Please cite the Published Version

Enquiries:

If you have questions about this document, contact openresearch@mmu.ac.uk. Please include the URL of the record in e-space. If you believe that your, or a third party's rights have been compromised through this document please see our Take Down policy (available from https://www.mmu.ac.uk/library/using-the-library/policies-and-guidelines)

EE Optimization for Downlink NOMA-based Multi-Tier CRANs

Ziad Qais Al-Abbasi, Khaled M. Rabie, Member, IEEE, and Daniel K. C. So, Senior Member, IEEE

Abstract—**Non-orthogonal multiple access (NOMA) is increasingly becoming very attractive in cloud radio access networks (CRANs) to further boost the overall spectral efficiency, connectivity and, capacity of such networks. This paper addresses optimizing the energy efficiency (EE) for the downlink of a NOMA-based two tiers CRAN. The stochastic geometry represented by Poisson Point Process (PPP) distribution is used to decide the number and locations of the base stations (BSs) in each tier within the coverage area. A numerical optimal solution is obtained and compared against a proposed subgradient solution, as well as another proposed unoptimized solution based on the false positioning method. For comparison purposes, two other power allocation techniques are presented to allocate different powers to various BS categories; one allocates the power to each BS based on their relative distances to the cloud-based central station and the other is the bisection based scheme. Two simulation scenarios are presented to examine the performance of the two-tier NOMA-CRANs with NOMA is adopted as the multiple access of each tier in both cases. The first scenario considers heterogeneous CRAN (NOMA-HCRAN) case by using two different BS categories in each tier, namely, the macro-BSs and the RRHs. The second scenario considers a homogeneous CRAN (NOMA-CRAN) case by using the RRHs in both tiers but each tier has different frequency layer to prevent cross tier interference. Simulation results show the promising performance gain can be achieved with the proposed techniques relative to the existing approaches. More specifically, it was illustrated that the proposed subgradient based NOMA CRAN offers better performance than the proposed false positioning based NOMA CRAN, which is in turn better than the existing techniques, in particular, the bisection and the distance based NOMA-CRAN.**

Index Terms—**Non-orthogonal multiple access (NOMA), cloud radio access network (CRAN), false positioning method, resource allocation, energy efficiency.**

I. INTRODUCTION

EFFECTIVE performance, low power consumption, smart traffic management, and high reliability, represent increasingly urgent imperatives that are mandatory to have in the design of the next generation networks. Therefore, it is necessary to maintain an efficient power management paradigm in combination with high demand for increased capacity, better coverage, and high throughput. There is also a compelling need to reduce energy cost and minimize the carbon footprint. These challenges could be overcame by an innovative Cloud-RAN architecture with low powered small cells to reduce the consumed power [1], [2]. This is due to the fact that major part of the consumed power is in the radio access networks, in particular, at BSs. They consume more

Ziad Qais Al-Abbasi, Middle Technical University (MTU), Baquba Technical Institute, Baghdad, Iraq. (e-mail: ziad.al-abbasi@mtu.edu.iq).

Khaled Rabie, Department of Engineering, Manchester Metropolitan University, Manchester, UK. (e-mail: k.rabie@mmu.ac.uk).

Daniel K. C. So, School of Electrical and Electronic Engineering, The University of Manchester, Manchester, UK. (e-mail: d.so@manchester.ac.uk).

than 75% of the total power required by a standard mobile network system [2]–[4].

Cloud radio access networks (CRAN) and non-orthogonal multiple access (NOMA) are regarded as two key technologies in finding a solution to the expected 5G challenges. Both are considered as vital to achieve high spectral efficiency (SE) and energy efficiency (EE) [5]–[7]. To effectively densify the mobile data network and to tackle a number of recent network challenges, it is essential to upgrade the current architecture of small-cell RAN (SC-RAN), for instance, CRAN. SC-RAN offers a scalable, cost effective, and self-organizing solution to solve a number of the network expansion issues. CRAN consist of Remote Radio Heads (RRHs) that contains include the radio equipment, amplification, filtering, and the antenna. The RRHs are connected to BBU through Common Public Radio Interface (CPRI) or Open Base Station Architecture Initiative (OBSAI) links. The combination of SC-RAN along with Heterogeneous Networks (HetNets) could be used to provide a mixed-network deployment scenarios.

The concept of CRAN suggests employing the low powered remote radio heads (RRHs) instead of the conventional large coverage area base stations (BSs). These RRHs are responsible for processing the radio frequency while other tasks such as resource allocation and signal processing are performed by the central processing unit at the cloud [5], [8]. CRAN architecture is promising as it allows several BSs to be aggregated and coordinated through the cloud based central station (CCS) [9], [10]. It also assists in radio capabilities establishment across a tier of RRHs with different types of services and tackling issues, such as, handover control and interference management [4], [9], [10]. CRAN architecture contributes in improving the reliability of the cloud-based services by playing the role of providing a cooperative interface for the non-homogeneous resources of its cloud-related applications [11].

Fig. 1 shows a two-tier heterogeneous CRAN (HCRAN) platform, which is regarded as efficient in allocating the power and frequency radio resources and tackling the interference issue in multi-cell environment. According to CRANs structure, the function of the macro-BSs is to continuously provide coverage and network control. The control operations include handover management and interference control to macro cells [1], [8]. The baseband processing unit (BBU) in the CCS coordinates the whole network performance aspects, such as EE and SE, establishing interference cancellation techniques among RRHs. In addition, all the baseband signal processing is performed by the BBUs through cloud computing technologies, where the BBUs act as computing servers whilst the high data rates are provided by the RRHs [1]–[3].

On the other hand, NOMA is proven to be an effective solution to meet the expected huge traffic demand and to scale up the number of serviceable users in 5G networks, for both, uplink and downlink. NOMA permits multiple users to occupy the same time and frequency resource. This is how it differs from traditional multiple access techniques such as

time division multiple access (TDMA) and frequency division multiple access (FDMA) in which users occupy their distinctive resources [12]–[14]. NOMA users can be multiplexed in the power domain or in the code domain [12], [15], [16]. In code domain NOMA, the users share the resources using spreading sequences [12], [13], [15]. On the other hand, in power domain NOMA, multiplexing is applied through power allocation depending on the channel gain differences among the multiplexed users. As a result, NOMA and CRAN could be tailored into applications where it fulfill the expected demands and challenges of 5G networks [12], [16], [17]. The available resources in the CRAN-NOMA architecture is considered to be the same as in the rest of cellular networks. It include computational resources and radio resources. The computational resources are usually needed at the BS and it consist of the processing power, bandwidth, memory, time, and data storage [18]. Managing the resources in the BBU pool assist in computational resources management. The radio resources, on the other hand, include the scarce radio frequency spectrum. Its management achieved through algorithms and strategies that establish dynamic channel allocation, transmit power control, joint optimization, spectrum and/or cache management [18].

Therefore, it is expected that the cooperation between CRAN and NOMA has the potential to fulfill the predicted high data rate demands of the future wireless networks as NOMA allows a complete share of the bandwidth resource; thanks to superposition coding in the power domain [3], [19]
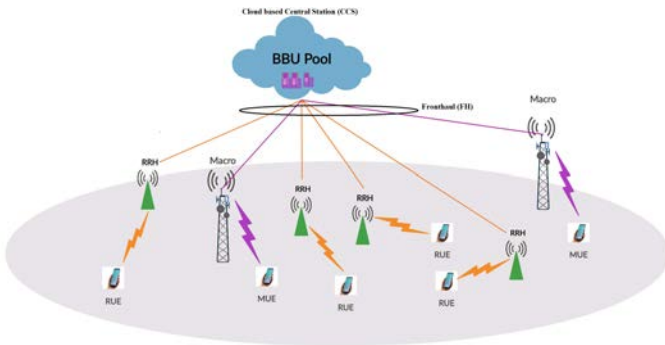

Fig. 1. Two-tier HCRAN platform, where MUE: Macro user equipment; RUE: RRH user equipment.

A. Prior Works and Motivation

Recent literature works have proposed NOMA as an energy and spectral efficient air interface for the next generations of wireless networks [7], [17], [19].

Generally speaking in regards to the multi-cell distribution throughout the coverage area, several schemes were applied to model the BS's locations, however, according to literature works, stochastic techniques offers the most realistic representations. For instance, the author in [20] presented a survey about using stochastic geometry to emulate the BSs distribution in multi-tier networks and they also classified the most common and applicable stochastic techniques and their challenges; these techniques included Grid model, Poisson Point Process (PPP), Hard core point process, binomial point process, and the cluster point process, their findings highlighted the vital role that the stochastic geometry plays in realistic modeling the multi-tier networks. In addition, the authors in [21] considered a homogeneous, multi-cell scheme based on PPP which was established to consider different BSs

placement strategies such as grid and Poisson distributions. Similarly, the authors in [22] presented an analytical representation for multi-tier heterogeneous cellular networks, were the Poisson distribution and the grid schemes were used to model each tier. The authors of both works concluded that the Poisson distribution is more realistic in modeling cellular networks inside a geographic area with variable number of BSs, hence, it is more reliable and close to reality to obtain the results using PPP for BS distribution rather than other distributions and that is the main reason behind adopting PPP in this article.

On the other side, it is worth mentioning that the idea of power domain NOMA is not new, actually it has been around for years. Nevertheless, it was idle for a while until advances in its processing appeared recently that made NOMA practically implementable. These advances rejuvenated a new literature research wave on NOMA performance. Some of this works discussed the combination of NOMA with multiple-input, multiple-output (MIMO) technology to enhance the overall system throughput [23]–[27]. In addition, a considerable number of works considered incorporating the NOMA concept into device-to-device (D2D) communication to fully exploit the bandwidth resources and boost the achievable spectral efficiency [28]–[31]. The detection techniques and the structure of NOMA receiver was also examined in a considerable number of works to highlight the SIC applicability and SIC decoding order at NOMA receiver [32]–[34]. Finally, the performance of NOMA uplink was also studied by a number of authors [35]–[38], these works aimed on clarifying how the user communicates with the serving BS and to highlight the difference between the uplink and the downlink. There are also other aspects of NOMA that have been discussed in literature to achieve higher performance; all this attention grabbed by NOMA is a clear evidence of its huge potential. Hence, this paper is a continuation in the same direction, however, with a new contribution.

The combination of both NOMA and CRAN was introduced in some literature works [5], [9], [10], [39], [40]. In particular, the authors of [41] derived a closed-form expression for the outage probability of a user pair in a NOMA-CRAN system. In addition, [40] considered the performance of wireless backhaul in a downlink Heterogeneous-CRAN (HCRAN). The authors investigated the optimal number of cells to be supported by HCRAN. Besides investigating the power consumption of different kinds of BSs, the authors in [9] studied NOMA-HCRAN and addressed determining the maximum number of BSs that each cell could support taking into account limited power resource. The authors in [10] investigated the problem of sharing the same time-frequency resource in a downlink of NOMA-HCRAN. They proved that NOMA-based HCRAN outperforms by four times its orthogonal based counterparts. Other related works are summarized in Table I.

B. Main Contribution and Organization

All these works besides the potential offered by the combination of NOMA and CRAN stood as the motivation backbone to produce this paper. The main contributions of this paper are as follows:

- Firstly, we propose two power allocation techniques for NOMA-HCRAN combination, namely, the optimized subgradient based and unoptimized false position method (FPM). The previous works on NOMA-CRAN did not take into account the effect of user's quality of service

Table I
RELATED LITERATURE WORKS REGARDING NOMA AND CRAN.

Authors [Reference]	The studied NOMA - CRAN aspect	Outcome	limitations
[5]	Green NOMA - CRAN	NOMA could assist CRAN to achieve higher EE than OMA	Slightly complex approach
[25]	Considered the outage probability of relay assisted MIMO-NOMA system	NOMA offers higher outage than OFDMA	The allocated power depended on UFPA plus it was unoptimized
[39]	The combination of NOMA and HCRAN	NOMA cooperation with HCRANs could greatly enhance the overall system EE	
[42]	Addressed the performance of NOMA-HCRAN	NOMA approach helps HCRAN to remove the inter-tier interference	The proposed approach is not simple as it decompose the joint optimization problem into two sub-problems.
[43]	A consideration of NOMA with SIC in practice	In terms of throughput, NOMA outperforms OFDMA by around 30%	Unoptimized power allocation
[44]	The combination of Fog radio access network (FRAN) and NOMA	NOMA could improve the overall network latency.	A number of challenges must be met before NOMA-FRAN combination could work
[45]	Analysis framework for NOMA-CRAN	NOMA could enhance the spectral efficiency of CRAN	two users are scheduled in the same resources
[46]	Performance analysis of NOMA-HCRAN	Higher imperfect SIC factor degrades the EE	
[47]	NOMA - cell-edge user throughput	NOMA can enhance the cell-edge user throughput while achieving fairness	
[48]	Introduced the concept of green NOMA system	NOMA offers energy efficient paradigm, however its application encounter several challenges	

(QoS) requirements on the power assigned to that BS serving those users.

- Secondly, unlike the previous works on NOMA-CRAN, the power allocated at each BS type will depend on the QoS requirement of that particular BS, which represents the summation of the minimum rate required by the users being served by that particular BS, i.e., its load. This contributes in enhancing the overall achievable throughput, and hence, improving the overall network EE.

- The obtained results confirm that the proposed power allocation schemes can achieve higher EE than the conventional techniques and offer close performance trends to the optimal one. Furthermost, applying the combination of NOMA with CRAN can achieve larger multiuser diversity gains and BS densification gains.

- Finally, two scenario cases are considered in this paper, where each scenario considers the overall performance by deploying different BSs types in each tier and NOMA is applied within that tier. The first scenario examines the performance of two types of BSs, in particular, the macro-BS and the RRHs; this is referred to throughout this paper as HCRAN scenario. The second scenario considers the overall performance when all the serving BSs are of the same type, which are chosen to be RRHs. In this scenario, the RRHs are still deployed in two tiers and NOMA is applied among these RRHs within their respective tier; however, each tier is allocated different frequency resources to avoid cross tier interference. The aim of these two scenarios is to address the drawbacks and the outcomes of replacing the high-transmit powered macro-BS tier with a lower-transmit powered RRHs.

The rest of this paper is organized as follows. Section II presents the considered NOMA-based system model. Section III presents the formulated EE optimization problem and the obtained solutions, as well as the proposed power allocation schemes. In Section IV, the simulation scenarios and the performance evaluation results are discussed, and finally, Section V presents the conclusion of the paper.

II. SYSTEM MODEL

The considered system consists of a τ-tiers downlink of NOMA-based HCRAN where each tier has a different type

of BSs and at the same time, different frequency layer; this is to suppress the inter-tier interference in the downlink. This is applicable in several scenarios, including air interface in smart cities, underlying the connection of a smart health system, industrial internet, wireless vehicle-networks and internet of things (IoT) infrastructure.

Each tier in the modeled system is assumed to have its own radio resource blocks and NOMA is applied within each tier. In other words, each tier has its own bandwidth and all tier BSs shares it. For a given radio resource block in the macro-BS tier, it is assumed that there are K single-antenna users being served by each macro-BS. On the other hand, in the RRH-tier, each RRH is assumed to be associated with a one single-antenna user.

Intra tier interference will exist since NOMA is applied within each tier; however, it will be removed by the application of the successive interference cancellation (SIC). SIC details will be skipped to avoid repetitive text as it has been deeply studied in a number of works, such as [7], [32], [33], [49]. Each tier has β_i BSs distributed in the coverage area according to the PPP with density λ_i. PPP is applied as it provides more realistic emulation of BSs deployment and closer to the actual scenarios. Without loss of generality, the considered model includes one macro tier, and the additional multiple tiers are for all RRH tiers. Each user in the i-th tier receives a combined signal z_i from all the BS in that respective tier. Lets consider the macro-tier and the RRH tier separately, then the superimposed signal $z_{i,b}$ in the i-th RRH tier is expressed as

$$z_{i,b} = \sqrt{P_{i,b}} x_{i,b}, \tag{1}$$

while that of the macro-tier is $z_1 = \sum_{k=1}^{K} \sqrt{P_{1,k}} x_{1,k}$ with the subscript 1 because only one macro-tier is considered in this model. Note that $P_{i,b}$ and $x_{i,b}$ refer to the transmission power and the transmitted signal of the b-th BS in the i-th tier, respectively.

On the other hand, the signal received by a user associated with the b-th BS in the i-th RRH tier is given by

$$y_{i,b} = \sum_{b=1}^{\beta_i} h_{i,b} z_{i,b} + n_{i,b}, \tag{2}$$

where $h_{i,b}$ represent the channel effect between the b-th BS and its respective user in the i-th tier, whilst the variable $n_{i,b}$ represents the noise power affecting them. It is worth mentioning that in case of macro-tier, the expression in (2) is $y_1 = \sum_{k=1}^{K} h_k z_k + n_k$. To examine the power allocation for the BSs in the i-th tier, let P_{T_i} denote the total transmission power that is available at the CCS for the BSs in the i-th tier, i.e.,

$$P_{T_i} = \sum_{b=1}^{\beta_i} P_{i,b}. \tag{3}$$

Overall, the total power consumed in the CRAN-downlink is [40]:

$$P_{tot} = \sum_{i=1}^{\tau} \left[\beta_i \left(P_b^{(C)} + P_b^{(FH)} \right) + \sum_{b=1}^{\beta_i} P_{i,b} \right], \tag{4}$$

where $P_i^{(C)}$ is the power consumed by each BS's circuit in the

i-th tier, whilst $P_b^{(FH)}$ denotes the front-haul (FH) consumed power that links the CCS to each BS in both tiers. This power is consumed in two parts, the power consumed in the aggregation switch of the CCU and the power consumed in the downlink interfaces of the wireless switches, and it can be given as [9], [46], [50]

$$P_b^{(FH)} = P_{int} + \frac{u_b P_b^{(SWmax)} + (1 - u_b) \frac{T_b^{(SW)}}{T_b^{(SWmax)}} P_b^{(SWmax)}}{S_b^{(int)}}, \tag{5}$$

where P_{int} denotes the consumed power at the interface in the aggregation switch, u_b stands for a weighting factor of the relative interaction between the power consumption of the switch ($P_b^{(SWmax)}$) and the amount of power in respect to $T_b^{(SW)}$, where $T_b^{(SW)}$ refers to the traffic quantity that passes within the switch while $T_b^{(SWmax)}$ is the maximum traffic that the switch can occupy. Finally, $S_b^{(int)}$ represents the number of interfaces in each switch. On the other hand, the achievable throughput by a single RRH is expressed as

$$R_{i,b} = W_T \log_2 \left(1 + \gamma_{i,b}^k \right), \tag{6}$$

where $\gamma_{i,b}^k$ stands for the signal-to-interference plus noise ratio (SINR) of the k-th user associated with the b-th RRH in the i-th tier, and it is given as $\gamma_{i,b}^k = \frac{P_{i,b} \left| h_{i,b}^k \right|^2}{\sum_{m=1}^{b-1} P_{i,m} \left| h_{i,b}^k \right|^2 + N_0 W_T}$. In addition, W_T represents the total bandwidth per each tier, and N_0 is the noise power spectral density. $\left| h_{i,b}^k \right|^2 = \xi |g_{i,b}|^2 d_b^{-\upsilon}$ is the channel gain which encompasses the Rayleigh flat fading $|g_{i,b}|^2$, the Log-normal shadowing ξ effect, while $d_b^{-\upsilon}$ is the distance from each RRH to the CCS with υ being the path loss exponent. As we considered one active user per each RRH, the upper-script ^{k} will be removed to ease the readibility of the equations.

III. EE OPTIMIZATION: PROBLEM FORMULATION AND THE PROPOSED SOLUTION

From (6) it can be concluded that the overall throughput is

$$R = \sum_{i=1}^{\tau} \sum_{b=1}^{\beta_i} R_{i,b}. \tag{7}$$

Then, the overall EE of NOMA-HCRAN is

$$EE = \frac{R}{\sum_{i=1}^{\tau} \left[\beta_i \left(P_i^{(C)} + P_i^{(FH)} \right) + \sum_{b=1}^{\beta_i} P_{i,b} \right]}, \tag{8}$$

which will be considered as the objective function of the formulated CRAN's EE optimization problem which can be expressed as

$$\underset{P_{i,b}}{maximize} \quad EE \tag{9}$$

$$Subject \ to \quad \sum_{i=1}^{\tau} \sum_{b=1}^{\beta_i} P_{i,b} \leq P_t \tag{10}$$

$$P_{i,b} \geq 0, \ \forall i, b \tag{11}$$

$$R_b \geq \Phi_b. \tag{12}$$

The constraint in (10) stands for the total transmission power constraint and (11) guarantees positive power allocation. It is worth to mention that the power $P_{i,b}$ denotes the power that the CCS use to transmit data and communicate with each BS in each tier. Hence, the constraint given in (10) is limiting the summation of this power of each BS within all tiers not to exceed the P_t, which is the total available power at the CCS. On the other hand, (12) denotes the minimum rate constraint per each BS, where Φ_b denotes a predefined minimum rate. From (9) to (12), it is obvious that the nature of the objective function is non-convex, fractional and non-linear. To solve such a problem, the objective function can be transformed into a subtractive by using the Dinkelbach approach [51], [52]. The idea of Dinkelbach is that for every objective function with a fractional form, there exists an equivalent subtractive form, keeping in mind that both forms have the same exact optimal solution [51], [53]. Applying the Dinkelbach approach encompasses introducing a weight factor α into the resulted objective function [51], [53], which would be expressed as

$$\Gamma(P_{i,b}, \alpha) = \chi(P_{i,b}) - \alpha \Lambda(P_{i,b}), \quad (13)$$

where $\Gamma(P_{i,b}, \alpha)$ represents the transformed objective function in terms of the optimizing variable $P_{i,b}$ and the introduced weight α. χ is the numerator of the original fractional objective function, which denotes the overall sum rate, and Λ is the denominator of the original objective function which is represented by the total consumed power in the CRAN-downlink. It is worth mentioning that $\chi(P_{i,b})$ acts as the overall CRAN system revenue gained by the information transmission while $\alpha \Lambda(P_{i,b})$ stands for the overall system tariff, which is the consumed power in all CRAN components. Finally, α represents a negative weight whose optimal value balances the tariff and the revenue and it affects the total consumed power within the CRAN. It must be noted that α could be set as a constant [52]. After applying the Dinkelbach approach, the formulated optimization problem is expressed as

$$\begin{align} maximize \quad & \Gamma(P_{i,b}, \alpha) \quad &(14) \\ Subject\ to \quad & (10), (11), (12). \quad &(15) \end{align}$$

A. The proposed Subgradient based Power Allocation Solution

By exploiting fractional programming transformation, the objective function could be converted into a parameterized objective function by Dinkelbach concept and, after that, a two-loop iterative algorithm is applied to overcome the non-convexity issue. Consider $P_{i,b}^*$ as the optimal solution of the problem in (14) and (15), from (13). As $\alpha = \frac{\chi(P_{i,b}^*)}{O(P_{i,b}^*)}$, according to the Dinkelbach scheme, the solution of the optimization problem in (14) and (15) is reached by finding the roots of $\Gamma(P_{i,b}^*, \alpha) = 0$. After that, the obtained value of $\Gamma(P_{i,b}^*, \alpha)$ should be tested, if positive then α is less than its optimal value, and if negative it means that α is larger than the optimal value. Nevertheless, the optimal value of α is found when α equals zero as the maximum EE is reached when $\chi(P_{i,b}^*) - \alpha^* \Lambda(P_{i,b}^*) = 0$. By examining the optimization techniques presented in [53], it is possible to express the Lagrangian function of the optimization problem in (14) and (15) as

$$\Gamma(P_{i,b}, \alpha) = W_T \sum_{i=1}^{\tau} \sum_{b=1}^{\beta_i} log_2(1 + \gamma_{i,b}) - \alpha(P_{tot})$$
$$-\kappa \left(\sum_{i=1}^{\tau} \sum_{b=1}^{\beta_i} P_{i,b} - P_T \right) - \Theta \sum_{b=1}^{\beta_i} \left(\Phi_b - W_T log_2(1 + \gamma_{i,b}) \right), \quad (16)$$

where κ and Θ represent the Lagrange multipliers. It is possible to simplify this expression as

$$\Gamma(P_{i,b}, \alpha) = W_T \sum_{i=1}^{\tau} \sum_{b=1}^{\beta_i} log_2(1 + \gamma_{i,b}) - \alpha(P_{tot}) -$$
$$\Theta \sum_{b=1}^{\beta_i} \left(\Phi_b - W_T log_2(1 + \gamma_{i,b}) \right) - \kappa \left(\sum_{i=1}^{\tau} \sum_{b=1}^{\beta_i} P_{i,b} - P_T \right). \quad (17)$$

Lets consider the expression in (17) for all BSs except the one with the highest channel gain (i.e., the β_i-th BS). The derivative with respect to $P_{i,b}$ is obtained as follows

$$\frac{d\Gamma}{dP_{i,b}} = \frac{W_T|h_{i,b}|^2 + W_T|h_{i,b}|^2\Theta}{\left(\sum_{m=b+1}^{\beta_i} P_{i,m}|h_{i,b}|^2 + N_0 W_T \right)(1 + \gamma_{i,b})} - \alpha - \kappa. \quad (18)$$

On the other hand, the derivative of (17) for the BS with highest channel gain is obtained as

$$\frac{d\Gamma}{dP_{i,\beta_i}} = \frac{|h_{i,\beta_i}|^2}{N_0(1 + \gamma_{i,\beta})} - \frac{|h_{i,\beta}|^2 \Theta_{\beta_i}}{N_0(1 + \gamma_{i,\beta})} - \alpha - \kappa. \quad (19)$$

To obtain the optimal solution, the necessary waterfilling expressions are found to be

$$P_{i,b} = \left[\frac{(\Phi_b + \Theta_b)}{\Phi_b(\alpha + \kappa)} - \frac{\sum\limits_{m=b+1}^{\beta_i} P_{i,m}|h_{i,b}|^2 + N_0 W_T}{|h_{i,b}|^2} \right]^+ \quad (20)$$

$$P_{i,\beta_i} = \left[\frac{\left(\Phi_{\beta_i} - \sum\limits_{b=1}^{\beta_i - 1} \varphi_b \right)}{\Phi_{\beta_i}(\alpha + \kappa)} - \frac{N_0 W_T}{|h_{i,\beta_i}|^2} \right]^+. \quad (21)$$

By applying the subgradient approach, the Lagrangian multipliers are updated using the following expressions

$$\kappa_b^{(j+1)} = \left[\kappa_b^{(j)} - \Theta^{(j)} \left(P_T - \sum_{i=1}^{\tau} \sum_{b=1}^{\beta_i} P_{i,b} \right) \right]^+ \quad (22)$$

$$\Theta_b^{(j+1)} = \left[\Theta_b^{(j)} - \Upsilon^{(j)} \left(\sum_{b=1}^{\beta_i} \left(\Phi_b - W_T log_2 \left(1 + \gamma_{i,b} \right) \right) \right) \right]^+,$$
(23)

where $\kappa^{(j)}$ and $\Theta^{(j)}$ are small step sizes to be updated at each iteration and chosen to be $0.1/\sqrt{j}$ [53]. In brief, the solution of the optimization problem is depicted in Algorithm 1.

Algorithm 1 The proposed subgradient-power allocation method.

- Initialize the max. tolerance Δ and the max. No. of iterations I_{max}
- Set the iteration index $j = 0$ and the initial weight value$\alpha_0 = 0$
- While —$\Gamma \left(P_{i,b}, \alpha \right)$— $\geq \Delta$ or $j \leq I_{max}$ do
    - For $i = 1$ to τ
        * For $b = 1$ to β_i
            · Initialize $P_{i,b} = 0$, $P_{i,\beta_i} = 0$, $\kappa_b^{(j)} = 0.01$, $\Theta_b^{(j)} = 1$,
            · While κ_b and Θ_b are not convergent, do
                · Determine $P_{i,b}$ and P_{i,β_i}, from (20) and (21), respectively.
                · update κ_b and Θ_b using (22) and (23), respectively.
            · end while
            · return $P_{i,b}^*$ and P_{i,β_i}^*
            · Solve (13) using known α_i and the obtained power P_{i,β_i}^*
        * end for
    - end for
    - Obtain $|\Gamma \left(P_{i,b}, \alpha \right)|$ and $\alpha_{j+1} = \frac{A\left(P_{i,b}^*\right)}{O\left(P_{i,b}^*\right)}$.
    - $j = j + 1$
- end while and Return $(P_{i,b}, \alpha)$

B. The proposed False Position Method (FPM)

FPM is exploited to calculate an appropriate transmission power per each BS subject to a predetermined level of rate requirements. We first modify the conventional false position technique by eliminating the negative values and as such we name the method as the FPM. Based on the triangular shapes depicted in Fig. 2, where the vertical axis represents the required data rate per each BS (i.e., RRH and macro in this paper) axis and the horizontal one represents the power, the method is derived as follows: Keeping in mind that FPM works by finding the best solution value between two boundaries, upper and lower. In the context of this paper, these boundaries are the lower and upper power values of the BS's transmission power, denoted as P_l and P_u, respectively.

Then, suppose that the range for the BS transmission power is varying from $P_l = 0$ dBm (0.001 W) to $P_u = P_{bs}$, where $P_{bs} = \frac{P_T}{\beta_i}$; while R_{pl} and R_{pu} represent respectively the lower and upper bit rates for the BS which are calculated based on the two power bounds P_l and P_u. Now, the FPM will determine a new power value, P_{fpm} that is within these bounds range and achieves a suitable bit rate for each BS such that the BS achieves the required target rate.

Next, by drawing a straight line from both R_{pu} and R_{pl} towards P_{fpm}, then two triangles would be formed, which are ABC and DCE as illustrated in Fig. 2. By applying the rules for triangle similarity between triangle 1 and triangle 2, we can determine the value for P_{fpm} as

$$\frac{DE}{AB} = \frac{CE}{CB} \longrightarrow \frac{R_{pl}}{R_{pu}} = \frac{P_{mfp} - P_l}{P_u - P_{fpm}}$$
(24)

$$R_{pu}.P_{mfp} - R_{pu}.P_l = R_{pl}.P_u - R_{pl}.P_{fpm}$$
(25)

$$P_{mfp}.(R_{pu} + R_{pl}) = R_{pl}.P_u + R_{pu}.P_l$$
(26)

$$P_{fpm} = \frac{R_{pl}.P_u + R_{pu}.P_l}{R_{pu} + R_{pl}},$$
(27)

and simply, P_{fpm} represents the final power value that should be assigned to the BS. The full algorithm steps are listed in Algorithm 2.
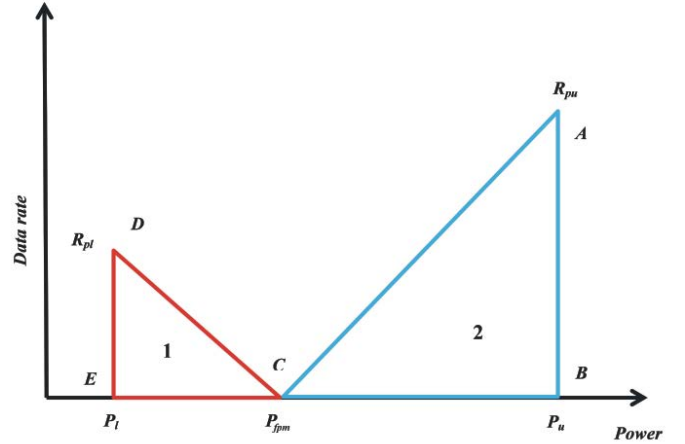

Fig. 2. The steps of deriving the FPM root.

The applicability of the proposed FPM approach is feasible as it merely represent a modified version of the false position method that is applied in root finding, as well as image and signal processing applications.

C. Comparison with Other Methods

To quantify the performance gains of the proposed techniques, we compare it against existing power allocation methods. These include the distance-based method which encompasses assigning the BS power depending on the distance between the BS and the CCS. In particular, if this distance is less than %20 of the radius, then the allocated power is %40 of the P_{bs}, which is $P_{bs} = \frac{P_T}{\beta_i}$ and if the distance is higher than %20 of the radius but less than %50 of the radius, then the allocated power is about %35 percent of P_{bs}. Finally, if this distance is higher than %50 of the radius, then the allocated power is %25 of P_{bs}. Secondly, we consider the bisection method, which, as the name implies, evaluates the function at the middle of an interval and replaces whichever limit has the same or higher value. The steps of applying the bisection are similar to the procedure in Algorithm 2, with the exception that (27) is replaced with $P_{i,b} = \frac{P_u + P_l}{2}$. The bisection method continues to halve the size of the interval until it eventually converges to an acceptable power value. However, when an
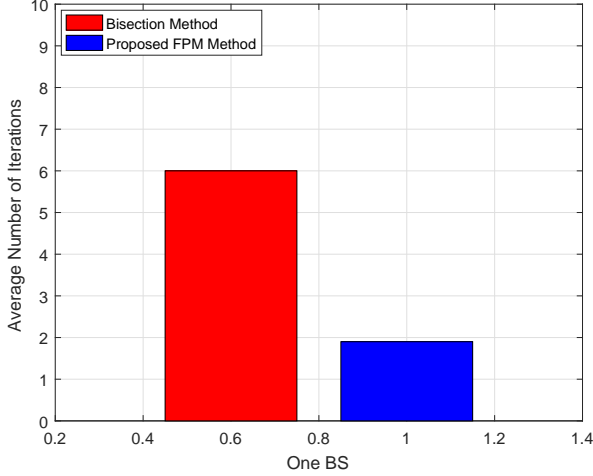
Fig. 3. Comparison between the FPM method and the Bisection method in terms of the number of iterations.

Algorithm 2 The Proposed FPM approach.

- Initialize the following parameters: $P_u = P_{bs}$ dBm, $P_l = 0$ dBm, $\epsilon ¿ 0$;
- For $b = 1$ to β_i do
  - Calculate the achievable data rate R_{pl} and R_{pu} using (6), for each given value of P_l and P_u, according to (27);
    * If $R_{pl} ¿ R_{target}$ then: Set $P_{fpm} = P_l$ dBm ;
    * else if $R_{pu} ¡ R_{target}$ then: Set $P_{fpm} = P_u$ dBm
    * else
      · while $P_u - P_l ¿ \epsilon$ do
      · Calculate P_{fpm} according to (27)
      · Calculate $R_{i,b}$ according to (6)
        · If $R_{i,b} ¿ R_{target}$ then: Update $P_u = P_{fpm}$;
        · else: Update $P_l = P_{fpm}$;
        · end if
      · end while
    * end if and set $P_{i,b} = P_{fpm}$;
- end for and Return $(P_{i,b}, \alpha)$

interval contains more than one solution, the bisection method could become slow. Also, even if the interval contains a single solution, the bisection method converges to that solution but sometimes the process can be slow as it has to examine each half interval to reach this solution. On the other hand, the proposed FPM has the advantage of searching along the range between the upper and lower limits of the BS power, where, the bisection method, it also depends on the two boundary values.

In Fig. 3, a comparison has been made between the FPM and bisection methods in terms of the average number of iterations required to reach the target power value. It is obvious from Fig. 3 that the FPM method requires, on average, about 2 iterations to reach the target value, while the bisection method requires on average 6 iterations to achieve the same goal. This occurs because the FPM takes into account the relative magnitudes of R_{pu} and R_{pl} unlike bisection which just uses the midpoint of P_l and P_u; that means the bisection method needs more time to investigate each half inside the interval $[P_l, P_u]$. In addition, FPM takes into account the interval $[P_l, P_u]$ as well as R_{pu} and R_{pl}, which makes it more accurate than the bisection method which focuses only on the interval $[P_l, P_u]$.

IV. PERFORMANCE EVALUATION

The considered scenario emulates a circular coverage area of diameter $= 2000m$ that has $\tau = 2$ tiers, namely, the RRHs tier and macro-BSs tier each has β_i BSs distributed following the PPP with density λ_i in each tier. The consumed power by the macro-BS is 1350W, and that by the RRH is 754.8W, the rest of the parameters are listed in Table II [50].

As mentioned earlier, the considered scenario examines two cases each with different BSs types in each tier, namely; heterogeneous and homogeneous CRANs. In the heterogeneous case, the simulated scenario includes two different types of BSs: macro-BSs and RRHs. The BSs in each tier use different frequency layer to that in the other tier to avoid cross tier interference. Nevertheless, NOMA is still applied within each tier.

In the second scenario, all the considered BSs are RRHs, even though they are of the same type, the RRHs in the second scenario are also deployed in two tiers and each tier uses

different frequency resources as compared to its counterpart to avoid cross tier interference. Intra tier interference will exist as the RRHs within each tier share the same bandwidth resources thanks to NOMA. Compared to the heterogeneous case, the macro-BSs are replaced with RRHs to reduce the required transmission power by the macro-BSs as compared to the RRHs which would contributes in reducing the power consumption of the overall network, and hence boosting the achievable EE performance.

Table II
SIMULATION PARAMETERS.

Parameter Name	Value
Path Loss Exponent (υ)	3.76
No. of tiers (τ)	2
$P_b^{(SWmax)}$	300W
P_{int}	1W
power consumed by each macro-BS	1350W
power consumed by each RRH	754.8W
coverage area of diameter	2000m
Shadowing standard deviation	8 dB
N_0	-174 dBm / Hz
I_{max}	50
$S_b^{(int)}$	24
u_b	0.5
$T_b^{(SW)}$	1Gbps
$T_b^{(SWmax)}$	24Gbps
Φ_b	2 Mbps
W_T	20MHz

Fig. 4 illustrates the performance of the first scenario where a heterogeneous network is simulated using macro and RRH tiers. In particular, the RRHs-tier has a density of λ which is double the density of its macro-tier counterpart that is simulated at a density of 0.5λ. The transmission power of these tiers, on the other hand, is simulated where the transmission power of the macro-BS is about 5-times that of

the RRHs tier. This figure shows that, in general, the trends of the compared schemes are decreasing gradually as the transmission power increases. This is a clear evidence that increasing the transmission power is not an effective measure to boost the achievable EE as each BSs tend to fulfill the required QoS level and neglect the extra (out of need) power which is reflected as inefficient use of the available resources.

It is also clear that the proposed Sub-NOMA-HCRAN offers a near optimal performance and is the best performing scheme relative to the other simulated techniques. It is also obvious that the proposed FPM performs better than the bisection-based HCRAN and the other compared techniques. This is because the FPM takes into account the relative magnitudes of R_{pu} and R_{pl} unlike bisection which just uses the midpoint of P_l and P_u; that means the bisection method needs more time to investigate each half inside the interval $[P_l, P_u]$. In addition, FPM considers the interval $[P_l, P_u]$ besides R_{pu} and R_{pl}, which makes it more precise than the bisection method.

Fig. 5. Case 2: The homogeneous scenario case: The achieved EE against various BS densities with $\lambda_1 = \lambda$ and $\lambda_2 = 0.5\lambda$ in each tier with $P_{T_1} = P_{T_2} = 35$ dBm.
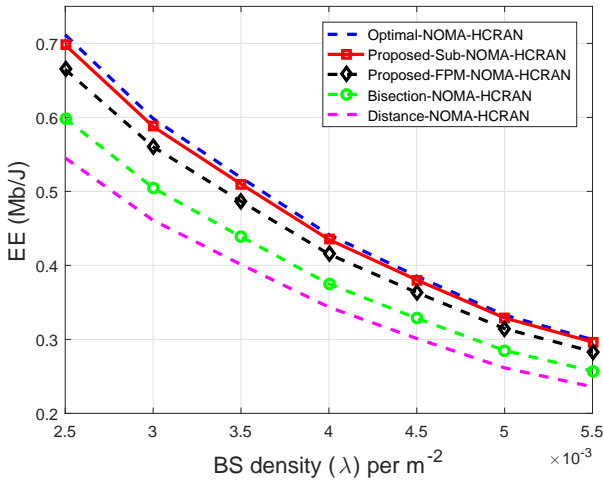

Fig. 4. Case 1: The heterogeneous scenario case: The achieved EE against various BS densities with $\lambda_1 = \lambda$ and $\lambda_2 = 0.5\lambda$ in each tier with $P_{T_1} = 35$ dBm and $P_{T_2} = 5P_{T_1}$.

Fig. 5 illustrates the performance of homogeneous case against various BS densities. The superiority of the compared power allocation techniques is similar to that of Fig. 4. By examining Figs. 4 and 5, it becomes clear that the heterogeneous case has a slight advantage over the homogeneous one. This is caused by the high transmission power that each macro-BS has compared to the RRH.

The trends of EE against various levels of P_T are depicted in Fig. 6. The overall trends shows that EE increases in proportion to P_T. The competition between the compared schemes is the same as in Fig. 4, where the proposed Sub-NOMA-HCRAN is very close to the optimal scheme and offers a better performance than all other techniques. This figure also shows that the proposed FPM is better than the bisection and distance-based methods, and despite offering lower performance behavior than the proposed Sub-NOMA-HCRAN, the proposed FPM scheme has less complexity.

The performance of the homogeneous counterpart of Fig. 6 is depicted in Fig. 7. These two figures confirm the slight privilege that the heterogeneous case has over the homogeneous case. From the homogeneous case perspective, this slight performance gain might be a considerable price to pay in order to earn flexible BS deployment (i.e., by deploying
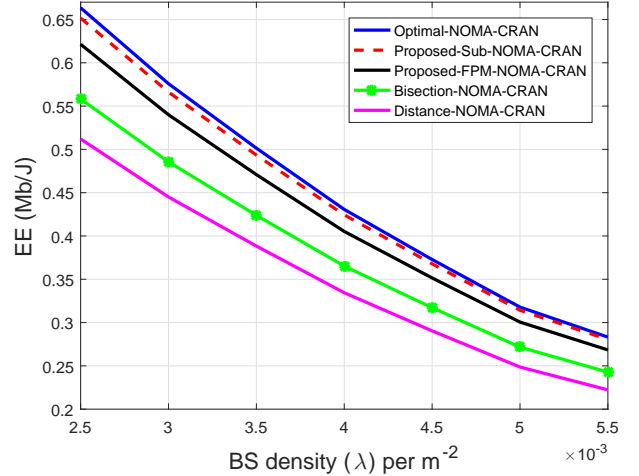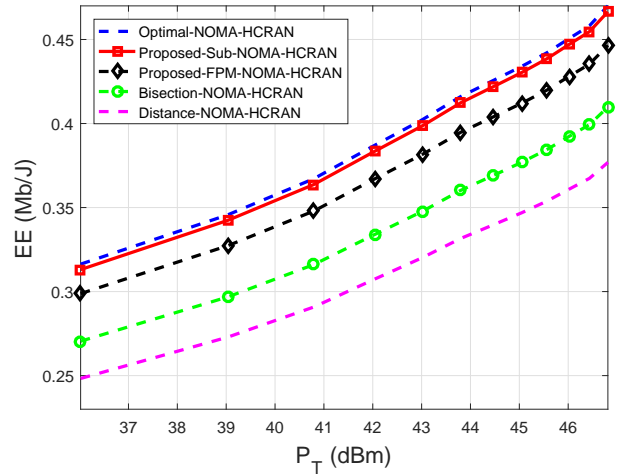

Fig. 6. Case 1: The heterogeneous scenario case: The achievable EE against various P_T levels with $P_{T_1} = P_T$, $P_{T_2} = 5P_T$, $\lambda_1 = 5 \times 10^{-3}\, m^{-2}$ and $\lambda_2 = 0.5\lambda_1$.

small-sized RRHs as compared to the macro-BS). In addition, RRHs cause less cross tier interference because of their lower transmission power.

Next, as a reflection of the quality of service (QoS), the trade-off between EE and the minimum rate requirement is depicted in Figs. 8 and 9, respectively. For a given value of the minimum rate requirement, it is noticeable from these figures that all schemes achieve increasing EE at low rate requirements. In addition, EE increases gradually in the low rate requirement region, this is because many users receive enough transmission power to comfortably meet their rate requirements, hence, there will be no need for extra energy consumption. On the other hand, at high QoS requirements, EE of all schemes start to decrease gradually. This is due to the fact that at high QoS (high rate requirements), more transmit power will be allocated to the BSs (in each tier) to achieve such minimum QoS, which leads to more energy consumption.

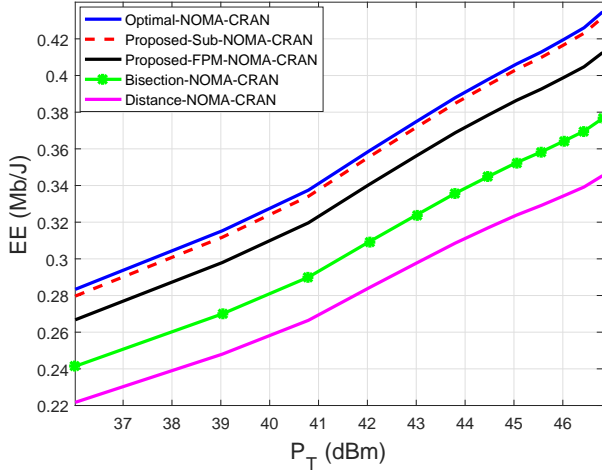As compared to Fig. 8, Fig. 9 proceeds in proving that

Fig. 7. Case 2: The homogeneous scenario case: The achievable EE against various P_T levels with $P_{T_1} = P_{T_2} = P_T$, $\lambda_1 = 5 \times 10^{-3}\,m^{-2}$ and $\lambda_2 = 0.5\lambda_1$.
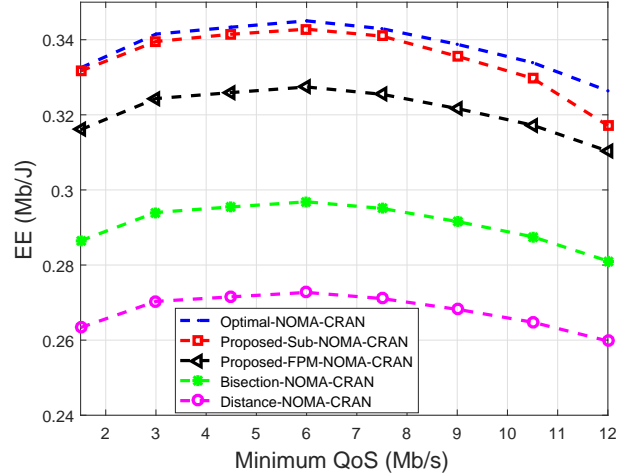

Fig. 9. Case 2: The homogeneous scenario case: The achieved EE against various minimum rate values with $\lambda_1 = 5 \times 10^{-3}$ and $\lambda_2 = 0.5\lambda$ in each tier with $P_{T_1} = P_{T_2} = 40\,\text{dBm}$.
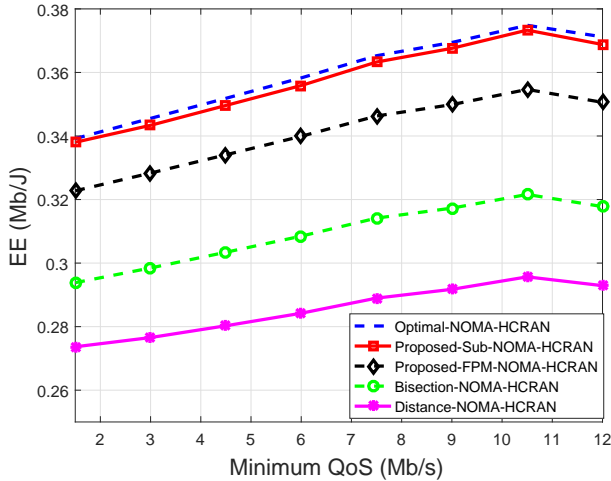

Fig. 8. Case 1: The heterogeneous scenario case: The achieved EE against various minimum rate values with $\lambda_1 = 5 \times 10^{-3}$ and $\lambda_2 = 0.5\lambda$ in each tier with $P_{T_1} = 40\,\text{dBm}$ and $P_{T_2} = 5P_{T_1}$.
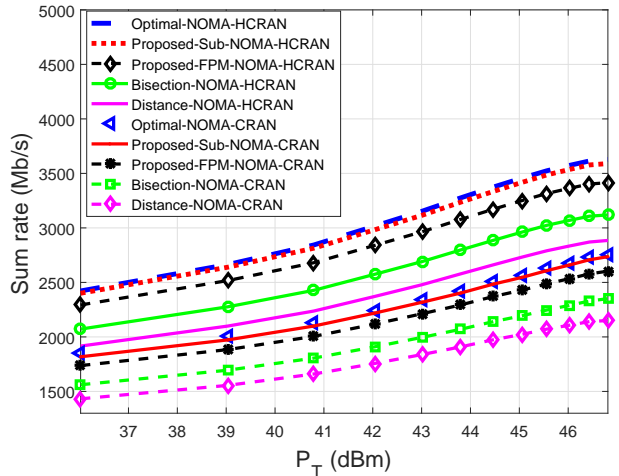

Fig. 10. The substituted tier behavior is compared in both scenarios in terms of the achievable sum rate against various levels of the transmission power

the performance of the homogeneous scenario case has lower performance levels than the heterogeneous case. To make a more specific comparison, the sum rate achieved by the macro-BS tier in the first scenario case (from the heterogeneous scenario case) is compared against that achieved by the substitute RRHs tier in the second scenario case (i.e., the one from the homogeneous scenario case). In Fig. 10, the sum rate of the heterogeneous scenario is produced depending on the same parameters adopted to produce Fig. 6. On the other hand, the sum rate of the homogeneous scenario case is obtained by using the same parameters as in Fig. 7.

Furthermore, Fig. 10 shows that the overall achievable sum rate increases in proportion to the overall transmission power, for both scenarios cases. In addition, this figure indicates clearly that the heterogeneous scenario has a small advantage over the homogeneous case. However, it should be noted that using RRHs would cause less interference to their counterparts as they use less transmission power. In addition, their nature

offers flexibility in their placements throughout the coverage area. It is obvious that both the homogeneous and the heterogeneous cases are applicable and the show slight differences between them, however, the preference between them could mainly depend on the platform, the properties of the covered area, and the type of services they are required to provide.

For further performance examination, the achievable sum rate of the substituted tier has also been compared in both scenario cases against various BS densities as shown in Fig. 11. The parameters used to produce both the heterogeneous and the homogeneous cases are the similar to those used in Figs. 4 and 5, this is mainly to ensure that this comparison reflects a clear reading for the outcomes of the simulated scenarios as well as providing a clear evaluation for the behavior of the compared techniques. Fig. 11 shows that in general, the achievable sum rate decreases in proportion the BS density increase. This is because increasing the number of the BSs, regardless of their type, would result in more competition over the available resources and also increases
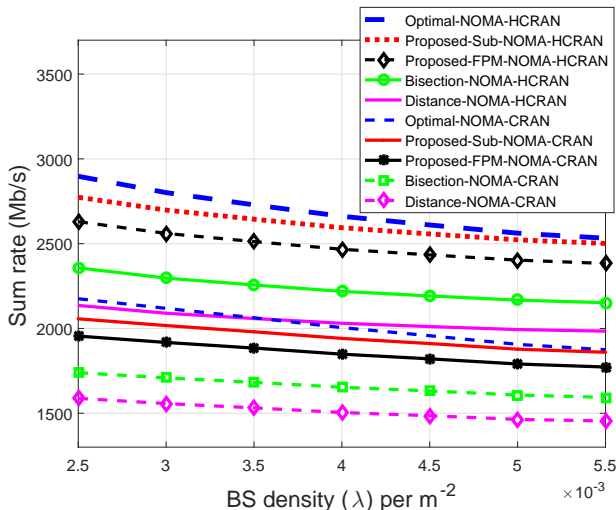
Fig. 11. The substituted tier behavior is compared in both scenario cases in terms of the achievable sum rate against different BSs densities.

the caused interference. In particular, Fig. 11 depicts that the heterogeneous case outperforms the homogeneous case as the macro-BS tier has clearly more transmission power than its substitute RRHs tier. It should be noted that the macro-BS placement is more complex and the caused interference is more detrimental as compared to their RRH peer.

It is worth mentioning that the users are randomly deployed and, to the best of the authors' knowledge, deploying the users with PPP along with Monte Carlo averaging the results, the outcome of the comparison among the considered techniques would not change much in terms of privilege.

V. CONCLUSION AND FUTURE SCOPE

This paper focused on resource allocation to optimize the achievable EE in two-tier NOMA-based CRANs. The paper presented new power allocation approaches, optimized and unoptimized. An EE optimization problem was formulated and a solution was reached using the subgradient approach with the Dinkelbach theorem assistance. The second allocation scheme was proposed based on the false position searching technique that finds the best solution to optimize the EE.

To address the effects of replacing the high-powered macro-BS tier with a low-powered RRHs, this paper presented two scenarios; the heterogeneous or NOMA-HCRAN scenario and the homogeneous or NOMA-CRAN scenario, with each scenario applies NOMA within each tier. The results obtained following both scenarios clarified that the proposed subgradient scheme offers a very close performance to the optimal scheme, this is because the subgradient approach has an iterative nature that performs deep search for the optimal value. Despite showing a slightly lower performance trends than the subgradient and the optimal techniques, the proposed FPM has less complexity and still offers a better performance relative to the bisection based and distance based NOMA-HCRAN (and so is the case with NOMA-CRAN). This is because the FPM takes into consideration both of R_{pu} and R_{pl} unlike bisection which just uses the midpoint of P_l and P_u.

In addition, FPM takes into account the interval $[P_l, P_u]$ as well as R_{pu} and R_{pl}, which makes it more accurate in finding the best roots than its bisection that solely focuses on the

$[P_l, P_u]$ interval. Finally, the results obtained from the two scenarios show that NOMA-HCRAN scenario offers better performance than NOMA-CRAN which state that exploiting macro-BS with NOMA-HCRAN offers better performance than using RRHs with NOMA-CRAN, however, on the expense of using higher transmission power and hence, causing higher interference. To sum up, applying the NOMA-CRAN scenario might be more useful than NOMA-HCRAN in a highly crowded areas such as football stadiums were the users are close to the serving BSs.

In terms of future research directions, one of the major motivations for introducing CRAN is to assign bandwidth amongst the RRHs such that, even with the non-uniform spatial distribution of the users' traffic, the QoS can be maintained. So this is a possible aspect to be addressed in the future works. It is also possible to optimize CRAN intelligence further by allowing its cooperation with innovative recent innovative technologies and platforms, for instance, Artificial Intelligence (AI), Blockchain and Internet of Things (IoT). However, this cooperation might face a number of challenges such as the scarce of cloud resources to perform an application within a certain run-time. In addition, future works could investigate extending CRAN boundaries through the decentralizing of resource management and transferring it from CRAN data centers to edge networks by exploiting the concept of Fog and/or Edge computing. One of the main features of AI is its ability to provide a backbone platform to optimize large systems that process large data amounts efficiently through adaptive-decision making mechanisms rather than human depended heuristics encoders. On the other hand, Blockchain technology has emerged recently to maintain data protection in IoT domain and real-time systems.

Due to the variety of IoT applications that are running on CRANs, it is almost mandatory to address the trade-off among different QoS requirements in the future research. This is to ensure that these applications are able to maintain optimum QoS properties with minimum overhead. Moreover, Container technologies could also be examined with CRAN in future research as it offers a lightweight cloud environment to deploy applications. This is because containers could limits data dependency among various units during the execution of user workloads as they are independent and self-contained. In addition, Container permits resource sharing among several applications whilst running simultaneously and independently.

REFERENCES

[1] L. Ferdouse, S. Erkucuk, A. Anpalagan, and I. Woungang, "Energy efficient SCMA supported downlink cloud-RANs for 5G networks," *IEEE Access*, vol. 8, pp. 1416–1430, 2020.

[2] Q. Fan, H. Lu, W. Jiang, P. Hong, J. Wu, and C. W. Chen, "Load coupling power optimization in cloud radio access networks," *arXiv preprint arXiv:1608.06749*, 2016.

[3] J. Zuo, J. Zhang, C. Yuen, W. Jiang, and W. Luo, "Energy efficient user association for cloud radio access networks," *IEEE Access*, vol. 4, pp. 2429–2438, 2016.

[4] F. N. C. INC., "The benefits of cloud-ran architecture in mobile network expansion." FUJITSU NETWORK COMMUNICATIONS INC., 2014.

[5] W. Hao, Z. Chu, F. Zhou, S. Yang, G. Sun, and K. Wong, "Green communication for NOMA-based CRAN," *IEEE Internet of Things Journal*, vol. 6, no. 1, pp. 666–678, Feb 2019.

[6] Q. Li, H. Niu, A. Papathanassiou, and G. Wu, "5G network capacity: Key elements and technologies," *Veh. Technol. Mag., IEEE*, vol. 9, no. 1, pp. 71–78, Mar. 2014.

[7] Z. Q. Al-Abbasi, D. K. C. So, and J. Tang, "Energy efficient resource allocation in downlink non-orthogonal multiple access (NOMA) system," in *2017 IEEE 85th Veh. Technol. Conference (VTC Spring)*, 2017, pp. 1–5.

[8] B. Han, L. Liu, J. Zhang, C. Tao, C. Qiu, T. Zhou, Z. Li, and Z. Piao, "Research on resource migration based on novel RRH-BBU mapping in cloud radio access network for HSR scenarios," *IEEE Access*, vol. 7, pp. 108 542–108 550, 2019.

[9] H. Q. Tran, P. Q. Truong, C. V. Phan, and Q. Vien, "On the energy efficiency of NOMA for wireless backhaul in multi-tier heterogeneous CRAN," in *2017 International Conference on Recent Advances in Signal Processing, TeleCommun. Computing (SigTelCom)*, Jan 2017, pp. 229–234.

[10] Q. Vien, T. A. Le, B. Barn, and C. V. Phan, "Optimising energy efficiency of non-orthogonal multiple access for wireless backhaul in heterogeneous cloud radio access network," *IET Commun.*, vol. 10, no. 18, pp. 2516–2524, 2016.

[11] S. S. Gill, S. Tuli, M. Xu, I. Singh, K. V. Singh, D. Lindsay, S. Tuli, D. Smirnova, M. Singh, U. Jain *et al.*, "Transformative effects of IoT, blockchain and artificial intelligence on cloud computing: Evolution, vision, trends and open challenges," *Internet of Things*, vol. 8, p. 100118, 2019.

[12] M. Vaezi and Y. Zhang, *Cloud mobile networks*. Springer, 2017.

[13] Z. Q. Al-Abbasi and D. K. C. So, "Power allocation for sum rate maximization in non-orthogonal multiple access system," in *Personal, Indoor, and Mobile Radio Communication (PIMRC), 2015 IEEE 26th Annual International Symposium on*, Aug 2015.

[14] A. A. Sabri and Z. Q. Abdulkareem AlAbbasi, "Performance analysis of framelet based OFDM system under different channel conditions," *Engineering and Technology Journal*, vol. 30, no. 5, pp. 795–810, 2012.

[15] B. Wang, K. Wang, Z. Lu, T. Xie, and J. Quan, "Comparison study of non-orthogonal multiple access schemes for 5G," in *Broadband Multimedia Systems and Broadcasting (BMSB), 2015 IEEE International Symposium on*, June 2015, pp. 1–5.

[16] Z. Q. Al-Abbasi and D. K. C. So, "User-pairing based non-orthogonal multiple access (NOMA) system," in *Proc. IEEE VTC Spring.*, May 2016.

[17] M. N. Dani, Z. Q. Al-Abbasi, and D. K. C. So, "Power allocation for layered multicast video streaming in non-orthogonal multiple access system," in *2017 IEEE Globecom Workshops (GC Wkshps)*, 2017, pp. 1–6.

[18] R. T. Rodoshi, T. Kim, and W. Choi, "Resource management in cloud radio access network: Conventional and new approaches," *Sensors*, vol. 20, no. 9, p. 2708, 2020.

[19] Y. Tao, L. Liu, S. Liu, and Z. Zhang, "A survey: Several technologies of non-orthogonal transmission for 5G," *Commun., China*, vol. 12, no. 10, pp. 1–15, Oct 2015.

[20] H. ElSawy, E. Hossain, and M. Haenggi, "Stochastic geometry for modeling, analysis, and design of multi-tier and cognitive cellular wireless networks: A survey," *Commun. Surveys Tutorials, IEEE*, vol. 15, no. 3, pp. 996–1019, Third 2013.

[21] J. Andrews, F. Baccelli, and R. Ganti, "A new tractable model for cellular coverage," in *Communication, Control, and Computing (Allerton), 2010 48th Annual Allerton Conference on*, Sept 2010, pp. 1204–1211.

[22] H. Dhillon, R. Ganti, and J. Andrews, "A tractable framework for coverage and outage in heterogeneous cellular networks," in *Information Theory and Applications Workshop (ITA), 2011*, Feb 2011, pp. 1–6.

[23] L. Han, R. Liu, Z. Wang, X. Yue, and J. S. Thompson, "Millimeter-wave MIMO-NOMA based positioning system for internet of things applications," *IEEE Internet of Things Journal*, pp. 1–1, 2020.

[24] H. Huang, Y. Yang, Z. Ding, H. Wang, H. Sari, and F. Adachi, "Deep learning-based sum data rate and energy efficiency optimization for MIMO-NOMA systems," *IEEE Trans. on Wireless Commun.*, pp. 1–1, 2020.

[25] J. Men and J. Ge, "Non-orthogonal multiple access for multiple-antenna relaying networks," *Commun. Letters, IEEE*, vol. 19, no. 10, pp. 1686–1689, Oct 2015.

[26] Q. Sun, S. Han, C. L. I, and Z. Pan, "On the ergodic capacity of MIMO NOMA systems," *IEEE Wireless Commun. Letters*, vol. 4, no. 4, pp. 405–408, Aug 2015.

[27] Z. Q. Al-Abbasi, D. K. C. So, and J. Tang, "Resource allocation for MU-MIMO non-orthogonal multiple access (NOMA) system with interference alignment," in *2017 IEEE International Conference on Commun. (ICC)*, May 2017.

[28] Q. Li, P. Ren, and D. Xu, "Security enhancement and QoS Provisioning for NOMA-Based Cooperative D2D networks," *IEEE Access*, vol. 7, pp. 129 387–129 401, 2019.

[29] J. Chen, J. Jia, Y. Liu, X. Wang, and A. H. Aghvami, "Optimal resource block assignment and power allocation for D2D-Enabled NOMA communication," *IEEE Access*, vol. 7, pp. 90 023–90 035, 2019.

[30] Y. Dai, M. Sheng, J. Liu, N. Cheng, X. Shen, and Q. Yang, "Joint mode selection and resource allocation for D2D-enabled NOMA cellular networks," *IEEE Trans. on Veh. Technol.*, vol. 68, no. 7, pp. 6721–6733, 2019.

[31] Young Bae Song, Hyeon Su Kang, and Duk Kyung Kim, "5G cellular systems with D2D assisted NOMA relay," in *2016 URSI Asia-Pacific Radio Science Conference (URSI AP-RASC)*, 2016, pp. 1–3.

[32] C. Yan, A. Harada, A. Benjebbour, Y. Lan, A. Li, and H. Jiang, "Receiver design for downlink non-orthogonal multiple access (NOMA)," in *Veh. Technol. Conference (VTC Spring), 2015 IEEE 81st*, May 2015, pp. 1–6.

[33] M. Kaneko, K. Hayashi, and H. Sakai, "Superposition coding based user combining schemes for non-orthogonal scheduling in a wireless relay system," *IEEE Trans. on Wireless Commun.*, vol. 13, no. 6, pp. 3232–3243, June 2014.

[34] K. Higuchi and Y. Kishiyama, "Non-orthogonal access with random beamforming and intra-beam SIC for cellular MIMO downlink," in *Veh. Technol. Conference (VTC Fall), 2013 IEEE 78th*. IEEE, 2013, pp. 1–5.

[35] M. Al-Imari, P. Xiao, M. A. Imran, and R. Tafazolli, "Uplink non-orthogonal multiple access for 5G wireless networks," in *2014 11th International Symposium on Wireless Commun. Systems (ISWCS)*, Aug 2014, pp. 781–785.

[36] X. Chen, A. Benjebbour, A. Li, and A. Harada, "Multi-user proportional fair scheduling for uplink non-orthogonal multiple access (NOMA)," in *2014 IEEE 79th Veh. Technol. Conference (VTC Spring)*, May 2014, pp. 1–5.

[37] S. Chen, K. Peng, and H. Jin, "A suboptimal scheme for uplink NOMA in 5G systems," in *2015 International Wireless Commun. and Mobile Computing Conference (IWCMC)*, Aug 2015, pp. 1429–1434.

[38] Y. Endo, Y. Kishiyama, and K. Higuchi, "Uplink non-orthogonal access with MMSE-SIC in the presence of inter-cell interference," in *2012 International Symposium on Wireless Communication Systems (ISWCS)*, Aug 2012, pp. 261–265.

[39] F. Zhou, Y. Wu, R. Q. Hu, Y. Wang, and K. K. Wong, "Energy-efficient NOMA enabled heterogeneous cloud radio access networks," *IEEE Network*, vol. 32, no. 2, pp. 152–160, Mar. 2018.

[40] Q. Vien, T. A. Le, C. V. Phan, and M. Opoku Agyeman, "An energy-efficient NOMA for small cells in heterogeneous CRAN under QoS constraints," in *European Wireless 2017; 23th European Wireless Conference*, May 2017, pp. 1–6.

[41] X. Gu, X. Ji, Z. Ding, W. Wu, and M. Peng, "Outage probability analysis of non-orthogonal multiple access in cloud radio access networks," *IEEE Commun. Letters*, vol. 22, no. 1, pp. 149–152, Jan 2018.

[42] R. Rai, H. Zhu, and J. Wang, "Coordinated scheduling and power control for non-orthogonal multiple access (NOMA) enabled H-CRAN," in *2020 IEEE 91st Vehicular Technology Conference (VTC2020-Spring)*, 2020, pp. 1–5.

[43] A. Benjebbour, Y. Saito, Y. Kishiyama, A. Li, A. Harada, and T. Naka-mura, "Concept and practical considerations of non-orthogonal multiple access (noma) for future radio access," in *Intelligent Signal Processing and Commun. Systems (ISPACS), 2013 International Symposium on*, Nov 2013, pp. 770–774.

[44] M. Kaneko, I. Randrianantenaina, H. Dahrouj, H. Elsawy, and M. S. Alouini, "On the opportunities and challenges of NOMA-based fog radio access networks: An overview," *IEEE Access*, vol. 8, pp. 205 467–205 476, 2020.

[45] F. J. Martin-Vega, Y. Liu, G. Gomez, M. C. Aguayo-Torres, and M. Elkashlan, "Modeling and analysis of NOMA Enabled CRAN with cluster point process," in *GLOBECOM 2017 - 2017 IEEE Global Communications Conference*, 2017, pp. 1–6.

[46] T. Nguyen, V. Phan, and P. Q. Truong, "Energy efficient performance analysis of NOMA for wireless down-link in heterogeneous networks under imperfect SIC," in *2019 International Conference on System Science and Engineering (ICSSE)*, July 2019, pp. 255–260.

[47] S. Tomida and K. Higuchi, "Non-orthogonal access with SIC in cellular downlink for user fairness enhancement," in *Intelligent Signal Processing and Commun. Systems (ISPACS), 2011 International Symposium on*, Dec 2011, pp. 1–6.

[48] V. Basnayake, D. N. K. Jayakody, V. Sharma, N. Sharma, P. Muthuchidambaranathan, and H. Mabed, "A new green prospective of non-orthogonal multiple access (noma) for 5g," *Information*, vol. 11, no. 2, p. 89, 2020.

[49] Z. Q. Al-Abbasi and D. K. C. So, "Resource allocation in non-orthogonal and hybrid multiple access system with proportional rate constraint," *IEEE Trans. on Wireless Commun.*, vol. PP, no. 99, pp. 1–1, 2017.

[50] G. Auer, V. Giannini, C. Desset, I. Godor, P. Skillermark, M. Olsson, M. Imran, D. Sabella, M. Gonzalez, O. Blume, and A. Fehske, "How much energy is needed to run a wireless network?" *Wireless Commun., IEEE*, vol. 18, no. 5, pp. 40–49, October 2011.

[51] W. Dinkelbach, "On nonlinear fractional programming," *Management Science*, vol. 13, no. 7, pp. 492–498, 1967.

[52] K. T. K. Cheung, S. Yang, and L. Hanzo, "Achieving maximum energy-efficiency in multi-relay OFDMA cellular networks: A fractional programming approach," *IEEE Trans. on Commun.*, vol. 61, no. 7, pp. 2746–2757, July 2013.

[53] S. Boyd and L. Vandenberghe, *Convex optimization*. Cambridge university press, 2004.