ICC 2022 - IEEE International Conference on Communications | 978-1-5386-8347-7/22/$31.00 ©2022 IEEE | DOI: 10.1109/ICC45855.2022.9838985

Resource Allocation for the 4G and 5G Dual-Connectivity Network with NOMA and NR

Tzu-Yu Chen†, Chih-Hang Wang♯, Jang-Ping Sheu†, Guang-Siang Lee♯, and De-Nian Yang♯
†Dept. of Computer Science, National Tsing Hua University, Taiwan
♯Institute of Information Science, Academia Sinica, Taiwan
E-mail: s107062562@m107.nthu.edu.tw, superwch7805@iis.sinica.edu.tw,
sheujp@cs.nthu.edu.tw, and {gslee9822802, dnyang}@iis.sinica.edu.tw

***Abstract*—3GPP has defined Dual Connectivity (DC) to allow a user to access a 4G and a 5G base station (BS) simultaneously. However, the resource allocation for DC is challenging because of not only the co-channel interference between 4G and 5G BSs but also different shapes of Resource Blocks (RBs) for New Radio (NR) and the reuse of RBs for Non-Orthogonal Multiple Access (NOMA). In this paper, we formulate Dual Connectivity Multi-dimensional Resource Allocation Problem and prove that it is NP-hard. We design an approximation algorithm with the ideas of 1) Zone Shaping, 2) Occupancy Indicator and Overlap Degree of RBs, 3) DC Slicing, and 4) DC Inter-Numerology Relation, to maximize the total throughput of heterogeneous user demands in the coexisting 4G and 5G network with NR, NOMA, and DC. Simulation results manifest that our algorithm outperforms the state-of-the-arts regarding throughput and resource efficiency.**

I. INTRODUCTION

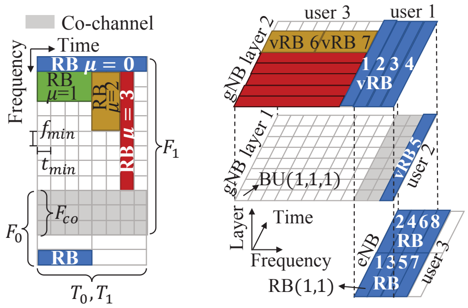
The Fifth Generation (5G) radio access has emerged for the need of scalable and faster transmission [1], but the Fourth Generation (4G) and the 5G networks are expected to coexist before the standalone 5G is widely deployed [2]. To this end, 3rd Generation Partnership Project (3GPP) has defined Dual Connectivity (DC) [3] to enable a user to simultaneously access a 4G base station (BS), called E-UTRAN NodeB (eNB), and a 5G BS, called Next Generation NodeB (gNB) [4], [5]. However, the co-channel interference occurs when the eNB and gNB share an identical frequency band [5], [6]. Shi et al. [5] minimized the co-channel interference to maximize system throughput in heterogeneous networks. Liu et al. [6] explored joint radio resource and power allocation in the network with DC and Non-Orthogonal Multiple Access (NOMA). However, the above works did not explore the Resource Block (RB) allocation in the coexisting 4G and 5G network with New Radio (NR) and NOMA.

3GPP introduced NR in 5G for scalable radio resource management [7]. Different from the single numerology in 4G [8], 5G includes multiple numerologies to enable different shapes of RBs with unique Subcarrier Spacing (SCS) and Transmission Time Interval (TTI) for supporting different service requirements [7]. For example, the RBs with a larger SCS, i.e., the height in the frequency domain, and a smaller TTI, i.e., the width in the time domain, are suitable for real-time services like vehicle-to-vehicle communications [2], [9]. In contrast, the RBs with a smaller SCS and a larger TTI are appropriate for low-volume and delay-tolerant sensing applications. However, when different numerologies are transmitted in the same frequency band in a time slot, Inter-Numerology Interference (INI) occurs and severely degrades the system performance [9], [10]. McWade et al. [9] analyzed the data rate under INI in a multi-numerology system, and the RB allocation under INI for maximizing the weighted sum rate was further explored in [10]. However, the above works did not consider the RB reuse between 4G and 5G by DC, where a user can be served by the two BSs simultaneously for a more flexible RB allocation.

To increase the spectrum efficiency in 5G, NOMA permits multiple users to be multiplexed on a single RB [11], [12]. The signal for different users is multiplexed in the power-domain at the transmitter and decoded with the support of Successive Interference Cancellation (SIC) at the receiver [12], [13]. Liu et al. [12] applied NOMA to enable massive device connectivity for energy-efficient networks. Ni et al. [13] allocated channels and power in multi-cell NOMA networks. NOMA can increase the network capacity more flexibly since the user data are allowed to be delivered by an RB shared with other users. However, the above works did not explore the DC network with NOMA and mixed numerology, where INI occurs if the RB is shared by the users with different numerologies.

In this paper, we make the first attempt to explore the RB allocation in the coexisting 4G and 5G networks with DC, NR, and NOMA to maximize the system throughput. However, the problem is challenging since it introduces the following new research issues. 1) *Resource fragmentation with mixed numerology in DC*. With NR, each numerology leads to a different RB shape, and the locations of virtual RBs (vRBs) with different shapes are allowed to be unaligned,[1] leading to scattered empty spaces and resource fragmentation. A small fragmented space is difficult to satisfy a data-intensive user. Hence, it is important to allocate the data to the eNB and gNB simultaneously by DC to properly exploit the fragmented spaces for maximizing throughput. For example, in Fig. 1(b), when user 3 obtains vRBs 6 and 7 in a fragmented space, if the data rate is insufficient, it also needs to allocate RBs 1-8 from eNB to avoid INI in the gNB (detailed in Section II).

[1] To demonstrate the radio reuse of NOMA in the gNB, an RB is logically regarded as multiple vRBs in different layers of a 5G resource grid [10].



(a) Resource grids (b) An example of DC

Fig. 1. An illustrative example of RB allocation with NOMA and DC.

2) *Reused RB locations with NOMA in DC.*[2] With NOMA, it is desired to overlap multiple vRBs with sufficiently large differences in channel gains for increasing system throughput, but INI occurs when those overlapped vRBs use different numerologies. DC satisfies a user rate requirement by the eNB and gNB simultaneously such that fewer RBs in the gNB are allocated for the user, which leads to smaller reused locations of vRBs and thereby reduces INI. Hence, it needs to jointly decide the portion of user data and their locations in the 4G and 5G resource grids to lower INI and co-channel interference. 3) *Heterogeneity of users under different BSs.* The feasible candidate numerologies, leading to different vRB shapes, and the number of vRBs required by users are heterogeneous and derived according to their QoS demands and channel quality under different BSs. It limits the flexibility in RB allocation since a user can only be served by the vRBs from the feasible candidate numerologies. Moreover, improper RB locations in the BSs will occupy a large space due to worse channel quality, and it thereby needs to carefully decide the portion and locations of RBs in the 4G and 5G resource grids for each DC user.

To tackle the above issues, we formulate Dual Connectivity Multi-dimensional Resource Allocation Problem (DC-MDRAP) to maximize the system throughput under heterogeneous user demands in the coexisting 4G and 5G network with NR, NOMA, and DC. Different from previous research exploring only NOMA-based [6] or OFDM-based DC [1], [2], DC-MDRAP considers the NOMA-based mixed-numerology DC system with different shapes of RBs. We prove that DC-MDRAP is NP-hard and design an approximation algorithm, named Dual Connectivity-Scattered Sharing Resource Manager (DC-SSRM), with the ideas of 1) Zone Shaping, 2) Occupancy Indicator and Overlap Degree of RBs, 3) DC Slicing, and 4) DC Inter-Numerology Relation, to properly address the aforementioned challenges. Simulation results manifest that DC-SSRM outperforms the state-of-the-art algorithms regarding throughput, resource efficiency (i.e., average bits/RB), and user satisfaction.

[2] Since an RB includes multiple Base Units (BUs) [10], we call that two vRBs are *overlapped* if they contain common BUs of the RB, and these BUs form a reused location of the RB. More details are explained in Section II-A. The term "*overlap*" is also regarded as the co-channeled RBs and vRBs in the eNB and gNB.

The rest of this paper is organized as follows. Section II introduces the system model and formulates DC-MDRAP. Section III presents the approximation algorithm DC-SSRM. Section IV summarizes the simulations, and Section V draws the conclusions.

II. System Model and Problem Formulation

A. Resource Grid

According to the 3GPP standard [8], the resource grid of 4G LTE comprises time and frequency domains. In the time domain, the duration of a frame is 10 ms. The frame is further equally divided into 10 subframes, and each subframe contains two equal-duration slots. The basic scheduling unit is RB, which is composed of one slot and 12 consecutive subcarriers with 15 kHz SCS in the frequency domain. The RB shape is fixed for 4G LTE, but it varies according to numerology for 5G NR [7].

An NR RB is constructed by 12 subcarriers, and the numerology μ defines SCS and the slot duration. SCS follows the formula $15 \times 2^{\mu}$ kHz, and the slot length follows $1 \times 2^{-\mu}$ ms for $\mu \in \mathbb{N} = \{0, 1, \dots, \mu_{max}\}$, where $\mu_{max} = 4$ denotes the maximum numerology [10], and $\mathbb{N}$ is the numerology set. According to [14], the basic expression unit of an RB is defined as Base Unit (BU). As shown in Fig. 1(a), each RB consists of $2^{\mu_{max}}$ BUs [14]. Each BU occupies the bandwidth of $f_{min} = 12 \times 15$ kHz and the time slot duration of $t_{min} = 1 \times 2^{-\mu_{max}}$ ms. Consequently, the numerology determines the frequency span and the slot duration of an RB, and the shape of the RB is a $(f_{min} \cdot 2^{\mu}) \times (t_{min} \cdot 2^{\mu_{max}-\mu})$ rectangle constructed by $2^{\mu} \times 2^{\mu_{max}-\mu}$ BUs. Note that an LTE RB is half the size of an NR RB with $\mu = 0$, which is $f_{min} \times (t_{min} \cdot 2^{\mu_{max}})$. Hence, an LTE RB is in the shape of $f_{min} \times (t_{min} \cdot 2^{\mu_{max}-1})$.

Since NOMA allows an RB to be reused by several users, an NR frame consists of multiple *layers* to represent different layers in the superposition coding scheme of NOMA [10]. Hence, an RB is regarded as multiple vRBs in different layers, and each vRB is mapped to the RB with the corresponding BUs. In Fig. 1(b), for example, vRB 5 on layer 1 overlaps with vRB 4 on layer 2, and the overlapped area is the *reused locations of RBs*.

B. Problem Formulation

Due to the space constraint, the notation table is presented in [15]. Following [4], we consider the network including an eNB and a gNB partially overlapping with each other, and a user can associate to two BSs simultaneously [16], where each BS has limited RBs [10]. Let F_k and T_k respectively be the frequency span and time duration in the resource grid of BS k, and the grid size of BS k is $F_k \times T_k$, where $k \in \{0, 1\}$ denotes {eNB, gNB}, respectively. To indicate the location of a vRB in the gNB, a specific BU is denoted by $\{(l, i, j), 1 \le l \le L, 1 \le i \le F_1, 1 \le j \le T_1\}$, where L is the maximum number of layers, and F_1 and T_1 are the number of BUs in the frequency and time domain, respectively. Since an RB consists of multiple adjacent BUs [14], a particular vRB with numerology μ in layer l is then mapping to BUs given by $\{(l,i,j), f \le i \le f+2^{\mu}-1, t \le j \le t+2^{\mu_{max}-\mu}-1\}$,[3] where (l,f,t) is the top-left BU forming the vRB, and f and t are the indexes for a vRB in frequency and time domains, respectively. Following [5], the bandwidth of the sharing spectrum in the eNB and gNB is $F_{co} \cdot f_{min}$, where F_{co} is the number of BUs in the sharing spectrum. We assume the sharing spectrum is at the lowest frequency of the eNB and at the highest frequency of the gNB [5], i.e., the gray part in Fig. 1(a). Let $\mathbb{U}=\{1,\dots,|\mathbb{U}|\}$ be the set of users. Let $\mathbb{D}_u$ represent the set of BSs that cover user u, and u can access the 4G and 5G BSs simultaneously if u is under their coverage. Each user u has a data rate demand q_u and a candidate set of numerologies $\mathbb{C}_u \subseteq \mathbb{N}$.

Equipped with the above model, DC-MDRAP is to allocate RBs to a subset of users to satisfy their rate demands with the following constraints. 1) *RB allocation constraint*[4] [10], [17]: Each RB b can be allocated to at most one user, i.e., $\sum_{u\in\mathbb{U}} \beta_{u,b} \le 1, \forall b \in \mathbb{B}_0 \cup \mathbb{B}_1$, where $\beta_{u,b}$ denotes whether RB b is allocated to user u, and $\mathbb{B}_0$ and $\mathbb{B}_1$ are the sets of RBs in the eNB and gNB, respectively. For 5G, each RB contains a fixed number of BUs, i.e., $\sum_{i=1}^{F_1}\sum_{j=1}^{T_1} \alpha_{b,l,i,j} = 2^{\mu_{max}}, \forall b \in \mathbb{B}_1, \forall 1 \le l \le L$, where $\alpha_{b,l,i,j}$ is a binary variable indicating if RB b contains BU (l,i,j). 2) *RB reuse constraint* [10], [18]: The number of layers for RB reuse is limited due to the hardware limitation in the NOMA-based system [10], i.e., $\sum_{b\in\mathbb{B}_1}\sum_{u\in\mathbb{U}} \alpha_{b,l,i,j}\beta_{u,b} \le L, \forall 1 \le i \le F_1, \forall 1 \le j \le T_1$. 3) *NR Bandwidth Part (BWP) constraint* [7]: User u can use at most one numerology in each time slot j in the gNB. 4) *Robust rate constraint* [17], [19]: The Modulation and Coding Scheme (MCS) of each RB in a BS allocated to a user is the lowest available MCS level among all the RBs allocated to the user leading to the most robust rate. Let $R_{u,b}$ be the data rate of RB b allocated to user u. If RB b is in the gNB, i.e., $b \in \mathbb{B}_1$, $R_{u,b} = f_{min} \cdot 2^{c_u} \cdot \log_2(1+\gamma_{u,b})$; otherwise, $R_{u,b} = f_{min} \cdot \log_2(1+\gamma_{u,b}), \forall b \in \mathbb{B}_0$, where c_u is the numerology of u and $\gamma_{u,b}$ is the received SINR of user u on RB b [6], [20].[5] Hence, the data rate of user u is $R_u = \sum_{k\in\mathbb{D}_u} M_{u,k} \cdot \min_{b\in\mathbb{B}_k, \beta_{u,b}=1}\{R_{u,b}\} \ge q_u$ for satisfying the user demand, where $M_{u,k}$ is the number of allocated RBs to user u by BS k. 5) *DC constraint* [3]: The connections of a user are set according to the coverage of BSs, i.e., $\sum_{k\in\{0,1\}} \delta_{u,k} \le |\mathbb{D}_u|, \forall u \in \mathbb{U}$, where $\delta_{u,k}=1$ indicates that user u is associated with BS k. In other words, for BS k, $\delta_{u,k}=1$ if and only if $M_{u,k}>0, \forall u \in \mathbb{U}$.

Definition 1. Given a set of users $\mathbb{U}=\{1,\dots,|\mathbb{U}|\}$ with data rate demand q_u, a candidate numerology set $\mathbb{C}_u$, and a set of BSs $\mathbb{D}_u$ that can be connected for each user $u \in \mathbb{U}$, and the resource grids with $F_0 \times T_0$ RBs for the eNB and $L \times F_1 \times T_1$ BUs for the gNB,[6] DC-MDRAP is to allocate RBs to a subset of users to satisfy the users' rate demands and the *RB allocation*, *RB reuse*, *NR BWP*, *Robust rate*, and *DC constraints*. The objective is to maximize system throughput $\sum_{u\in\mathbb{U}} R_u$.

Theorem 1. *DC-MDRAP is NP-hard.*

Proof. The detailed proof is in [15] due to the limited space. □

III. Algorithm

To address DC-MDRAP, an intuitive method is to iteratively allocate the RBs from the top-left of the resource grid in each layer[7] to the users with the largest SINR on these RBs, and the numerology is the maximum one chosen from the users' candidate numerologies [13]. If a user can simultaneously access to the eNB and gNB, the gNB RBs are first allocated since the user can benefit from NOMA to increase throughput, and the remaining user demand is served by eNB RBs. However, the reused locations of RBs and INI are not carefully examined and thereby severely degrade system throughput.

In the following, we propose DC-SSRM, which includes the following new ideas to address the challenging issues listed in Section I. For the first challenge, DC-MDRAP derives a bounded allocation area, named *zone*, for each user, and Zone Shaping (ZS) merges small zones into a larger one according to their used numerologies for avoiding resource fragmentation. For the second challenge, we introduce Occupancy Indicator (OI) and Overlap Degree (OD) of vRBs to examine their occupied space and overlapped areas for increasing resource reuse efficiency, and DC Slicing (DCS) iteratively replaces the RBs in the gNB of a DC user by those with higher data rates in the eNB, to further enhance system throughput. For the last one, DC Inter-Numerology Relation (DC-INR) evaluates the occurrence of numerologies in users' candidate numerology sets to prioritize them for avoiding INI due to heterogeneous users. The overall complexity of DC-SSRM is $O(|\mathbb{U}|F^2T^2)$, where $F=\max\{F_0,F_1\}$ and $T=\max\{T_0,T_1\}$. Due to the space constraint, detailed complexity analysis and pseudocode of DC-SSRM are presented in [15].

A. Zone Shaping (ZS)

To reduce INI and fragmented RBs, for each user, ZS first finds the feasible numerology by examining DC-INR and creates a *zone*. Specifically, we define DC-INR of numerology μ as $N_\mu = \sum_{u\in\mathbb{U},\{1\}\subseteq\mathbb{D}_u} |\mathbb{C}_u \cap \{\mu\}| \cdot \frac{|\mathbb{N}-\mathbb{C}_u|}{|\mathbb{D}_u|}, \forall \mu \in \mathbb{N}$, which implies that μ has higher priority when it can be utilized by more users to avoid INI.[8] Moreover, if the user can connect to multiple BSs, INI can be alleviated by reducing the number of allocated RBs in the gNB via DC. Hence, the user u with a larger $|\mathbb{D}_u|$ is associated with a lower priority to μ since it

[3] For ease of presentation, we logically regard the BUs with the same location in the frame but in different layers as different BUs.

[4] For ease of presentation, unless stated otherwise, we use "RB" to refer to the RB in 4G and vRB in 5G, which are the smallest allocation units.

[5] Due to the space constraint, the detailed calculations of SINR are provided in [15].

[6] Since the shapes of 5G RBs vary, we use BU as the minimum unit to indicate the locations of RBs.

[7] The allocation in eNB can be regarded as a single layer.

[8] Since μ can be utilized by more users, there are more chances to overlap these users with the identical numerology in different layers to lower INI.

induces a smaller INI due to DC. For each user u, ZS then extracts the numerology μ with the largest DC-INR in $\mathbb{C}_u$ and forms a zone z by evaluating the number of required RBs in the worst case for satisfying the user demand.[9] Specifically, let $E_u = \lceil \frac{q_u}{r^k_{min}} \rceil$ be the maximum number of required RBs, where r^k_{min} is the minimum achievable data rate of an RB with the most robust MCS in BS k. A zone z for a user is a rectangle region to ensure that the required RBs in the worst case can be allocated to the user, where the width W_z in the time domain is $\min(E_u, T_0)$ ($W_z = \min(E_u \cdot 2^{\mu_{max}-\mu_z}, T_1)$) for the eNB (gNB),[10] the height H_z in the frequency domain is $\lceil \frac{E_u}{T_0} \rceil$ ($H_z = 2^{\mu_z} \lceil \frac{E_u \cdot 2^{\mu_{max}-\mu_z}}{T_1} \rceil$) for the eNB (gNB), and μ_z is the numerology used for zone z. The RBs in each zone are first placed along the time domain, to avoid spanning the channel with poor quality (inducing worse MCS).

In order to avoid resource fragmentation, ZS merges the zones with the identical numerology according to the time duration of required RBs, and these zones will be allocated in the resource grids in the next phase ZPA. Specifically, for each BS k, ZS first finds each zone z with $W_z < T_k$, and it then sorts them by W_z in the descending order, because a zone with smaller W_z is easier to be merged without exceeding T_k. ZS then iteratively chooses a zone to be concatenated along the time axis to form a new zone $\hat{z}$ no wider than T_k, until no zone can be concatenated. The above process repeats until there is no zone that can be merged. Afterwards, ZS classifies all zones (i.e., merged and individual zones) into two sets $\mathbb{Z}_{large}$ and $\mathbb{Z}_{small}$ according to W_z for the allocation in the next phase ZPA. If $W_z \geq \frac{1}{2}T_k$,[11] ZS adds z to $\mathbb{Z}_{large}$; otherwise, z is added to $\mathbb{Z}_{small}$.

Example 1. Let $F_0 \times T_0 = 7 \times 2$, $F_1 \times T_1 = 13 \times 8$, $L = 3$, $F_{co} = 3$, and $|\mathbb{U}| = 11$. Fig. 2(a) presents the configurations of users. The DC-INR value of $\mu = 0$ is $N_0 = \frac{3}{2} + \frac{2}{1} + \frac{3}{1} = 6.5$, and $(N_0, N_1, N_2, N_3) = (6.5, 5.5, 5.5, 8.5)$. For u_2, $\mu = 3$ is selected since $N_3 > N_2$ with $\mathbb{C}_2 = \{2, 3\}$. The zone z with $W_z < 8$ is merged as listed in Fig. 2(b) (with detailed derivation in [15]) with the renumbered zone index $\hat{z}$. For example, z_5 and z_4 are merged into $\hat{z}_8$, since z_6 is too wide to be merged with either of them. Finally, zones $\hat{z}$ with $W_{\hat{z}} \geq 4$ are added to $\mathbb{Z}_{large} = \{1, \dots, 8\}$. Fig. 2(c) shows the result for eNB with the identical merging process.

u, z	1	2	3	4	5	6	7	8	9	10	11
$\mathbb{D}_u$	{0,1}	{1}	{1}	{0,1}	{1}	{1}	{1}	{1}	{1}	{0}	{0}
q_u	110	175	108	53	79	112	48	40	60	18	74
$\mathbb{C}_u$	{0}	{2,3}	{0,1}	{1,2,3}	{3}	{3}	{0}	{1}	{2}	{}	{}
c_u, μ_z	0	3	0	3	3	3	0	1	2	-	-

(a) Input settings and configuration of users

$\hat{z}$	1	2	3	4	5	6	7	8
$\mu_{\hat{z}}$	0	3	0	0	1	2	3	3
$\mathbb{U}_{\hat{z}}$	{1}	{2}	{3}	{7}	{8}	{9}	{6}	{5,4}

(b) Merged zones in the gNB

$\hat{z}$	1	2
$\mathbb{U}_{\hat{z}}$	{10}	{11}

(c) Merged zones in the eNB

Fig. 2. An illustrative example of ZS.

B. Zone Piling and Allocation (ZPA)

For each resource grid of the eNB and gNB, ZPA first allocates the zones in $\mathbb{Z}_{large}$ to the first layer according to OI,[12] by evaluating the percentage of occupied space. Specifically, for each zone z of gNB, $OI_z = \frac{2^{\mu_{max}} \cdot \sum_{u \in \mathbb{U}_z} E_u}{H_z \cdot T_1}$, where $H_z \cdot T_1$ is the number of BUs preserved for zone z, and $2^{\mu_{max}} \cdot \sum_{u \in \mathbb{U}_z} E_u$ is the number of BUs occupied by the RBs in zone z.[13] On the other hand, $OI_z = \frac{2^{\mu_{max}-1} \cdot \sum_{u \in \mathbb{U}_z} E_u}{H_z \cdot T_0}$ for zone z of the eNB. A zone with a larger OI means that it contains less empty space (i.e., most BUs in the zone are occupied to serve users) for maximizing throughput. To minimize the empty space in the resource grid for the first layer, ZPA iteratively allocates the zone with the largest OI to the vacant location with the smallest index in the frequency domain.[14]

For each layer $2 \leq l \leq L$,[15] ZPA starts from the lowest layer and iteratively allocates the remaining zones in $\mathbb{Z}_{large}$ to the vacant location with the smallest index according to OD. Specifically, $OD_{l,z} = \sum_{l'=1}^{l-1} \frac{G_{l',\mu_z}}{H_z \cdot T_1}$ represents the total proportion of BUs with the numerology identical to zone z in each previous layer $1, 2, \dots, l-1$ overlapping with z. G_{l',μ_z} is the number of BUs with numerology μ_z overlapping z on layer l' if z is allocated. A zone with a higher OD potentially induces a smaller INI, since it overlaps with more BUs with the same numerology for increasing RB reuse efficiency. For each layer l from 2 to L, ZPA iteratively allocates the zone z with the largest $OD_{l,z}$ to the vacant location with the smallest index, in order to reduce INI. ZPA stops when there is no zone can be allocated into any layer.

Example 2. Following Example 1, $\hat{z}_1$ and $\hat{z}_2$ are allocated to layer 1 because they have the highest OI and smallest index. Next, $\hat{z}_3$ and $\hat{z}_8$ are allocated to layer 2 because they have the largest OD among the zones in $\mathbb{Z}_{large}$ (with detailed derivation in [15]). The allocation of gNB after ZPA is shown in Fig. 3(a). The allocation of eNB is illustrated in Fig. 3(b) by OI with detailed derivation in [15].

[9]If a user can only be associated to eNB, it does not need to determine the numerology. If the user can be simultaneously associated to eNB and gNB, ZS considers only gNB when forming the zone, since 5G can provide a higher data rate, and DC will be examined in DCS.

[10]We present W_z with the unit of BU to depict the size of zones with different numerologies.

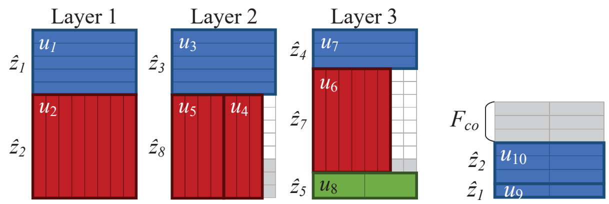
[11]Due to the space constraint, we prove that $\frac{1}{2}$ is crucial to ensure the approximation ratio in [15].

[12]Due to the high capacity of gNB with the NOMA technique, the co-channel spectrum will be preserved for the gNB in ZPA for higher resource efficiency. DC and unallocated users in $\mathbb{Z}_{small}$ will be examined in DCS.

[13]Recall that after merging zones in ZS, the height of z is $H_z = \lceil \frac{\sum_{u \in \mathbb{U}_z} E_u}{T_0} \rceil$ in the eNB or $H_z = 2^{\mu_z} \lceil \frac{\sum_{u \in \mathbb{U}_z} E_u \cdot 2^{\mu_{max}-\mu_z}}{T_1} \rceil$ in the gNB, where $\mathbb{U}_z$ is the set of users forming zone z.

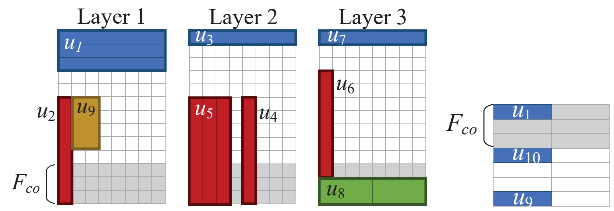
[14]The zones are allocated along the frequency domain, since the width of each zone is at least $\frac{1}{2}T_k$ in $\mathbb{Z}_{large}$ (detailed in ZS).

[15]Since NOMA is unavailable in 4G, there is only a single layer for the eNB, and the remaining process of ZPA is only for the gNB.



(a) gNB (b) eNB

Fig. 3. An illustrative example of ZPA.



(a) gNB (b) eNB

Fig. 4. An illustrative example of DCS.

C. DC Slicing (DCS)

DCS adjusts the MCS of each allocated user to release more available space for maximizing the system throughput. Then, DCS slices the RBs of DC users to be allocated in each BS such that the RBs incurring INI can be substituted by the RBs in another BS. Specifically, DCS assigns MCS $\min_{b\in\mathbb{B}_k,\beta_{u,b}=1}\{R_{u,b}\}$ for each allocated user u in BS k according to the channel condition of the occupied RBs.[16] Then, DCS releases $\lceil\frac{q_u}{r^k_{min}}\rceil-\lceil\frac{q_u}{\min_{b\in\mathbb{B}_k,\beta_{u,b}=1}\{R_{u,b}\}}\rceil$ RBs from the largest to the smallest index in the time domain, because r^k_{min} is the worst MCS and fewer RBs are required to satisfy the user after improving MCS.

For the DC users, DCS replaces some RBs in the gNB with those in the eNB to serve the users with better channel quality.[17] For each DC user, DCS iteratively examines the data rate of contiguous RBs from the smallest index to the largest one in the frequency domain to partition the RBs into two sets. Let $\mathbb{RB}^{SF}_u$ and $\mathbb{RB}^{LF}_u$ be the set of contiguous RBs of u in the locations with the smaller and larger indexes in the frequency domain, respectively. To reduce INI, DCS finds the frequency index that leads to the largest total throughput increment on the RBs overlapping with $\mathbb{RB}^{LF}_u$ if $\mathbb{RB}^{LF}_u$ is removed. Then, DCS allocates user u to the empty rectangle location in the eNB, that can provide at least the data rate of $\mathbb{RB}^{LF}_u$ with the best channel quality or the lowest co-channel interference if the location is in the sharing spectrum. The above process repeats until every DC user is examined. For the rest of the empty space, DCS iteratively allocates the remaining users (zones) with the least number of RBs to the location that results in the largest channel gain difference for maximizing total throughput by NOMA.

Example 3. Following Example 2, Fig. 4 shows the resource grids after DCS, and we take u_1 as an example for DCS. Since $\{0\}\subset\mathbb{D}_1$ and removing the two vRBs in higher frequency can improve the efficiency of vRBs for u_6 by avoiding INI, u_1 obtains a PRB at $(1,1)$ in the eNB as illustrated in Fig. 4(b) with the lowest co-channel interference. Lastly, a new user u_9 is allocated to BU $(1,6,2)$ in Fig. 4(a), since it overlaps with u_5 and induces the highest channel gain difference.

[16]Recall that each zone is formed with MCS r_{min} in ZS, and DCS maximizes the throughput with higher MCS to ensure the approximation ratio.

[17]In order to achieve a higher resource efficiency by NOMA, ZPA has carefully examined the RB allocation in the gNB for DC users. In DCS, we further improve the allocation by the channel with better quality (i.e., less interference) in the eNB.

Theorem 2. *DC-SSRM is a $\frac{1}{8}c$-approximation algorithm for DC-MDRAP, where $c=\min_{k\in\{0,1\}}\frac{r^k_{min}}{r^k_{max}}$, and r^k_{min} and r^k_{max} respectively denote the lowest and the highest achievable rate of an RB in BS k.*

Proof. The detailed proof is in [15] due to the limited space. □

IV. Simulation

A. Simulation Settings

We evaluate the performance of DC-SSRM in a DC mixed numerology NOMA-based system with one eNB partially covering a gNB, the radii of the BSs are both 500 m [21]. We randomly distribute 300 users in the whole BS coverage area [5]. The users' rate demands are set from 100–500 Kbps [14]. Following [10], we consider four numerologies with (SCS, TTI) respectively being (15 kHz, 1 ms), (30 kHz, 0.5 ms), (60 kHz, 0.25 ms), and (120 kHz, 0.125 ms) [7], and the number of layers for RB reuse is $L=4$ [10]. The default bandwidth of the two BSs is 40 MHz [22] with 5 MHz bandwidth shared between the two BSs. Following [5], [10], [12], the maximum transmit power of the BSs is set to 46 dBm by default, and the total noise power spectral density is -174 dBm/Hz. Meanwhile, we consider the log-normal shadowing with a standard deviation 8 dB [5], [10], and the path loss model is the macro propagation in outdoor urban areas [5], [10]. The MCS is in the range of QPSK-16QAM [17], [19].

Since there is no previous research exploring the RB allocation in 4G and 5G networks with DC, NOMA, and NR, we compare DC-SSRM with the baseline (i.e., the intuitive method described in Section III) and the RB allocation algorithms: 1) FRSA [10] and 2) MSEMA [6] integrated with MCUP [1], which determines the user association for DC. We change the following parameters: 1) the percentage of DC users, 2) co-channel bandwidth, 3) the bandwidth of gNB, and 4) the number of layers, to evaluate the performance metrics of system throughput, resource efficiency (i.e., average bits/RB), and the satisfaction ratio of users. Each result is averaged over 100 times. Due to the limited space, we provide more simulations in [15].

B. Simulation Results

In Figs. 5(a) and 5(b), the performance gaps between DC-SSRM and the other algorithms grow as the number of DC users increases, because DCS carefully determines the proportion and locations of RBs allocated in the two BSs,

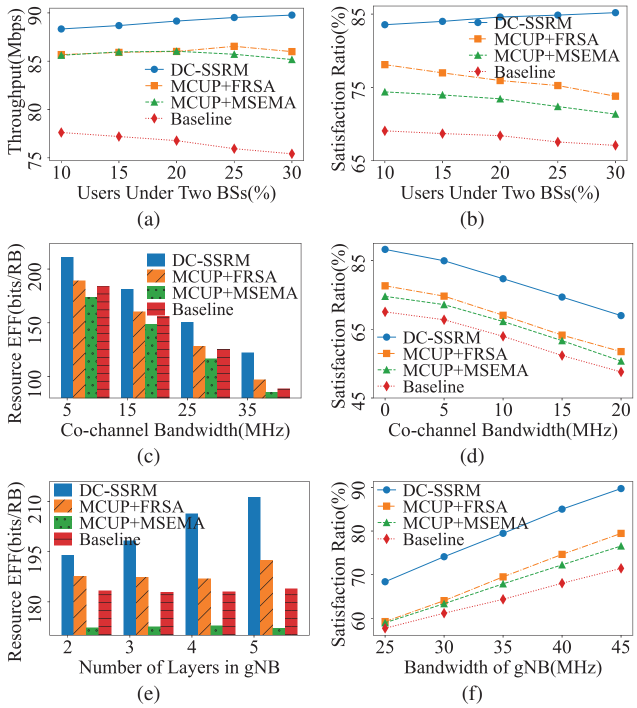


Fig. 5. Performance of different algorithms.

to avoid INI and co-channel interference for increasing the system throughput with a smaller number of RBs. In contrast, FRSA and MSEMA with MCUP maximize the throughput for every single user and ignore the interference between eNB and gNB during the association of DC users, leading to worse throughput and user satisfaction. In Fig. 5(c), when the co-channel bandwidth grows, higher interference occurs and degrades resource efficiency. However, DC-SSRM outperforms other algorithms since it exploits OI to overlap the zones with a smaller empty space for maximizing the throughput. Moreover, it overlaps the zones with the identical numerology for avoiding INI by OD. With a better RB reuse, DC-SSRM serves more than 20% of users compared with the baseline in Fig. 5(d). Fig. 5(e) shows that DCS generates much higher resource efficiency when the number of layers increases, since it properly exploits NOMA to allocate the user to the location of RBs with a sufficiently large channel gain difference. In Fig. 5(f), DC-SSRM can serve more users when the bandwidth increases, since it creates zones to avoid resource fragmentation by evaluating DC-INR. In contrast, the baseline allocates the RBs to the user with the largest SINR for throughput maximization, but it ignores INI and NOMA for serving a number of users by properly reusing RBs in different locations.

V. Conclusions

To the best of our knowledge, this paper makes the first attempt to explore the RB allocation in the coexisting 4G and 5G network with NOMA, NR, and DC. We formulate DC-MDRAP and prove its NP-hardness. Then, we design an approximation algorithm DC-SSRM with the ideas of DC-INR, OI, OD, and DCS to allocate RBs for the selected users with heterogeneous rate demands to maximize system throughput. Simulation results show that DC-SSRM outperforms state-of-the-art algorithms regarding throughput, resource efficiency, and user satisfaction. In the future, we will further explore RB allocations for user mobility across multiple BSs with DC.

References

[1] G. Simsek, H. Alemdar, and E. Onur, “Multi-connectivity enabled user association,” in *Proc. IEEE 30th PIMRC*, 2019, pp. 1–6.

[2] M. Agiwal, H. Kwon, S. Park, and H. Jin, “A survey on 4G-5G dual connectivity: Road to 5G implementation,” *IEEE Access*, vol. 9, pp. 16 193–16 210, 2021.

[3] 3GPP, “Universal mobile telecommunications system (UMTS); LTE; 5G; NR; multi-connectivity; overall description; stage-2,” 3GPP, Technical Specification (TS) 37.34, Apr. 2021.

[4] C. Li, H. Wang, and R. Song, “Intelligent offloading for NOMA-assisted MEC via dual connectivity,” *IEEE Internet of Things J.*, vol. 8, no. 4, pp. 2802–2813, 2021.

[5] Y. Shi, H. Qu, and J. Zhao, “Dual-connectivity enabled resource allocation approach with eICIC for throughput maximization in hetnets with backhaul constraint,” *IEEE Wireless Commun. Lett.*, vol. 8, no. 4, pp. 1297–1300, 2019.

[6] B. Liu and M. Peng, “Joint resource block-power allocation for NOMA-enabled fog radio access networks,” in *Proc. IEEE ICC*, 2019, pp. 1–6.

[7] 3GPP, “NR; physical channels and modulation,” 3GPP, Technical Specification (TS) 38.211, Jan. 2020.

[8] ——, “LTE; evolved universal terrestrial radio access (E-UTRA); physical channels and modulation,” 3GPP, Technical Specification (TS) 36.211, Apr. 2020.

[9] S. McWade, M. F. Flanagan, L. Zhang, and A. Farhang, “Interference and rate analysis of multinumerology NOMA,” in *Proc. IEEE ICC*, 2020, pp. 1–6.

[10] R.-J. Wang, C.-H. Wang, G.-S. Lee, D.-N. Yang, W.-T. Chen, and J.-P. Sheu, “Resource allocation in 5G with NOMA-based mixed numerology systems,” in *Proc. IEEE GLOBECOM*, 2020, pp. 1–6.

[11] J. Choi, B. Kim, K. Lee, and D. Hong, “A transceiver design for spectrum sharing in mixed numerology environments,” *IEEE Trans. Wireless Commun.*, vol. 18, no. 5, pp. 2707–2721, 2019.

[12] B. Liu, C. Liu, and M. Peng, “Resource allocation for energy-efficient MEC in NOMA-enabled massive IoT networks,” *IEEE J. Sel. Areas Commun.*, vol. 39, no. 4, pp. 1015–1027, 2021.

[13] W. Ni, X. Liu, Y. Liu, H. Tian, and Y. Chen, “Resource allocation for multi-cell IRS-aided NOMA networks,” *IEEE Trans. Wireless Commun.*, vol. 20, no. 7, pp. 4253–4268, 2021.

[14] L. You, Q. Liao, N. Pappas, and D. Yuan, “Resource optimization with flexible numerology and frame structure for heterogeneous services,” *IEEE Commun. Lett.*, vol. 22, no. 12, pp. 2579–2582, 2018.

[15] T.-Y. Chen, C.-H. Wang, J.-P. Sheu, G.-S. Lee, and D.-N. Yang, “Resource allocation for dual-connectivity with NOMA-based mixed numerology system (full version),” Oct. 2021. [Online]. Available: http://hscc.cs.nthu.edu.tw/paper/NTHUTechRepo2021.pdf

[16] P.-J. Hsieh, W.-S. Lin, K.-H. Lin, and H.-Y. Wei, “Dual-connectivity prevenient handover scheme in control/user-plane split networks,” *IEEE Trans. Veh. Technol.*, vol. 67, no. 4, pp. 3545–3560, 2018.

[17] K. Arshad and S. Rostami, “Resource allocation for multi-carrier cellular networks,” in *Proc. IEEE WCNC*, 2018, pp. 1–6.

[18] L. Dai, B. Wang, Z. Ding, Z. Wang, S. Chen, and L. Hanzo, “A survey of non-orthogonal multiple access for 5G,” *IEEE Commun. Surv. Tutorials*, vol. 20, no. 3, pp. 2294–2323, 2018.

[19] 3GPP, “5G; NR; physical layer procedures for data,” 3GPP, Technical Specification (TS) 38.214, Apr. 2021.

[20] N.-T. Le, L.-N. Tran, Q.-D. Vu, and D. Jayalath, “Energy-efficient resource allocation for OFDMA heterogeneous networks,” *IEEE Trans. Commun.*, vol. 67, no. 10, pp. 7043–7057, 2019.

[21] T. T. Nguyen, V. N. Ha, and L. B. Le, “Wireless scheduling for heterogeneous services with mixed numerology in 5G wireless networks,” *IEEE Commun. Lett.*, vol. 24, no. 2, pp. 410–413, 2020.

[22] 3GPP, “5G; NR; user equipment (UE) radio transmission and reception; part 1: Range 1 standalone,” 3GPP, Technical Specification (TS) 38.101-1, May 2021.