluated by using a mean opinion score (MOS) [15]. There are several MOS models for different services. In this paper, the common MOS given by eq. (6) in [15] is used, which is appropriate for video services.

The simulation parameters are set as these used in Section III. Fig. 5 is given to show the minimum MOS among UEs versus the minimum EE requirement of each UEs in a femto cell with five downlink UEs under different MA schemes and different numbers of antennas. The simulation results are obtained under the object that the minimum MOS of UEs is maximized subject to the minimum EE requirement of each UEs. The number of antennas is set as $N = 128$ or $N = 150$. The total transmit power of RRHs is set as 20 dBm. It is seen that the minimum MOS among UEs achieved by using NOMA is larger than that obtained by using TDMA. The reason is that NOMA can provide capacity gains compared with TDMA and the MOS is related to the capacity. This indicates that a better QoE can be obtained by using NOMA. However, it is also observed that the MOS decreases with the increase of the EE requirement. It can be explained by the fact that the MOS model given by eq. (6) in [15] can only reflect the user

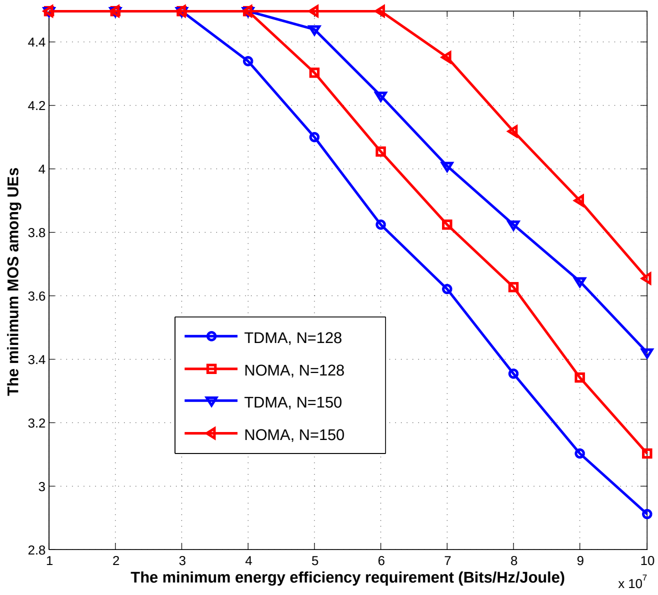


Fig. 5. (b) The minimum MOS among UEs versus the minimum EE requirement in a femto cell with five downlink UEs under different MA schemes and different numbers of antennas, $N = 128$ and $N = 150$.

experience in terms of QoS. It cannot reflect the user experience from the EE perspective. To the authors' best knowledge, an appropriate model that can reflect UEs' QoE from the EE perspective has not been found up to now. It still be an open and interesting issue to be studied.

V. CONCLUSIONS

The article proposed a new framework for NOMA H-CRANs that aims to provide a high energy efficiency, a massive UEs connectivity, and high system capacity. Several promising enabling technologies for NOMA H-CRANs were presented. These technologies include combining massive MIMO, CR, wireless charging, mmW communications, cooperative transmission, and D2D communications with NOMA in H-CRANs. The implementation challenges and potential solutions were discussed. The preliminary performance study shows that the proposed NOMA H-CRANs framework and the presented enabling technologies are promising in achieving high energy efficiency in the advanced 5G wireless communication systems.

REFERENCES

[1] J. G. Andrews *et al.*, "What will 5G be?," *IEEE J. Sel. Areas Commun.*, vol. 32, no. 6, pp. 1065-1082, Jun. 2014.

[2] R. Q. Hu and Y. Qian, "An energy efficient and spectrum efficient wireless heterogeneous network framework for 5G systems," *IEEE Commun. Mag.*, vol. 52, no. 5, pp. 94-101, May 2014.

[3] 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, pp. 126-135, Dec. 2014.

[4] H. Dahrouj, *et al.*, “Resource allocation in heterogeneous cloud radio access networks: Advances and challenges,” *IEEE Wireless Commun.*, vol. 22, no. 3, pp. 66-73, June 2015.

[5] M. A. Marotta *et al.*, “Resource sharing in heterogeneous cloud radio access networks,” *IEEE Wireless Commun.*, vol. 22, no. 3, pp. 74-82, June 2015.

[6] G. Wang *et al.*, “Acquisition of channel state information in heterogeneous cloud radio access networks: Chanlleges and research directions,” *IEEE Wireless Commun.*, vol. 22, no. 3, pp. 100-107, June 2015.

[7] Z. Ding *et al.*, “Application of non-orthogonal multiple access in LTE and 5G network,” *IEEE Commun. Mag.*, to be appeared.

[8] L. Dai *et al.*, “Nonorthogonal multiple access for 5G: Solutions, challenges, opportunities, and future research trends,” *IEEE Commun. Mag.*, vol. 53, no. 9, pp. 74-81, Sep. 2015.

[9] Z. Ding *et al.*, “A general MIMO framework for NOMA downlink and uplink transmission based on signal alignment,” *IEEE Trans. Wireless Commun.*, vol. 15, no. 6, pp. 4438-4454, June 2016.

[10] Y. Zhang *et al.*, “Energy-efficient transmission design in Non-orthogonal multiple access,” *IEEE Trans. Veh. Technol.*, to be appeared, 2016.

[11] L. Lu *et al.*, “An overview of massive MIMO: Benefits and challenges,” *IEEE J. Sel. Topics Signal Process.*, vol. 8, no. 5, pp. 742-758, Oct. 2014.

[12] Y. Liu *et al.*, “Non-orthogonal multiple access in large-scale underlay cognitive radio networks,” *IEEE Trans. Veh. Technol.*, 2016, to be appeared.

[13] X. Lu *et al.*, “Wireless networks with RF energy harvesting: A contemporary survey,” *IEEE Commun. Surveys Tuts.*, vol. 17, no. 2, pp. 757-789, 2015.

[14] S. Rangan *et al.*, “Millimeter wave cellular networks: Potentials and challenges,” *Proc. IEEE*, vol. 102, no. 3, pp. 366-385, Mar. 2014.

[15] M. Rugelj *et al.*, “Novel cross-layer QoE-aware radio resource allocation algorithms in multiuser OFDMA systems,” *IEEE Trans. Commun.*, vol. 62, no. 9, pp. 3196-3208, Sept. 2014.