Resource Allocation and Cluster Formation for Imperfect NOMA in DL/UL Decoupled HetNets


Abstract—Being capable of serving multiple users with the same radio resource, non-orthogonal multiple access (NOMA) can provide desirable performance enhancements in a fair and spectral efficient manner. In this paper, we investigate the resource allocation (RA) and cluster formation (CF) aspects of NOMA for downlink (DL) uplink (UL) decoupled (DUDe) heterogeneous networks (HetNets). A non-ideal NOMA scheme is considered with power disparity and sensitivity constraints (PDSCs), delay tolerance, and residual interference after cancellation. Taking the PDSCs into account, we analytically show that using the DL decoding order limits UL-NOMA performance by that of OMA, while employing an inverse order result in a performance gain that is mainly determined by the channel gain disparity of users. Therefore, a generic CF method is proposed for any type of user graph, which iteratively forms clusters using Blossom algorithm. Finally, highly non-convex RA problem is converted into a convex form by employing geometric programming (GP) where power and bandwidth are optimized to maximize network sumrate and max-min fairness objectives.

I. INTRODUCTION

EVER increasing number of communications devices with ambitious quality of service (QoS) demands puts forward challenging goals for fifth-generation (5G) networks such as massive connectivity, high spectral and power efficiency, low latency, etc. To fulfill these demands, ultra-dense HetNets have already been considered as a promising solution since densification of the network have the ability to boost network coverage and capacity, while reducing operational and capital expenditures of mobile operators [1]. However, traditional orthogonal multiple access (OMA) schemes dedicate radio resources to a certain user either in time or frequency domains, which is not adequately spectral efficient for expected massive number of users. Furthermore, traditional HetNets associate both DL and UL user equipments (UEs) with base stations (BSs) as per a biased received signal strength information (RSSI) such that traffic is offloaded from highly congested macro BSs (MBs) to small BSs (SBs). However, this may not yield a desirable performance for UL as it has different traffic conditions and UEs have close maximum permissible transmission powers. Alternatively, decoupling the DL and UL user associations are recently considered as a promising solution for an improved traffic offloading and performance [2]. [3].

Having its root in multi-user detection, NOMA has recently gained attentions with its ability to serve multiple users by multiplexing them into the same radio resource which may be realized either in power or code domain [4]. In particular, power domain NOMA ensures a certain reception power for each user such that while some users are enforced to operate in low power levels while being granted to cancel interference of others using successive interference cancellation (SIC), some others granted with high power levels but with (limited) interference cancellation (IC) allowance. Since the channel gain disparity is a limiting factor of the achievable NOMA gain, CF and RA problems are of the utmost importance to maximize achievable NOMA gain, which is addressed throughout this paper for DUDe HetNets.

Related works on NOMA can be exemplified as follows: System-level performance evaluation has been conducted in [5]. [6] for two-user downlink NOMA with superposition transmission at BS and SIC at the user terminals. Authors of [7] proposed a user pairing and power allocation for DL-NOMA based on proportional fairness. User clustering and power-bandwidth allocation of DL-HetNets is studied in [8] where clusters are formed using sequential multi-partite matching to achieve maximum sumrate, max-min fairness, and energy-spectrum cost minimization objectives. The impact of channel gain disparity on DL-NOMA is investigated in [9] for a two-user system with fixed and cognitive radio inspired power allocation. A near optimal solution was proposed by combining Lagrangian duality and dynamic programming for joint power and channel allocation in [10]. The work in [11] proposed a DL power allocation scheme by first grouping users into a single cluster, and then optimize the power allocation. For UL-NOMA, on the other hand, the work in [12] proposed a greedy-based algorithm to improve the throughput. In [13], the authors derived closed-form expressions for the outage probability of two-user UL-NOMA assuming fixed powers of different users. Aside from above research efforts, our contributions on NOMA operated UL-HetNets can be enlisted as follows:

1) We consider a non-ideal UL-NOMA scheme due to the power disparity and sensitivity constraints (PDSCs), delay tolerance, and residual interference. Although [12] states that decoding order does not affect the sumrate performance, we show that this may not be applicable when PDSCs are taken into account. Analytic findings prove that employing DL decoding order in the UL-NOMA limit NOMA sumrate performance by that of OMA, while employing an inverse DL decoding order in the UL-NOMA provide a NOMA gain which is determined by the channel gain disparity of users.

2) A generic fast yet high performance CF algorithm is proposed to be used for any type of user graphs. As channel gain disparity is a limiting factor on NOMA gain, edge weights
are first determined based on channel gain ratio and QoS requirement of users. Then, clusters are formed iteratively employing Blossom algorithm.

3) For given clusters, highly non-convex power and bandwidth allocation problem is put into convex form by modeling it as an geometric programming (GP) problem. For sumrate maximization and max-min fairness objectives, impacts of decoding, cluster size, and user density are investigated to reveal potential of NOMA in UL-HetNets.

The remainder of paper is organized as follows: Section II introduces the system model for DUDe HetNets. Constraints and imperfection of SUC receiver is presented in Section III. Then, Section IV provide proposed CF and RA methods. Numerical results are illustrated in Section V and Section VI concludes the paper with a few remarks.

II. SYSTEM MODEL

We consider an UL-HetNet where tier-1 consists of a single macrocell and tier-2 consists of smallcells. The spatial distribution of SBSs in tier-2 follows a Poisson point process (PPP) $\Phi_s$ of density $\lambda_s$. We denote the index set of all BSs by $C = \{c \mid 0 \leq c \leq S \}$ where $S$ represent the number of SBSs, $c = 0$ is the MBS index, and $0 < c \leq S$ are indices for SBSs, respectively. The spatial distribution of UEs follows a PPP $\Phi_u$ of density $\lambda_u$. Maximum transmission powers of UEs and BSs are generically denoted as $P_u$ and $P_s$, respectively, where $P_u$ equals to $P_m$ and $P_s$ for MBS and SBSs, respectively.

Fig. 1: Demonstration of DL/UL decoupling and user clustering.

DUCo scheme associates UEs with the same BS for both DL and UL transmission based on RSSI, which yields a significant traffic load on macrocells due to MBSs’ high transmission power. In order to offload DL traffic from MBSs to SBSs, UE are associated by introducing a bias factor, $0 \leq b \leq 1$. Nonetheless, requiring UEs to follow the same association in both UL and DL may not always yield an optimal performance. Keeping DL association method the same as in DUCo, on the other hand, DUDe scheme alternatively determines the UL association based on channel gain such that a UE can be associated with a nearby SBS in the UL even if it is associated with the MBS in the DL. Denoting the composite channel gain from UE to BS as $h^c_u$, DUDe association of UE is determined by

$$\Psi_u = \arg \min_{v \in C} \{h^c_u\} \tag{1}$$

where $h^c_u = PL^c_u 10^{C10/10} + \mathbb{E}[\{q^c_u\}]$, $PL^c_u = K^c_u d^{-\epsilon} u$, $K^c_u$ is a constant related to antenna parameters, $d_u$ is the distance between UE and BS, $\eta_u$ is the path loss exponent, $10^{C10/10} \sim \ln(0, \sigma_u^2)$ represents the log-normally distributed shadowing, $\xi^c_u$ is a normal random variable representing the variation in received power with a standard deviation of $\sigma_u^2$, i.e., $\xi^c_u \sim N(0, \sigma^2_u)$. Please note that we especially average the small scale fading in order to avoid the ping-pong effects in user associations. For a bias factor of $b = 0.7$, a simple DUDe scenario is depicted in Fig. 1 where decoupled UEs, shown as circled hexstars, are associated with the MBS and nearby SBS for DL and UL, respectively. In DUCo scheme, on the other hand, they are associated with MBS for both DL and UL.

Fig. 1: Demonstration of DL/UL decoupling and user clustering.

For given $\Psi_u, \forall u$, $U_c$ denotes the index set of $U_c$ UEs associated with the BS$_r$. Furthermore, index set of all $U = \sum u U_c$ UEs is given as $U = \bigcup u U_c$. $U_c$ is partitioned into clusters such that set of users within cluster $r$ is symbolized as $K_c \equiv \{\delta_{c,r}^u \mid 0 \leq \beta_r \leq 1\}$, $\delta_{c,r}^u$ is a binary indicator for the association of UE with BS$_c$. Denoting the set of all clusters as $R$, cluster $r$ of cell $c$ is allowed to utilize $0 \leq \beta_r \leq 1$, $\sum_{c \in R} \beta_r \leq 1, 1 \leq r \leq R$, portion of the entire UL bandwidth, $B$. Therefore, received signal by BS$_c$ from UE$_i \in K^c_c$ can be expressed as

$$y^c_{c,r} = \sqrt{P_u \varpi^c_{u,r}, h^c_u \delta^c_r x_u + \sum_{\nu \neq u} \sqrt{P_u \varpi^c_{u,r}, h^c_u \delta^c_r x_u} + n^c_{c,r} \tag{2}$$

where $\varpi^c_{u,r}$ is the power weight of UE$_u$ on cluster $r$, $x_u$ is the transmitted message, $n^c_{c,r} \sim N(0, \sigma_{c,r}^2)$ is the additive white Gaussian noise with variance $\sigma_{c,r}^2 = N_0 \beta_r B$, and $N_0$ is the thermal noise power spectral density.

III. CONSTRAINTS AND IMPERFECTIONS OF SIC

A. Impacts of Decoding Order in UL-NOMA

In order to extract desired signal, the SIC receiver first decodes the strongest interference, then re-generates the transmitted signal, and finally subtracts it from the received composite signal, which is repeated for succeeding interference components. Even though DL-NOMA decodes user signals in descending order of their channel gains, employing the same direction in UL-NOMA may not give the desired performance since UEs have limited transmission power and SIC requires a certain reception power disparity, $p_{th}$, between different messages [11]. To be more specific, we focus on a simple cluster with unit bandwidth allocation and users $UE_k$ and $UE_\ell$ such that $h_k \geq h_\ell$. Employing the descending decoding order as in DL-NOMA (i.e., $UE_k$ cancel the interference of $UE_\ell$),
OMA and NOMA sumrate capacities can be respectively given as

\[
\mathcal{C}_i^O = \frac{1}{2} \left\{ \log_2 \left( 1 + \rho \bar{h}_k \right) + \log_2 \left( 1 + \rho \bar{h}_\ell \right) \right\}
\]

\[
\mathcal{C}_i^N = \log_2 \left( 1 + \rho \bar{w}_k \bar{h}_k \right) + \log_2 \left( 1 + \frac{\omega \bar{h}_\ell}{\bar{h}_k + 1/\rho} \right)
\]

where \( \rho = P/\sum_{n} P_n \) and \( 0 \leq \bar{w}_k \leq 1, 0 \leq \bar{w}_\ell \leq 1 \) are power weights. As \( \rho \to \infty \), asymptotic gain can be expressed as

\[
\tilde{\mathcal{C}}_i^O = \lim_{\rho \to \infty} \mathcal{C}_i^O = \frac{1}{2} \log_2 \left( \rho^2 \bar{h}_k \bar{h}_\ell \right)
\]

\[
\tilde{\mathcal{C}}_i^N = \lim_{\rho \to \infty} \mathcal{C}_i^N = \log_2 \left( \rho \bar{w}_k \bar{h}_k \right)
\]

Accordingly, UL-NOMA gain with descending decoding order, \( \Delta_1 = \mathcal{C}_i^N - \mathcal{C}_i^O \), can be put in the asymptotic form as follow

\[
\tilde{\Delta}_1 = \lim_{\rho \to \infty} \Delta_1 = \log_2 \left( \rho \bar{w}_k \bar{h}_k \right) - \frac{1}{2} \log_2 \left( \rho^2 \bar{h}_k \bar{h}_\ell \right)
\]

\[
= \log_2 \left( \frac{\rho \bar{w}_k \bar{h}_k}{\sqrt{\bar{h}_k \bar{h}_\ell}} \right) = \log_2 \left( \bar{w}_\ell \sqrt{\frac{h_k}{h_\ell}} \right) \leq 0
\]

which follows from \( \bar{w}_\ell \leq 1 \) and \( h_k \geq h_\ell \). That is, even if the receiver has perfect sensitivity (\( p_\ell \to 0 \)), NOMA performance is upper bounded by that of OMA. On the contrary, asymptotic NOMA gain for an ascending order case (i.e., UE\(_\ell\) cancel the interference of UE\(_k\)) can be obtained as

\[
\tilde{\Delta}_1 = \log_2 \left( \frac{\rho \bar{w}_k \bar{h}_k}{\rho \sqrt{h_k h_\ell}} \right) = \log_2 \left( \bar{w}_\ell \sqrt{\frac{h_k}{h_\ell}} \right)
\]

which follows similar steps in \([3]\) - \([7]\). In the ascending order case, the power disparity constraint requires \( P_u \bar{w}_k \bar{h}_k - P_u \bar{w}_\ell h_k \geq p_\ell \) constitutes the upper bound on \( \bar{w}_\ell \) as \( \bar{w}_\ell \bar{h}_k - p_\ell \bar{w}_k \bar{h}_k \geq \bar{w}_\ell \), thus the upper bound on \( \tilde{\Delta}_1 \) as

\[
\tilde{\Delta}_1 \leq \log_2 \left( \frac{\bar{w}_\ell h_k - p_\ell \bar{w}_k}{\sqrt{h_k h_\ell}} \right) \leq \log_2 \left( \bar{w}_\ell \sqrt{\frac{h_k}{h_\ell}} \right) \leq \frac{1}{2} \log_2 \left( h_k - \log_2 \left( h_k \right) \right)
\]

where upper bound of NOMA gain is determined by power weight of high channel gain user and channel gain disparity of users. Especially, in the best case scenario where the receiver has perfect sensitivity (\( p_\ell \to 0 \)) and UE\(_k\) transmit at the maximum power (\( \bar{w}_k \to 1 \)), NOMA performance is mainly determined by the channel gain disparity as shown in \([11]\).

### B. Constraints and Imperfections of SIC Receivers

Let us consider a generic UE cluster of BS\(_c\) \( K^c \) = \( \{ j \mid h^j_k > h^j_{\ell} \geq h^j_{\ell+1}, U_c \geq j \geq 1 \} \) which is allocated to subcarrier \( r \) with bandwidth \( \beta^r \). Fig. 4 illustrates exemplary clusters for different BSs in cyan colored star shapes. Therefore, UL decoding order set for UE\(_j\) is given by

\[
\{ p_{c,r}^j < \ldots < p_{c,r}^{j+1} < p_{c,r}^{j+1} < \ldots < p_{c,r}^1 \}
\]

where \( K^c_j \) is the cluster size, \( p_{c,r}^j \) is the power received from UE\(_j\), \( \mathcal{O}_{c,j}^j = \{ j + 1, \ldots, K^c_j \} \) is the lower rank decoding order set, and \( \mathcal{O}_{c,j}^{j-\ell} = \{ 1, \ldots, j-1 \} \) is the higher rank decoding order set for UE\(_j\). Furthermore, power disparity and sensitivity constraints (PDSCs) can be given in linear scale as

\[
p_{c,r}^j - \sum_{l=k+1}^j p_{c,r}^l \geq p_{\delta}, \forall UE_j, \forall UE_k \in \mathcal{O}_{c,j}^{j-\ell}
\]

where the intuition is that during the IC process of UE\(_k\) in \( \mathcal{O}_{c,j}^h \), receiver observes undecoded signals of UE\(_j\) in \( \mathcal{O}_{c,j}^h \), \( l < k \), as noise. On the other hand, the latency and complexity order of SIC receiver is mainly determined by the total number of IC stages and users \([14]\). Thus, BS\(_c\) can tolerate delay of at most \( M_c \) interference cancellations for all clusters. That is, BS\(_c\) is able to cancel at most \( M_c \) UEs among UE\(_k\) in \( \mathcal{O}_{c,j}^h \). Thus, set of cancelled UEs can be given as the first \( \min \{ M_c, \mathcal{O}_{c,j}^h \} \) elements of \( \mathcal{O}_{c,j}^j \) (i.e., \( \min \{ M_c, \mathcal{O}_{c,j}^{j-\ell} \} \) most dominant interference), which is denoted as \( \mathcal{O}_{c,j}^{j-\ell} \).

Aside from the channel gain disparity, another performance limiter for NOMA gain is generated by errors during the IC process which yields a residual interference at the very end. The first source of cancellation error is caused by propagation of bit decision errors of previous symbols during the decoding process. The second error source is imperfect CSI which leaves a residual interference right after IC. Accordingly, a generic signal-to-interference-plus-noise-ratio (SINR) representation of the imperfect SIC receiver can be given by

\[
\gamma_{c,r}^j = \sum_{i\in\mathcal{K}_c^c} \rho_{c,r}^i \sum_{k\in\mathcal{K}_c^c} \rho_{c,r}^k + \sum_{i\in\mathcal{K}_c^c} \rho_{c,r}^i \sum_{k\in\mathcal{K}_c^c} \rho_{c,r}^k + \sigma_{c,r}^2
\]

where \( 0 \leq \epsilon_c \leq 1 \) is the residual interference factor of BS\(_c\). The first, second, and third terms in denominator represent the residual interference after cancellation, uncancelled interference due to delay constraint, and uncancelled lower rank interference, respectively. For a desirable performance, therefore, a cluster member should be able to cancel the dominant interference while tolerating the SIC imperfection and interference induced from lower rank UEs. Hence, the capacity of UE\(_j\) is given by

\[
C_{c,j} = \sum_{c,r} \delta_{c,r}^j \beta^r \log_2 (1 + \gamma_{c,r}^j)
\]

### IV. CLUSTER FORMATION AND RESOURCE ALLOCATION

#### A. Oimal Problem Formulation

Optimization of the integer-valued cluster number/sizes and user-cluster associations induces infeasible time complexity even for moderate sizes of HetNets. Thus, this combinatorial
problem can be put into a tractable form by making following assumptions: 1) Each cluster have a dedicated band and each band has a dedicated SIC receiver at the BSs so that inter-cluster interference is avoided within a macrocell coverage area, and 2) The cluster size is determined by orthogonalizing the lowest channel gain user, i.e., lowest channel gain cluster member can cancel interference of all other cluster members. Joint CF and RA problem can then be formulated as in $P_0$ where vectorized form of power allocation variables is denoted as $\varpi = [\varpi_1^u, \ldots, \varpi_r^u, \ldots, \varpi_{S+1}^u]$, $\varpi_c^u = [\varpi_{c,1}^u, \ldots, \varpi_{c,r}^u, \ldots, \varpi_{c,R_c}^u]$, and $\varpi_{c,r}^u = [\varpi_{c,t}^u, \ldots, \varpi_{c,r}, \ldots, \varpi_{c,R_c}]$. $\delta$ follows the same notation of $\varpi$. Similarly, bandwidth allocation variables can be vectorized as $\beta = [\beta_1^c, \ldots, \beta_r^c, \ldots, \beta_{S+1}^c]$ where $\beta_c = [\beta_1^c, \ldots, \beta_r^c, \ldots, \beta_{R_c}^c]$.

$$\begin{align*}
P_o & : \max_{\delta, \beta, \varpi} \quad \varpi \delta \beta \varpi \quad \varpi (\delta, \beta, \varpi) \\
C_1: & \quad \sum_{c,u} \delta_{c,r}^u \leq M_c, \quad \forall c, u \\
C_2: & \quad \sum_{c,u} \delta_{c,r}^u = 1, \quad \forall c, u \\
C_3: & \quad \beta_{c,r}^e \leq 1, \quad \forall c, r \\
C_4: & \quad p_c \leq p_{c,r}^k - \sum_{l=k+1}^{u} p_{c,r}^l, \quad \forall c, r, k \in \mathcal{O}_{u}, \forall u \\
C_5: & \quad q_u \leq \mathcal{C}_u, \quad \forall u \\
C_6: & \quad \delta_{c,r}^u \in \{0, 1\}, \quad 0 \leq \beta_{c,r}^e \leq 1, \quad 0 \leq \varpi_{c,u} \leq \delta_{c,r}^u.
\end{align*}$$

In $P_o$, $C_1$ ensures that a UE is assigned to at most one cluster and $C_2$ limits the number of UEs within a cluster by $M_c$, which is a design parameter. $C_3$ constrains the total bandwidth allocation weights to unity. PDSCs and QoS requirements are introduced in $C_4$ and $C_5$, respectively. $C_6$ defines optimization variables’ domains where the power allocation for $\mathcal{U}_u$ on cluster $r$ is set to zero ($0 \leq \varpi_{c,r}^u \leq \delta_{c,r}^u$) if $\mathcal{U}_u \notin \mathcal{C}_r$. In what follows, we deal with this mixed integer non-linear programming (MINLP) problem by decoupling it into CF and RA subproblems.

In what follows, we deal with this mixed integer non-linear programming (MINLP) problem by decoupling it into CF and RA subproblems.

**B. Cluster Formation Using Blossom Algorithm**

As already analyzed in Section III-A, channel gain disparities between cluster members determine the achievable performance gain by NOMA. Therefore, we develop a fast yet high performance method for user clustering. Let us consider a generic user graph of BS,$c$, $G_c(\mathcal{U}_c, \mathcal{E}_c, \mathcal{O}_{c})$ where $\mathcal{U}_c$, $\mathcal{E}_c$, and $\mathcal{O}_{c}$ denote the set of vertices (i.e., users), edges, and edge weights, respectively. An exemplary graph for a BS with 12 users is shown in Fig. 2.

For a cluster size of 2, an optimal cluster formation can be obtained using a weighted bi-partite matching method, e.g., Hungarian algorithm with complexity $O(U_c^2)$ [15]. However, a generic cluster size necessitates a universal method which can work on any type of graph. As an alternative, Blossom algorithm takes a general graph $G$ and constructs a maximum matching $\mathcal{M}$ by iteratively improving initial empty matching by adding edges to build augmenting paths in $\mathcal{M}$, which has a complexity of $O(|\mathcal{E}_c|U_c)$ [16]. Unlike Hungarian algorithm which cannot handle odd-length cycles, Blossom algorithm contract odd-length cycles (blossoms) to a single vertex, from where the search continues iteratively in the contracted graph. Edge weights between UE,$i \in \mathcal{U}_c$, and UE,$j \in \mathcal{U}_c$, is given as

$$\Omega_{i,j}^c = \max(h_i^c, h_j^c)/|q_i - q_j|$$

which favors for users with high channel gain disparity for higher NOMA gains as explained in Section III-A. In the denominator, on the other hand, users with close QoS requirement is favored in order not to allocate unnecessarily high bandwidths.

**Algorithm 1 Cluster Formation for BS,$c$, $\forall c$.**

**Input:** $G_c(\mathcal{U}_c, \mathcal{E}_c, \mathcal{O}_{c})$

**Output:** Clusters of size $L_c + 1$

1: $\mathcal{M}_0 \leftarrow \emptyset$
2: $\mathcal{K}_c^r \leftarrow \emptyset$, $\forall r \in \mathcal{R}_c$
3: $G_0 \leftarrow G_c(\mathcal{U}_c, \mathcal{E}_c, \mathcal{O}_{c})$
4: for $i = 1 : L_c$ do
5: \quad $(\mathcal{M}_i, F_i) \leftarrow \text{Blossom}(G_{i-1})$
6: \quad $\tilde{F}_i \leftarrow \text{sort}(F_i)$ & select max, $R_c$ weighted matches
7: \quad $\mathcal{K}_c^r \leftarrow \mathcal{K}_c^r \cup \tilde{F}_i[r]$; update clusters,
8: \quad $A_i \leftarrow \text{Elect the highest channel gain vertices from } \tilde{F}_i$
9: \quad $G_i \leftarrow G_{i-1} \setminus A_i$; eliminate $V_i$ for the next iteration.
10: end for
11: return $\mathcal{K}_c^r$, $1 \leq r \leq R_c$

Accordingly, Algorithm 1 presents the proposed CF method where matching sets, clusters, and graph are initialized between lines 1-3. Thereafter, clusters are formed iteratively within the for loop between lines 4-10 where number of clusters within a cell is determined as $R_c = \lceil \frac{|\mathcal{U}_c|}{L_c} \rceil$. In line 5, we first run the Blossom Algorithm to obtain the maximum weight perfect matching set $\mathcal{M}_i$ and edge weights of $\mathcal{M}_i$, $F_i$. From $F_i$, we select the best $R_c$ matching with the highest weights $\tilde{F}_i$ where $r^{'th}$ matching contains two members of $r^{'th}$ cluster. Therefore, cluster sets are updated in line 6. Before proceeding the next iteration, higher channel
gain vertices in \( \tilde{\mathcal{F}}_1 \) is removed from the graph in lines 8-9 as clusters admit new members with respect to lower channel gain member. For \( M_c = 2 \), a simple example of Algorithm 1 is demonstrated in Fig. 2 where first iteration result in \( \tilde{\mathcal{F}}_1 = \{\{1,4\},\{3,2\},\{5,12\},\{6,9\}\} \). Then, we remove the higher channel gain users, \( \mathcal{A}_1 = \{1,3,5,12\} \), from the graph before the second iteration where we get \( \tilde{\mathcal{F}}_2 = \{\{4,8\},\{2,7\},\{12,10\},\{9,11\}\} \), which finalizes the cluster formations. Due to the reduction in graph size at each iteration, overall complexity of Algorithm 1 is \( O(\sum_{i=1}^n |\mathcal{L}_i||\mathcal{V}_i|^{1/2}) \) where \( |\mathcal{V}_i| = U_c - (i-1)R_c \) and \( |\mathcal{L}_i| = |\mathcal{V}_i|(|\mathcal{V}_i|-1)/2 \). In the average case complexity, we have \( \overline{|\mathcal{V}|} = U_c - R_c (L_c - 1)/2 \) and \( O(\overline{|\mathcal{E}|}^2 \overline{|\mathcal{V}|}^{1/2}) \).

C. Resource Allocation Using Geometric Programming

GP can handle RA problems by exploiting useful theoretical and computational properties of monomials and posynomials which are defined as follows \[17\]

**Definition 1.** A function \( f : \mathbb{R}_+^n \rightarrow \mathbb{R} \) is defined as a monomial such that

\[
 f(x|\kappa, e) = \kappa x_1^{e(1)} x_2^{e(2)} \ldots x_n^{e(n)} = \kappa \prod_j x_j^{e(j)}
\]

(17)

where the multiplicative constant \( \kappa \geq 0 \) and the exponential constants \( e(j) \in \mathbb{R}, j = 1,2,\ldots,n \). On the other hand, positive sum of monomials is defined as a posynomial

\[
 g(f) = \sum_i f_i(x_i|\kappa_i, e_i) = \sum_i \kappa_i \prod_j x_j^{e(j)}
\]

(18)

Standard form of GP problems (GPPs) require a posynomial objective and accept monomial and posynomial equality and inequality constraints, respectively. Even though GPPs in standard form are non-convex due to the non-convexity of posynomials, it can easily be converted into an equivalent convex problem based on following features of posynomials: 1) Posynomials are closed under addition, multiplication, and positive scaling, 2) If \( f \) is a monomial and \( g \) is a posynomial, then \( g/f \) is also a posynomial, and 3) Posynomials can be converted into a convex form with logarithmic change of all the variables and multiplicative constants \[18\].

Based on GP preliminaries and a given CF, we now focus on optimal power-bandwidth allocation problem

\[
P_1(\varpi, \beta|\delta) : \min_{\varpi, \beta} \varpi \text{ s.t. } C_3-C_7
\]

where we consider following objectives: 1) Network Sumrate Maximization: \( \varpi_2(\varpi, \beta) = \sum_{c,r,u} \varpi_{c,r,u} \)\textsuperscript{u}, and 2) Min-Max Fairness: \( \varpi_3(\varpi, \beta) = \min(C_u) \). In the remainder, we explain how \( P_1^o \)s can be put into standard GP form starting from their common set of constraints \( C_3 - C_7 \).

Total power and bandwidth allocation constraints in \( C_3 \) and \( C_4 \) are simply sum of variables and can easily be recognized as posynomials. Received signal power, \( p_{u,r}^u = \bar{P}_u \delta_{u,r} \varpi_{c,r,u} \nu_{c,r,u} \) is a monomial as it is a single variable multiplied with a non-negative constant. Please note that \( \varpi_{c,r,u} \leq \delta_{u,r} \) in \( C_7 \) is embedded in the multiplicative constant of the \( p_{u,r}^u \). Hence, the nominator of the SINR expression in (14) is a monomial while the interference power in the denominator is a posynomial. Noting that \( \sigma_{c,r,u}^2 = N_0 B \beta \) is a monomial, inverse SINR (ISINR) \( \rho_{u,r}^u \pm 1/\gamma_{u,r}^u \) is a posynomial since it is ratio of a posynomial to a monomial. PDSCs in \( C_9 \), on the other hand, can be put into a posynomial form as \( p_{u,r}^u + \sum_{k \neq u} \rho_{k,r}^u \rho_{u,k}^u \leq 1 \) which is nothing but a posynomial to monomial ratio. Assuming that cluster are formed to ensure high SINR regime, we have \( \log_2(1 + \gamma_{u,r}^u) \approx \log_2(\tilde{\gamma}_{u,r}^u) \). Accordingly, QoS constraints in \( C_6 \) can be written in the form of a generalized posynomial as \( k_{\beta,c} \rho_{u,r}^u \leq 1 \) where \( k_u = 2^{q_u/\beta} \). For \( P_1^o \), \( \varpi_i(\varpi, \beta) \) can be transformed as

\[
 \max_{\varpi, \beta} \sum_{c,r,u} \beta_{c,r} \log_2 \left( 1 + \gamma_{u,r}^u \right) \approx \max_{\varpi, \beta} \sum_{c,r,u} \log_2 \left( \tilde{\gamma}_{u,r}^u \right)^{\beta_{c,r}}
\]

(19)

where the approximation (**) is based on high SINR assumption. Due to the monotonicity of the logarithm, maximization of \( \log(f(\cdot)) \) is equivalent to maximization of \( f(\cdot) \) which yields (**) by changing the objective from maximization of SINRs to minimization of ISINRs. As mentioned before, non-integer powered posynomials can be handled by introduction of auxiliary variables and constraints, i.e., \( \rho_{u,r}^u \leq \varpi_{u,r}^u \) which naturally yields the transformation of objective as in (**•••), \( \min \prod_{c,r,u} \left( \varpi_{u,r}^u \right)^{\beta_{c,r}} \). Finally, \( P_2^o \) can be adjusted to GP framework by maximizing a scalar auxiliary variable \( \varphi \) which is set as a common QoS threshold for all users, i.e., \( \varpi_u > \varphi, \forall u \). Notice that it is not possible for \( \varpi_2 \) and \( \varpi_3 \) to jointly optimize power and bandwidth since the base and power of \( \left( \varpi_{u,r}^u \right)^{\beta_{c,r}} \) are both optimization variables. Fortunately, this can be handled by primal-dual decomposition and projected subgradient method \[19\].

V. NUMERICAL RESULTS AND ANALYSIS

For the simulations, average macrocell size is arranged to be 1 km \( \times \) 1 km. QoS requirements of users are uniformly distributed with a mean of 1.5 Mbps. Unless it is stated explicitly otherwise, we use the default simulation parameters given in Table I. Moreover, we prefer to present results in a normalized scale for a better comparison of different cases within the same plot.

**TABLE I: Table of Simulation Parameters**

<table>
<thead>
<tr>
<th>Par.</th>
<th>Value</th>
<th>Par.</th>
<th>Value</th>
<th>Par.</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>( \eta' )</td>
<td>( 2 )</td>
<td>( \sigma_{u}^2 )</td>
<td>( 10 ) dBm</td>
<td>( E[</td>
<td>H_u^c</td>
</tr>
<tr>
<td>( N_0 )</td>
<td>( -174 ) dBm</td>
<td>( \epsilon_{10} )</td>
<td>( 10^{-10} )</td>
<td>( P_{u} )</td>
<td>( -100 ) dBm</td>
</tr>
<tr>
<td>( P_{u} )</td>
<td>( 25 ) dBm</td>
<td>( \delta_{10} )</td>
<td>( 46 ) dBm</td>
<td>( \lambda_u )</td>
<td>( 30 ) dBm</td>
</tr>
<tr>
<td>( B )</td>
<td>20 MHz</td>
<td>( \lambda_{u} )</td>
<td>10</td>
<td>( \lambda_u )</td>
<td>100</td>
</tr>
</tbody>
</table>

Fig. 3 compares the analytical findings obtained in (11) with the simulations where the reference user, UE\textsubscript{r}, placed 10 m away from the MBS with a deviation of 2 dB shadow fading. UE\textsubscript{r} is placed in \( (100, 300, 1000) \) m away with \( (0, 4, 8) \) dB deviation which results in \( \mu_B = \{10, 15, 20\} \) dB and \( \sigma_B = \{2, 6, 10\} \) dB, respectively. As \( \rho = \tilde{\gamma}_{u,r}^u/N_0B \) reaches up to 100 dB, simulation converges to the upperbound in (11). Please note that for \( P_u \) and \( N_0 \) given in Table I practical values
of $\rho$ ranges from 245 dB to 320 dB for bandwidths ranging from unity to 20 MHz. That is, analytical upperbound is tight enough for practical values of $\rho$.

The impacts of receiver sensitivity and residual interference is demonstrated in Fig. 4 where the NOMA performance starts to decrease after 90 dBm receiver sensitivity. On the other hand, the sumrate monotonically decreases with increasing error factor. Fig. 5 compares the NOMA performance under DUDe and DUCo schemes for different cluster sizes. In Fig. 5a, network sumrate increases as number of IC increases. DUDe gives the best performance since it associates UEs with nearby BSs. On the contrary, performance of DUCo reduces as more users are associated with the MBS. In Fig. 5b, our findings show that max-min fairness decreases as $M$ increases since max-min fairness tends to close the SINR gap between good and bad users which is limited by PDSCs. Putting more users into the same cluster open this gap wider as it yields more binding constraints on power weights.

Finally, Fig. 6 demonstrates the impact of UE density for different $M$ values. Similar to Fig. 5a increasing the number of IC, $M$, provides a better performance in Fig. 6a. Unlike the DL-NOMA where power is limited by MBS hardware, each additional user comes with its own power allowance which yields a monotonic increases with respect to density. However, it is worth noting that performance increase with user density start to become gradual as more users come up with extra interference some of which cannot be cancelled. On the other hand, Fig. 6b illustrates the effect of user density on max-min fairness. Even though cluster size is fixed for a given $M$, increasing user density increase the number of clusters of size $M + 1$. Therefore, limited bandwidth is shared by more clusters, that is why we observe a decrease in max-min fairness as UE density increases.

VI. CONCLUSION

In this paper, we investigated CF and RA for NOMA operated DUDe HetNets. Taking the constraints and imperfections of SIC receivers into consideration, we analytically show that using the same decoding order of DL case limit the NOMA performance by OMA. However, employing an inverse order provide a NOMA gain proportional to channel gain disparity of UEs. Based on this finding, a generic CF method is developed for any graph type, where clusters are iteratively formed using Blossom algorithm. Thereafter, we explain how to put highly non-convex power and bandwidth allocation problem into a convex form using GP. The potential of NOMA in DUDe HetNets is then demonstrated with numerical results.

REFERENCES


