Cost-Effective Deployment for Fully-Decoupled RAN: A Techno-Economic Approach

Jiwei Zhao, *Member, IEEE*, Jiacheng Chen, *Member, IEEE*, Bo Cheng, *Member, IEEE*, Bo Qian, *Member, IEEE*, Yunting Xu, *Student Member, IEEE*, and Haibo Zhou, *Senior Member, IEEE*

***Abstract*—With the advent of the Internet of Everything, the forthcoming 6G networks will necessitate an extensive deployment of base stations to achieve comprehensive network coverage. This development presents a formidable challenge for operators due to the surging expenses associated with each base station and the exponential increase in their quantity. Consequently, formulating strategies to mitigate network deployment costs has become an imperative challenge that operators must address urgently. The fully-decoupled radio access network (FD-RAN), through its innovative separation of uplink and downlink base stations, provides the flexibility for deployment in response to asymmetrical service demands, offering a promising solution to the deployment cost dilemma. In this study, we first propose a techno-economic cost model (TECM) for FD-RAN deployment based on the techno-economic approach and then formulate a cost-minimization problem for decoupled network deployment. To tackle the complexity of massive terminals, we present an adjacent gravity-based optimal link clustering algorithm (AGOC) to transform the significant 0–1 programming problem into a mixed integer program. Leveraging FD-RAN's uplink-downlink decoupling, we decompose the original problem into separate subproblems for uplink and downlink network deployments. For the decoupled problem, we further propose a branch-and-cut based network deployment (BCND) algorithm to solve the two decoupled deployment subproblems, respectively. Simulation results indicate that FD-RANs can offer significant cost benefits when facing differentiated service demands. Furthermore, the analysis reveals that power consumption and rental costs are the primary factors influencing network expenses.**

***Index Terms*—Branch and cut, clustering, FD-RAN, network deployment, techno-economic.**

I. Introduction

With the proliferation of mobile terminals, the significant surge in user traffic demand has driven the available spectrum in mobile communications to evolve towards higher frequencies, resulting in more intensive base station density [1], [2], [3], [4], [5]. Concurrently, the escalating adoption of emerging technologies, such as massive MIMO, metasurfaces, and artificial intelligence, has led to a persistent increase in the cost of individual base stations and a corresponding surge in network power consumption. [6], [7], [8], [9], [10]. As a result, substantial capital investment has emerged as a enormous obstacle for operators. The rapid proliferation of future large-scale terminals, estimated to reach trillions, will result in a significant increase in uplink traffic. Consequently, the service model of wireless networks will undergo substantial transformations. The widespread deployment of full-function base stations will result in a considerable waste of spectrum, energy, and economic resources. Hence, the formidable cost challenge and the utilization efficiency of network resources have arisen as substantial impediments that pose a threat to the sustainable development of operators.

To tackle the challenges of deployment cost and resource utilization efficiency in forthcoming networks, a green network architecture, known as the fully-decoupled radio access network (FD-RAN), has been put forward [3]. In the FD-RAN architecture, uplink base stations (UBSs) and downlink base stations (DBSs) are physically separated, resulting in changes in the reception mechanism and service mode. In conventional cellular networks, the requirement for users to use the same base station (BS) for both uplink and downlink services is no longer applicable in the context of FD-RAN. Thus, the deployment of network resources based on differentiated service demands emerges as the optimal approach to addressing the cost-related challenges in mobile communication. However, the effort to achieve cost-efficient deployment of FD-RAN networks, which requires accommodating differentiated and personalized service requirements in both uplink and downlink, remains an unexplored avenue.

Manuscript received 23 November 2023; revised 7 March 2024 and 19 May 2024; accepted 18 June 2024. Date of publication 11 July 2024; date of current version 7 November 2024. This work was supported in part by the National Natural Science Foundation Original Exploration Project of China under Grant 62250004, in part by the National Natural Science Foundation of China under Grant 62271244, in part by the Natural Science Fund for Distinguished Young Scholars of Jiangsu Province under Grant BK20220067, in part by High-level Innovation, in part by the Entrepreneurship Talent Introduction Program Team of Jiangsu Province under Grant JSSCTD202202, in part by the China National Postdoctoral Program for Innovative Talents under Grant BX20230163, and in part by China Postdoctoral Science Foundation under Grant 2023M731832. The review of this article was coordinated by Dr. Lin X. Cai. *(Corresponding author: Haibo Zhou.)*

Jiwei Zhao is with the College of Information Science and Electronic Engineering, Zhejiang University, Hangzhou 310058, China (e-mail: jackokie@zju.edu.cn).

Jiacheng Chen is with the Department of Strategic and Advanced Interdisciplinary Research, Peng Cheng Laboratory, Shenzhen 518000, China (e-mail: chenjch02@pcl.ac.cn).

Bo Cheng is with the School of Electronic Science and Engineering, Nanjing University, Nanjing 210023, China, and also with the China United Network Communications Corporation Guangzhou Branch, Guangzhou 510630, China (e-mail: bocheng@smail.nju.edu.cn).

Bo Qian is with the Information Systems Architecture Science Research Division, National Institute of Informatics, Tokyo 101-8430, Japan (e-mail: boqianmt@gmail.com).

Yunting Xu and Haibo Zhou are with the School of Electronic Science and Engineering, Nanjing University, Nanjing 210023, China (e-mail: yuntingxu@nju.edu.cn; haibozhou@nju.edu.cn).

Digital Object Identifier 10.1109/TVT.2024.3424547

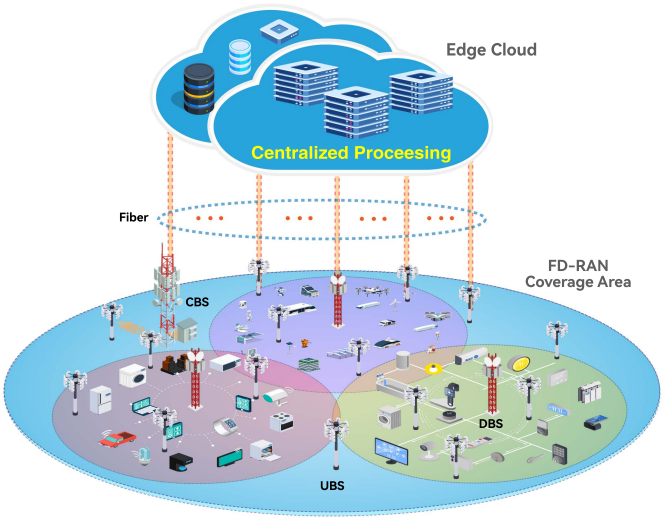

Fig. 1. Decoupled deployment of FD-RAN to satisfy the differentiated service requirements for large-scale IoT users.

Extensive research has been conducted on the deployment of wireless network base stations [11], [12], [13]. However, the FD-RAN network architecture introduces notable changes in the costs associated with different network components, especially UBSs and DBSs, which have distinct functional attributes. Therefore, the issue of devising a comprehensive and rational approach to modeling the cost of FD-RAN networks emerges as an unavoidable challenge. Furthermore, the distinctive separation of uplink and downlink services in FD-RAN introduces new opportunities and complexities for deploying fully decoupled base stations [14]. Modeling the cost of FD-RAN components and achieving minimal-cost network deployment while ensuring personalized service satisfaction is essential for operators to optimize network resource allocation, provide personalized services, and maintain revenue growth. Additionally, it serves as a crucial foundation for validating the evolution of future network architectures.

Optimizing network deployment can enhance the efficiency of infrastructure investments [15]. In the context of future networks, particularly differentiated service scenarios involving large-scale personalized terminals, we conduct a comprehensive analysis of the primary factors impacting network costs. Additionally, we propose a techno-economic cost model (TECM) that can be utilized for the effective deployment of FD-RAN. Furthermore, a cost-minimization deployment method is proposed for decoupled networks. The primary contributions of this paper are summarized as follows:

- We propose a novel techno-economic cost model (TECM) for cost-effective FD-RAN deployment, addressing large-scale terminals with personalized services. Additionally, we develop a cost-minimization-based decoupled deployment model specifically tailored for FD-RAN.
- To cope with the complexity of the deployment cost minimization problem, we propose the adjacent gravity-based optimal link clustering (AGOC) algorithm, which effectively converts the initial large-scale 0-1 programming problem into a mixed integer linear programming problem through channel clustering.
- In light of FD-RAN's decoupling characteristics, we introduce the branch-and-cut based network decoupled deployment (BCND) algorithm, which tackles the issues of differentiated service demand and network deployment cost by separately solving the uplink and downlink network deployment problems.

The rest of this paper is organized as follows. In Section II, we discuss related work on network deployment and techno-economics. In Section III, we propose the TECM cost model and formulate the cost-minimization network deployment problem. Then, in Section IV, we propose the AGOC clustering algorithm to reduce problem complexity. Additionally, in Section V, we introduce the B&C network deployment algorithm. Simulation settings and results are presented in Section V. Finally, Section VI concludes this work.

II. Related Works

Since the implementation of the first generation of mobile communication network systems, wireless network deployment, which aims to improve infrastructure investment efficiency, has been widely studied by researchers. In the following, we intend to elucidate research related to FD-RAN deployment from three aspects: network cost modeling, wireless network deployment, and techno-economic analysis.

A. Network Cost Modeling

Network cost assessment plays a crucial role in operators' strategic decision-making, network planning, and engineering implementation. Academic researchers have also shown a strong interest in network cost modeling [16], [17], [18], [19], [20], [21]. Literature [17] implemented a reliable backhaul design based on mesh networks and conducted a detailed capital expenditure (CAPEX) and operating expense (OPEX) economic evaluation. Xie et al. compared the deployment costs of three typical backhaul technologies, namely wireless backhaul, fiber, and passive optical network (PON), and the deployment costs of fixed wireless access and home fiber [16]. Literature [22] studied the problem of millimeter-wave base station deployment with limited connectivity in Manhattan-type geometric structures. The deployment modes of 5G standalone and non-standalone networks have been thoroughly analyzed regarding coverage range, network capability, network deployment complexity, and cost [23]. However, this work did not address the network service guarantee problem. Prior cost models have predominantly focused on fully integrated base stations, neglecting the modeling for deployments where uplink and downlink are decoupled, which is a central advantage for substantiating the feasibility of FD-RAN.

B. Wireless Network Deployment

Since the implementation of the first generation of mobile communication network systems, the wireless network deployment problem has been widely studied [11], [12], [13], [24], [25],

[26]. Liu et al. proposed a deployment strategy for multiple types of demands to solve deterministic and grid-based deployment problems [24]. Lin et al. studied the deployment of an intelligent computing system composed of a cloud core, gateways, fog devices, edge devices, and sensors attached to logistics center facilities [27]. Literature [12] proposed a low-complexity algorithm that minimizes the required number of small base stations while ensuring user satisfaction and meeting network capacity. Dong et al. proposed a method for millimeter-wave base station deployment problem based on a minimum deployment cost criterion and subject to a user equipment interruption constraint [13]. Additionally, minimizing network CAPEX under coverage constraints by jointly optimizing base station location and transmission power has also attracted attention [25]. Contemporary efforts in network deployment have largely ignored the critical perspective of network coverage from a service demand viewpoint. In contrast, our research investigates cost-effective deployment in response to differentiated uplink and downlink service demands.

C. Techno-Economic Analysis

Techno-economic analysis is a powerful technique that helps analyze pricing models and illustrates the relationship between cost and various parameters. In the field of communication, techno-economic analysis has been widely applied [20], [28], [29], [30], [31]. Yaghoubi et al. proposed a generic and comprehensive techno-economic framework for evaluating the total cost of ownership (TCO) and business feasibility of heterogeneous network deployment [20]. The study [30] quantified network costs by considering both CAPEX and OPEX, and established a technical and economic framework for broadband network deployment in rural and remote areas. Bouras et al. developed mathematical models for these two technologies. The experiments used prices from the Greek market, and they used sensitivity analysis to identify the most expensive cost parameters, which may deter providers from investing [29]. Literature [31] explored the techno-economic feasibility of cell-free networks with segmented fronthaul links based on bus, star, ring, and tree topologies. The results showed that the tree topology with a low serial level is the best topology for balancing network cost. Literature [32] provided a detailed overview of the applications of techno-economic analysis. Techno-economic methods have been widely applied to the evaluation of 5G networks, such as software-defined networking (SDN) technology [33], DAS techno-economic model [29], [34]. In addition, Oughton et al. introduced an open-source Python simulator for integrated modeling of 5G networks, evaluating engineering and cost metrics in a unified framework [28]. In [35], the authors studied the deployment efficiency of wireless networks; while in [36], [37], they examined the interrelationship between different parameters of wireless communication. Network deployment planning is one of the most critical steps to ensuring investment feasibility. Diverging from traditional methodologies, we employ techno-economic analysis to ascertain the influence of various sub-components and categories of network costs. This can yield strategic insights for actual network deployment and improve infrastructure investment efficiency.

III. System Model and Problem Formulation

As depicted in Fig. 1, we consider a large-scale Internet of Things (IoT) scenario, where numerous low-power IoT nodes are evenly dispersed throughout the region within a typical FD-RAN coverage area. The edge cloud serves as the control center for state acquisition, with all base stations connected to it through low-latency optical fibers. One of the critical components in this system is the control base station, which plays a crucial role in facilitating control plane signaling interactions among all users within the network's coverage area. In contrast, the data base station, which is physically divided into UBS and DBS, is responsible for delivering reliable services that meet the specific service requirements within the designated area. The uplink network employs a two-level merging technique to achieve terminal data collecting, aggregation, and forwarding [4]. The provision of personalized services for users is facilitated through the deployment of edge cloud infrastructure and the collaborative efforts among multiple base stations. The achievement of efficient and cost-effective coverage in the downlink network can be facilitated by leveraging edge cloud control to enable resource collaboration across numerous base stations [38].

In this paper, considering user mobility and the dynamics of service requests, we assume that users' service traffic is randomly distributed across the area. First, we use the subscript i to denote the downlink ($i = D$) and uplink ($i = U$) for distinct parameters and variables. In order to analyze the spatial distribution of traffic, the entire FD-RAN coverage area is partitioned into equilateral rectangles of size K, denoted as $\mathcal{K}$. Each individual sub-region k within $\mathcal{K}$ is then represented as a virtual service demand point (VSP) for service abstraction. It is assumed that the service demand of each VSP follows a normal distribution. Furthermore, we assume that the channel between the base station located in the coverage area and each VSP $k \in \mathcal{K}$ is obtained through extensive road tests. The symbol m is used to represent each road test point (RTP), and the collection of all these road test points constitutes the entire set of potential base station sites, denoted as $\mathcal{M}$. Hereafter, we employ the notation $x_m^i = 1$, where i belongs to the set $\{D, U\}$, to represent whether m deploys a UBS ($i = U$) or a DBS ($i = D$). Consequently, the challenge of deploying base stations is reformulated as a service guarantee issue for all VSPs $\mathcal{K}$ within the FD-RAN coverage area.

A. Wireless Transmission Model

We assume that both base stations and consumers operate at their respective maximum power levels, denoted as $P_{\max}^i$. To account for the distinct service requirements of uplink and downlink transmissions, we assume that users are serviced by independent DBSs in the downlink direction, while multiple UBSs collaborate to serve users in the uplink direction. Moreover, it is postulated that several UBSs employ maximum ratio combining to receive signals, and the resultant combined signal power at the edge cloud is determined by the sum of signal powers originating

from various base stations within a virtual service cluster (VSC). Denoting the received power at VSP k as p_k^i, we have

$$p_k^i = \sum_{m=1}^{M} a_{m,k}^i \cdot \rho_i \left(d_{m,k}, P_{\max}^i\right). \tag{1}$$

where $a_{m,k}^i$ is a binary variable indicating whether VSP k is served by the UBS/DBS located at m, and $\rho_i(\cdot)$ denotes channel fading, following the 3GPP channel model [39].

B. Coverage Model

Let N_k^i denote the interference and noise power of VSPs. Then, we can calculate the expected throughput for each VSP

$$R_k^i = B^i * \mathbb{E}\left[\log_2\left(1 + p_k^i / N_k^i\right)\right]. \tag{2}$$

where N_k^i denotes the expectation of interference and noise power. Since our primary concern is the VSP service guarantee, the sum of interference and noise power for each VSP, N_k^i, is assumed to be constant over time,

$$N_k^i = \mathbb{E}\left[\sum_{a \in \mathcal{M}-m} x_m^i \cdot \rho_i \left(d_{m,k}, P_{\max}^i\right)\right] + \mathrm{N}_0. \tag{3}$$

where N_0 denotes thermal noise power. To simplify system analysis, we assume that N_k^i is identical for all users, i.e., $N^i = N_k^i, \forall k \in \mathcal{K}$.

Considering personalized service for each user, we assume that VSP k is served only when the achievable rate surpasses a certain threshold, denoted as w_k^i,

$$B^i \cdot \log_2\left(1 + p_k^i / N^i\right) \geq w_k^i, \tag{4}$$

where $w_k^i \sim \mathcal{N}\left(\mu_i, \sigma_i^2\right)$. Since each base station has limited resources, we abstract the total resources of each base station as Π_i, leading to the following constraint:

$$\sum_{k=1}^{K} a_{m,k}^i \leq \Pi_i, \tag{5}$$

where Π_i denotes the maximum number of users each UBS/DBS can serve, considering spectrum or time resource utilization. For the uplink network, due to multiple base stations providing joint service, constraints on the UBS deployment decision x_m^U and the user uplink access decision $a_{m,k}^U$ are

$$\sum_{k=1}^{K_i} a_{m,k}^U \geq x_m^U, a_{m,k}^i \leq x_m. \tag{6}$$

For the downlink network, we assume that user downlink service is provided by a single DBS, then the base station deployment decision x_m^D satisfies the following constraints

$$\sum_{k=1}^{K_i} a_{m,k}^D = x_m^D. \tag{7}$$

C. Network Cost Model

Estimating the cost of mobile communication networks is an extraordinarily complex and challenging task that involves various expenses, including equipment, labor, and system operation. Based on the operator's field data, we propose a cost model for FD-RAN deployment, TECM [28], [34], which systematically divides the TCO across the life cycle into two primary categories: CAPEX and OPEX.

1) Capital Expenditure: CAPEX refers to the aggregate capital expenditure incurred in the process of constructing a network and providing wireless communication services to users. The primary components of the system are comprised of three elements: the access network C_{RAN}^i, the backhaul link C_{FH}^i, and the equipment installation C_{IT}^i

$$C_{CAPEX} = C_{RAN} + C_{FH} + C_{IT}, \tag{8}$$

The cost of access network C_{RAN} mainly consists of three parts: the antenna (AAU) C_{AAU}^i, the remote radio unit (RRU) $C_{R,RRU}^i$, and the baseband processing unit (BBU) $C_{R,BBU}^i$

$$C_{RAN} = \sum_{m=1}^{M} x_m^D \cdot C_{AAU}^D + \sum_{m=1}^{M} x_m^D \cdot C_{R,RRU}^D + \sum_{m=1}^{M} x_m^U \cdot C_{RU}^U + C_{R,BBU}. \tag{9}$$

The access network also includes the cost of fronthaul link for data backhaul, such as fiber fronthaul link $C_{B,FB}^i$ and access network router $C_{B,RT}$

$$C_{BH} = \sum_{i \in \{D,U\}} \sum_{m=1}^{M} x_m^i \cdot C_{B,FB}^i + C_{B,RT}. \tag{10}$$

In addition, considering the deployment of equipment, the access network includes expenses associated with labor, specifically for the installation of equipment such as base stations ($C_{I,TW}^i$) and fiber ($C_{I,FB}^i$), as well as their deployment and debugging processes:

$$C_{IT} = \sum_{i \in \{D,U\}} \left(\sum_{m=1}^{M} x_m^i \cdot C_{I,TW}^i + \sum_{m=1}^{M} x_m^i \cdot C_{I,FB}^i\right) \tag{11}$$

It should be noted that the builder of base station towers and the network operator may differ. For example, there are tower companies that specialize in tower construction and maintenance. Therefore, from the operator's perspective, we assume that the operator rents base stations from third-party companies. Our primary focus lies in the assessment of site rental expenses, rather than the costs associated with constructing base stations.

2) Operation and Expense: OPEX refers to the routine operating, management, and maintenance costs of the system. It chiefly incorporates three components: power consumption (C_{PW}), rental (C_{RT}), and operation and maintenance (C_{OM}).

$$C_{OPEX} = C_{PW} + C_{RT} + C_{OM} \tag{12}$$

The power expenditure of the system essentially encompasses the combined energy usage of the BBU, RRH/AAU, room equipment, air conditioning, etc. For simplification, the aggregate power consumption of various components within the central server room, including cooling systems and the centralized BBU resource pool (C-BBU), is represented as a consolidated power consumption variable $C_{P,Fix}^i$.

Since different network architectures use varied terminologies for base stations, their costs are uniformly denoted as $C^i_{P,RF}$. The power consumption associated with fiber fronthaul/backhaul for each base station is denoted as $C^i_{P,FB}$. Let T denote the cumulative duration of system operation until final withdrawal. Subsequently, the cumulative expenditure on power throughout the network's life cycle is

$$C_{PW} = T \cdot \sum_{i \in D,U} \left\{ C^i_{P,FIX} + \sum_{m=1}^{M} x^i_m \cdot C^i_{P,RF} + \sum_{m=1}^{M} x^i_m \cdot C^i_{P,FB} \right\}. \tag{13}$$

The rental costs chiefly entail two elements: infrastructure rental cost $C^i_{R,TW}$ required for base station deployment, and rental cost for the central server room $C_{R,RM}$

$$C_{RT} = T \cdot \left(\sum_{i \in D,U} C^i_{R,TW} + C_{R,RM} \right). \tag{14}$$

For micro base stations (MicBS) or UBSs, they are assumed to be directly installed on buildings, such as indoors and on apartment rooftops, due to their compact size and ease of deployment. In this case, the rental expense C_{RT} is expected to be significantly less than that of renting a tower.

The maintenance cost C_{OM} involves many factors, such as random repairs, system inspections, and routine maintenance, making it difficult to quantify accurately. Therefore, we adopt a holistic perspective on operating and maintenance costs, quantifying the average maintenance cost across the system.

D. Problem Formulation

In the following, we will formulate the network deployment cost optimization problem by incorporating the specified constraints and the TECM cost model. The primary objective is to minimize the TCO while simultaneously satisfying the personalized service requirements for both uplink and downlink. According to the expression of CAPEX (8) and OPEX (12), we can derive the expression of the objective function (15), shown at the bottom of this page. According to the derivation in the previous section, we can obtain the network deployment constraints (4), (5), (6), and (7). Based on the TECM expression, we can model the original base station deployment problem as a coverage problem based on minimum network cost

$$\min_{\mathcal{X},\mathcal{A}} \quad \phi(\mathcal{X},\mathcal{A}) \tag{16}$$

$$\text{s.t.} \quad B^i \cdot \log_2\left(1 + p^i_k / N^i\right) \geq w^i_k, \tag{16a}$$

$$a^i_{m,k} \leq x^i_m, \tag{16b}$$

$$\sum_{m=1}^{M} a^U_{m,k} \geq 1, \tag{16c}$$

$$\sum_{m=1}^{M} a^D_{m,k} = 1, \tag{16d}$$

$$\sum_{k=1}^{K} a^i_{m,k} \leq \Pi_i, \tag{16e}$$

$$a^i_{m,k} \in \{0,1\}, x^i_m \in \{0,1\}, \tag{16f}$$

$$\forall m \in \mathcal{M}^i, k \in \mathcal{K}^i, i \in \{D,U\}. \tag{16g}$$

The nonlinear constraint in (16a) makes the whole optimization problem a nonlinear programming problem. In order to address this issue, we transform variables in (16a) to the outside of the logarithm using an exponential operation, thus converting it into an equivalent linear constraint.

$$-\sum_{m=1}^{M} a^i_{m,k} \cdot P^i_{\max} \cdot g^i_{m,k}(d_{m,k}) + N^i \cdot 2^{w^i_k / B^i} \leq N^i. \tag{17}$$

Therefore, the nonlinear constraint in (16a) is transformed into a linear constraint in (17). At this point, we obtain a new optimization problem:

$$\min_{\mathcal{X},\mathcal{A}} \quad \phi(\mathcal{X},\mathcal{A})$$

$$\text{s.t.} \quad (17), (16b) \sim (16g). \tag{18}$$

The selection of base station deployment locations is crucial for guaranteeing network coverage and serves as a foundation for evaluating the performance of the FD-RAN. Given that m belongs to the set $\mathcal{M}$, which represents a division of the coverage area, and considering the large number of users $\mathcal{K}$ in Internet of Things scenarios, the binary decision variable $a^i_{m,k} \in R^{M \times K}$ can potentially reach tens of millions. This poses a significant challenge in directly solving the original problem (18). Considering the characteristics of base station deployment, users in adjacent areas are likely to share the same service base station. We divide the original problem into two

$$\begin{aligned} \phi(\mathcal{X},\mathcal{A}) = & \sum_{m=1}^{M} x^D_m \cdot \left(C^D_{AAU} + C^D_{R,RRU} + C^D_{B,FB} + C^D_{I,TW} + C^D_{I,FB} + T \cdot \left(C^D_{P,RF} + C^D_{P,FB}\right)\right) \\ & + \sum_{m=1}^{M} x^U_m \cdot \left(C^U_{RU} + C^U_{B,FB} + C^U_{I,TW} + C^U_{I,FB} + T \cdot \left(C^U_{P,RF} + C^U_{P,FB}\right)\right) \\ & + \sum_{i \in \{D,U\}} T \cdot \left(C^i_{RT,TW} + C^i_{OM}\right) + T \cdot C_{P,FIX} + C_{B,RT} + C_{R,BBU} \end{aligned} \tag{15}$$

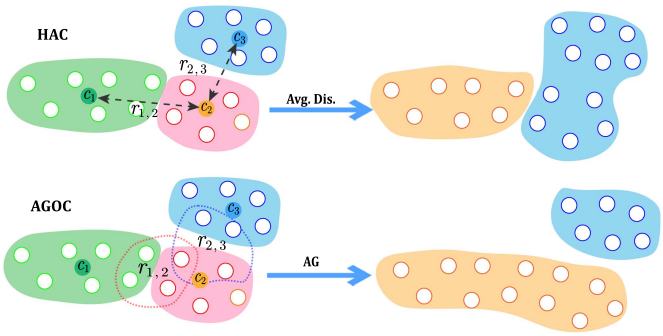

Fig. 2. Comparison of clustering outcomes between the average distance metric and the adjacent gravity metric.

subproblems: the first involves channel clustering to reduce the size of the problem, while the second focuses on solving the decision variable x_m^i to determine the deployment locations.

Given our dedication to ensuring reliable service for a substantial users $\mathcal{K}$, the decision variables $a_{m,k}^i \in \mathbb{R}^{M\times K}$ outlined in problem (18) could potentially exceed millions, making direct solution of the problem challenging. Considering the characteristics of wireless channels, it is common for users residing in adjacent areas to share similar base stations. In this paper, we decompose the original problem into two subproblems: one aims to mitigate problem complexity through channel clustering, while the other determines the deployment locations of base stations by resolving the decision variables x_m^i [40].

IV. Optimal Connectivity Clustering Based on Adjacent Gravity

To address the base station deployment subproblem, we propose the AGOC algorithm, which implements an agglomerative hierarchical clustering [41], [42].

A. Adjacent Gravity

The gravity is an excellent measure to quantify the clustering relationship, which considers both the connectivity and cohesion in determining the similarity between different sets of nodes. [43]. Given two centroids a and b, their gravity is defined as:

$$r_{a,b} = \frac{G_0 \mathrm{m}_a \mathrm{m}_b}{d_{a,b}^2}, \tag{19}$$

where G_0 denotes the universal gravitational constant, and m_a and m_b represent the masses of the two centroids respectively.

Let us consider two clusters, labeled as C_i and C_j. Within these clusters, let a be an element belonging to C_i and b be an element belonging to C_j. The distance set between the two clusters is denoted as $\mathcal{R}_{i,j}$, which is defined as $\mathcal{R}_{i,j} = \{r_{a,b}, a \in C_i, b \in C_j\}$.

As shown in Fig. 2, in the hierarchical clustering algorithm based on average linkage (HAC), the distance between different clusters $s_{i,j}$ is calculated based on the cluster centers C_i. In this case, the inter-cluster distance is defined as $\frac{1}{|C_i|\cdot|C_j|}\sum_{a\in C_i, b\in C_j} r_{a,b}$. Consequently, the clustering result is easily affected by data points far away from each other within corresponding cluster. To mitigate this issue, we propose using adjacent gravity (AG), which measures the distance by utilizing partial closest points between two clusters.

Between two clusters, the data points that are closer to each other tend to have a greater influence, especially in non-Euclidean datasets such as circular or linear datasets. The gravity values $\mathcal{R}_{i,j}$ between any two points in two clusters are sorted in descending order to obtain a permutation of the inter-cluster gravity, $\hat{\mathcal{R}}i,j$. We define a gravity coefficient ϵ, which represents the proportion of samples in $\mathcal{R}_{i,j}$ used to calculate the inter-cluster gravity, i.e., $\lceil \epsilon \cdot |\hat{\mathcal{R}}_{i,j}| \rceil$. Next, we define the AG set used for clustering as the first $\lceil \epsilon \cdot |\hat{\mathcal{R}}_{i,j}| \rceil$ elements of $\hat{\mathcal{R}}_{i,j}$

$$\widetilde{\mathcal{R}}_{i,j} = \left\{ r_{i,j}^{(k)} \mid k = 1, 2, \ldots \lceil \epsilon \cdot |\hat{\mathcal{R}}_{i,j}| \rceil \right\}, \tag{20}$$

where $r_{i,j}^{(k)}$ denotes the k-th element in the sequence. At this stage, we can define the adjacent gravity between C_i and C_j

$$s_{i,j} = \mathbb{E}\left(\widetilde{\mathcal{R}}_{i,j}\right). \tag{21}$$

As shown in Fig. 2, the samples used to determine the AG are represented by the dashed line at the bottom. Additionally, it can be observed that the clustering results based on AG are more rational in this scenario.

B. Optimal Connectivity Clustering

In social relationships, the nearest neighbor is a sufficient statistic to measure clustering relationships [44]. Based on this assumption, we propose a clustering algorithm based on optimal connectivity. According to the definition of AG $s_{i,j}$ in (21), we can build the AG matrix $\mathcal{S}$. Letting $l_{i,j}$ indicate whether cluster i and cluster j are optimally connected, we define the optimal connection of cluster i as χ_i

$$\chi_i \triangleq \arg\max_j s_i, \tag{22}$$

where $s_i = [s_{i,1}, \ldots, s_{i,|\mathcal{S}_i|}]$ represents the i-th row of $\mathcal{S}$, and j is the index of the j-th element of vector s_i. At this point, we have $l_{i,\chi_i} = 1$. Based on $\mathcal{S}$, we can construct the optimal connection matrix (OCM), denoted as $\mathcal{L} \in \mathbb{R}^{|\mathcal{S}|\times|\mathcal{S}|}$

$$\mathcal{L} = \{l_{i,j} \mid i, j \in \{0, \ldots |\mathcal{S}| - 1\}\}. \tag{23}$$

Literature [44] shows the chain effect of optimal connections in clustering. Clusters C_i and C_j tend to form a new cluster if they have an optimal connection, namely, $l_{i,j} = 1$ or $l_{j,i} = 1$. Based on this principle, we can obtain the optimal connection extended matrix (OCE), denoted as $\widetilde{\mathcal{L}} \triangleq \{\widetilde{l}_{i,j}\}$, where the element $\widetilde{l}_{i,j}$ is defined as

$$\widetilde{l}(i,j) = \begin{cases} 1 & \text{if } j = \chi_i^1 \text{ or } \chi_j = i \text{ or } \chi_i = \chi_j \\ 0 & \text{otherwise.} \end{cases} \tag{24}$$

By associating each point i with its first neighbor via χ_i, we derive the optimal OCE matrix $\mathcal{L}$. Furthermore, the COE symmetry is preserved by $\widetilde{\mathcal{L}}$ through the assignment of $\chi_j = i$ and $\chi_i = j$. Additionally, $\widetilde{\mathcal{L}}$ establishes connections between points (i, j) that have a shared neighbor by assigning $\chi_i = \chi_j$. Hence, $\widetilde{\mathcal{L}}$ will

be a symmetric sparse matrix defining a graph whose connected components converge to a cluster. The criterion in (24) integrates 1-nearest neighbor clustering and shared nearest neighbor graph. Employing only the integer index of the first neighbor, namely OCE, (24) specifies and detects the relationships among the samples, further yielding the partition of clusters.

C. Handling of Outliers

We first define the outlier coefficient γ, which is primarily used for the construction of the AG matrix $\mathcal{S}$ and the masking matrix $\widetilde{\mathcal{L}}$. After constructing $\mathcal{S}$, we compute the lower triangular matrix $\mathcal{S}_L$ of $\mathcal{S}$ and set the diagonal elements to 0. At this point, we sort $\mathcal{S}_L$ in ascending order and obtain a new sequence $\hat{\mathcal{S}}_L$. Subsequently, we determine the threshold for identifying outliers by employing the outlier coefficient.

$$\tau_\gamma = \hat{\mathcal{S}}_L[\lfloor \gamma * |\hat{\mathcal{S}}_L| \rfloor], \tag{25}$$

where $\lfloor \cdot \rfloor$ denotes the floor function. Based on τ_γ, we obtain the mask matrix $\gamma_M = [\mathcal{S} \geq \tau_\gamma]$. Then, we use γ_M to mask the OCE matrix

$$\mathcal{L}_M = \gamma_M \cdot \widetilde{\mathcal{L}}. \tag{26}$$

At this point, we can ensure that the links below the outlier threshold cannot achieve clustering in the current iteration, eliminating outliers' influence. Furthermore, let us consider the scenario where we collect the outlier count N_i^γ for each cluster C_i throughout each iteration of the clustering process. This count represents the total number of times that the cluster is identified as an outlier in the current iteration. Then we have the outlier matrix

$$N^\gamma = \text{diag}\left(N_1^\gamma, \ldots N_i^\gamma, \ldots, N_{|S|}^\gamma\right) \tag{27}$$

After computing the OCM matrix $\mathcal{S}$, we continue to use N^γ to process $\mathcal{S}$ within current iteration

$$\mathcal{S} = N^\gamma \cdot \mathcal{S}. \tag{28}$$

The procedure, multiplying each row vector by the outlier count, increases the clustering priority for those identified as outliers more frequently, preventing them from joining the clustering at the final stage. Additionally, when multiple clusters combine to form a new cluster, the count of outliers is reset to 1.

D. AGOC Algorithm

Balcan et al. have undertaken a comprehensive investigation into the issue of identifying a benchmark clustering [45]. Successive clusters are obtained through multiple iterations, allowing for the capture of the underlying data structure at various levels of granularity. In Algorithm 1, we first initialize the cluster $\mathcal{C}$, the outlier count N^γ, and the cluster partition Γ. In the clustering loop, we first compute the AG $\mathcal{S}$ between different samples and then weight $\mathcal{S}$ using the outlier statistics N^γ. Next, we compute the optimal connection matrix $\mathcal{L}$ and mask it according to (25) and (26), excluding outliers that cannot be clustered. We then perform chain clustering, obtaining the clustering results $\mathcal{C}$ and Γ for the current iteration. Moreover, we also introduce

Algorithm 1: Optimal Connection Clustering based on Adjacency Gravity (AGOC).

Input: Feature matrix H, gravity coefficient ϵ, outlier coefficient γ, desired number of clusters $N_\mathcal{E}$, maximum number of clusters $N_{\max}$, minimum number of clusters $N_{\min}$

Output: Clustering result $\mathcal{C}$ and Γ

```
Initialize clustering C = 1^K, outlier count N^γ = 1^K
Cluster partition Γ = 1^K, and clustering tree Π_C = {C} and Π_Γ = {Γ}
while true do
    Calculate S according to Equations (19), (20), and (21)
    Perform outlier weighting according to Equation (28)
    Calculate the OCM L according to Equation (23), and perform masking according to Equation (26)
    Perform chain clustering according to Equation (24), update C and Γ
    Calculate current number of clusters N_C = |Γ|, the number of clusters in the previous iteration N_C^- = |Π_Γ[−1]|
    if N_C < N_C^- then
        Π_C = {Π_C, C}, Π_Γ = {Π_Γ, Γ}
        if (N_C < N_max) | (N_C < N_E) | (N_C ≤ 2) then
            break
        end if
    else
        if N_C < 2 then
            break
        else
            return error
        end if
    end if
    if (N_C > N_min) | (N_C = N_E) then
        N_C = max(N_max, N_E)
        C, Γ = SORT_CONVERGE(Π_Γ[−2], N_C)
        Π_C = {Π_C, C}, Π_Γ = {Π_Γ, Γ}
    end if
end while
```

several input parameters for the AGOC algorithm, such as the desired number of clusters $N_\mathcal{E}$, the maximum number of clusters $N_{\max}$, and the minimum number of clusters $N_{\min}$. To meet the user's requirements for the number of clusters, we propose Algorithm 2 to achieve the desired number of clusters. In addition, if the algorithm is provided with an upper and lower bound for the number of clusters, it will determine the termination condition based on the number of clusters in the iterative process and execute Algorithm 2 to obtain a clustering result within the interval $[N_{\min}, N_{\max}]$.

In Algorithm 2, given the current cluster partition Γ and the target number of clusters $\mathcal{N}_\mathcal{E}$, we achieve the clustering goal by $|\Gamma| - \mathcal{N}_\mathcal{E}$ iterations. In step 1, we first compute the AG between each category and obtain $\mathcal{S}$. Then, in each iteration, we select two clusters corresponding to the maximum AG value and merge them to obtain $\hat{\gamma}_i$, i.e., update the cluster partition

Algorithm 2: SORT_CONVERGE.

Input: Cluster partition Γ, target number of clusters $\mathcal{N}_{\mathcal{E}}$
Output: $\widetilde{\Gamma}, \mathcal{C}$

1: For current cluster partition Γ, calculate $\mathcal{S}$ according to Equations (19), (20), and (21)
2: **for** $k = N_{\mathcal{E}}, \ldots, |\Gamma|$ **do**
3: According to $\mathcal{S}$, calculate the indices i and j of the two closest clusters in Γ, $i < j$
4: Merge the i-th cluster γ_i and the j-th cluster γ_j in Γ into a new cluster $\hat{\gamma}_i$
5: Replace γ_i with $\hat{\gamma}_i$, delete the j-th cluster γ_j in Γ, and form new Γ
6: Construct the latest clustering $\mathcal{C}$ according to Γ
7: Perform incremental update of $\mathcal{S}$ according to $\mathcal{S}$ and the latest cluster $\hat{\gamma}_i$
8: **end for**

Γ and the cluster list $\mathcal{C}$. In the iteration process, updating $\mathcal{S}$ is incremental, i.e., we only need to compute the AG between the new merged cluster $\hat{\gamma}_i$ and other clusters to construct a new $\mathcal{S}$ without computing AG between any two clusters, which meets computational efficiency requirements.

The fundamental procedure of the AGOC algorithm is straightforward, involving the initial partitioning of samples followed by the recursive merging of these clusters. This process necessitates the identification of the first neighbor for each cluster. To identify these neighbors, it is essential to compute the inter-cluster distances. The complexity of the proposed AG computation is $O(N_a^2)$, where N_a represents the number of samples used to compute AG for each cluster. This will lead to high complexity. Given that $N_a < \lceil \epsilon \cdot N \rceil$, we can control the computational complexity within certain limits. Subsequently, the nearest neighbor is identified using these distance metrics, approaching a complexity of $O(N \log(N))$ due to the merging of clusters in each recursion.

V. Branch&cut Based Network Deployment

Following the AGOC clustering described above, we obtain a series of user clusters Γ through channel clustering. Next, we will explore how to use Γ to solve the original problem.

A. Problem Transformation

By partitioning users with Γ, we obtain a set of new virtual user nodes (Virtual User Cluster, VUC) $\{C_1, C_2, \ldots, C_{|\mathcal{C}|}\}$, where each virtual node C_n comprises multiple users $\mathcal{C}_n = \{k_1, k_2, \ldots, k_{|\mathcal{C}_n|}\}$. The received power for each user in C_n from base station m is

$$\phi^i_{m,n} \triangleq \left[\rho^i\left(d_{m,k_1}, P^i_{\max}\right), \ldots, \rho^i\left(d_{m,k_{|\mathcal{C}_n|}}, P^i_{\max}\right)\right]^{\mathrm{T}} \quad (29)$$

Assuming the personalized service requirements for each user in C_n are $\hat{T}^i_n \triangleq [w^i_{k_1}, w^i_{k_2}, \ldots, w^i_{k_{|\mathcal{C}_n|}}]$, $\alpha_{m,n}$ represents the decision variable of base station m serving VUC n. The condition for C_n to be served is

$$\log_2\left(1 + \sum_{m=1}^{M} \frac{\alpha^i_{m,n} \cdot \phi^i_{m,n}}{N^i}\right) \geq \hat{T}^i_n \quad (30)$$

where the operation applies the constraint to each element in $\phi m, n^i$, meaning the decision $a_{m,n}$ must satisfy the rate constraints for each element within $\mathcal{C}_n$. This precisely explains the purpose of clustering, namely, transforming independent decisions into collective decisions, thereby reducing the dimensionality of the decision space. Then, the original constraint for each VSP is tightened into a collective constraint. Constraint (30) can be transformed into the linear constraint:

$$\sum_{m=1}^{M} \alpha^i_{m,n} \cdot \phi^i_{m,n} \geq N^i \cdot \left(2^{\hat{T}^i_n} - 1\right) \quad (31)$$

Equation (31) implies that the received power for each user exceeds the threshold required by its personalized service demand. A VUC node n is considered to be served only if all the services offered by the base stations satisfy its virtual demand u^i_n. Let $z_{m,n}$ denote the time-slot resource that base station m provides to VUC node n, then we have the constraint:

$$\sum_{m=1}^{M} z^i_{m,n} \cdot \alpha^i_{m,n} \geq u^i_n. \quad (32)$$

However, constraint (32) renders the problem non-linear, significantly increasing its complexity. Since $\alpha^i_{m,n}$ can only take values of 0 or 1, when $\alpha^i_{m,n} = 0$, it effectively sets $z^i_{m,n}$ to 0. This means it only considers the resources assigned by the connected base stations. To enforce the condition that $z^i_{m,n}$ takes a value of $z^i_{m,n} = 0$ when $\alpha^i_{m,n} = 0$, we can impose the following constraint:

$$z^i_{m,n} \leq \alpha^i_{m,n} \cdot u^i_n \quad (33)$$

Thus, constraint (32) can be reformulated as

$$\sum_{m=1}^{M} z^i_{m,n} \geq u^i_n. \quad (34)$$

However, since each base station assigns different time-frequency resources to users, the capacity of each base station will also be limited. According to constraint (16e), we have the following base station capacity constraint:

$$\sum_{n=1}^{N_C} z^i_{m,n} \leq \Pi_i. \quad (35)$$

These constraints ensure that the aggregate service provided by all base stations is sufficient to meet the demand of all VUCs. At this point, we can recast problem (18) as (36):

$$\min_{\mathcal{X},\mathcal{Z}} \quad \phi\left(\mathcal{X}, \alpha, \mathcal{Z}\right) \quad (36a)$$

$$\text{s.t.} \quad \sum_{m=1}^{M} \alpha^i_{m,n} \cdot \phi^i_{m,n} \geq N^i \cdot \left(2^{\hat{T}^i_n} - 1\right), \quad (36b)$$

$$\sum_{m=1}^{M} z_{m,n}^{i} \geq u_{n}^{i}, \tag{36c}$$

$$\sum_{n=1}^{N_C} z_{m,n}^{i} \leq M_{K}^{i}, \tag{36d}$$

$$z_{m,n}^{i} \leq u_{n}^{i} \cdot \alpha_{m,n}^{i}, \tag{36e}$$

$$z_{m,n}^{i} \leq x_{m}, \tag{36f}$$

$$x_{m}^{i} \in \{0,1\}, z_{m,n}^{i} \in \left\{1,\ldots,M_{K}^{i}\right\}, \tag{36g}$$

$$\forall a \in \{1,\ldots,M\}, n \in \left\{\mathcal{C}^{i}\right\}, i \in \{D,U\}. \tag{36h}$$

In this paper, the objective function of problem (36) is related only to $\mathcal{X}$ and not to $\mathcal{A}$ and $\mathcal{Z}$. Then, $\phi(\mathcal{X},\alpha,\mathcal{Z})$ is equivalent to $\phi(\mathcal{X},\mathcal{A})$. Each user's uplink and downlink services must be provided by the same base station in the existing network, while uplink and downlink services are provided through two independent networks in FD-RAN. Currently, for problem (36), the uplink and downlink constraints are independent, and the uplink and downlink variables are also independent in the objective function. Hence, problem (36) can be decomposed into two independent subproblems: one for downlink network deployment and another for uplink network deployment. The subproblem for UBS deployment can be expressed as follows

$$\begin{aligned} \min_{\mathcal{X}^{U}} \quad & \sum_{m=1}^{M} x_{m}^{U} \\ \text{s.t.} \quad & \text{(36b)} \sim \text{(36f)}, \\ & x_{m}^{D} \in \{0,1\}, z_{m,n}^{D} \in \left\{1,\ldots,M_{K}^{i}\right\}, \\ & i = D, \forall a \in \{1,\ldots,M\}, n \in \left\{\mathcal{C}^{D}\right\}, \end{aligned} \tag{37}$$

while the subproblem for DBS deployment is

$$\min_{\mathcal{X}^{U}} \quad \sum_{m=1}^{M} x_{m}^{D} \tag{38a}$$

$$\text{s.t.} \quad \text{(36b)} \sim \text{(36f)}, \tag{38b}$$

$$x_{m}^{U} \in \{0,1\}, z_{m,n}^{U} \in \left\{1,\ldots,M_{K}^{i}\right\}, \tag{38c}$$

$$i = U, \forall a \in \{1,\ldots,M\}, n \in \left\{\mathcal{C}^{U}\right\}. \tag{38d}$$

The objectives and constraints of subproblems (37) and (38) are linear, making them both mixed integer linear programming (MILP) problems. Next, we will describe how to solve this problem.

B. Network Deployment With Branch&cut Algorithm

The branch and Cut (B&C) concept consists of two optimization methods: branch&bound and cutting plane. With these two methods, the B&C algorithm generates upper bounds by relaxing the problem and incorporating these bounds as constraints into the original problem, then iteratively finding the optimal solution. Relaxing problems can simplify complex problems so they can be solved more easily.

1) Cutting Plane: A cutting plane is an additional linear constraint that contradicts the linear programming (LP) solution after relaxation but does not eliminate any feasible integer solutions. Specifically, the cutting plane method (sometimes called valid inequality) repeatedly adds constraints to the LP problem, reducing the feasible region and tightening the LP relaxation, ensuring the current relaxed solution remains within the feasible space. In contrast, the optimal integer solution remains within the feasible space after the cutting planes are applied. The difference between branching and bounding to tighten LP relaxation is illustrated in Fig. 3.

The black points in Fig. 3 represent the feasible integer solutions of the MILP problem, while the area enclosed by line segments represents the feasible region of the relaxed mixed-integer programming problem. The reduced feasible region is depicted as the blue area in the diagram, with the green point representing the optimal solution of the updated LP problem. During the iterative procedure, the feasible region is partitioned by successively adding cutting planes until the optimal solution of the LP problem is an integer. At this point, the optimal solution of the original LP problem is achieved. Solving MILP by relying solely on the cutting plane method is not feasible. Consequently, it is customary to be consistently integrated with B&B. The following section will provide a concise overview of the B&C algorithm for addressing the base station deployment problem.

2) Branch&cut Based Network Deployment Algorithm: Algorithm 3 delineates the network deployment algorithm BCND using B&C adopted in this paper. First, in Step 3, a subproblem $\mathcal{P}$ is selected from the queue. In Step 4, the LP relaxation of $\mathcal{P}$, excluding integer constraints, is solved using the simplex method. If the optimal solution $\bar{x}(\mathcal{P})$ contains non-integer values, cutting plane algorithms can be utilized to identify additional linear constraints that satisfy all feasible integer points while violating the current LP relaxation. In Step 16, these inequalities are added to the linear program for re-solving, yielding a different solution. In the branch and bound part of Step 18, the problem $\mathcal{P}$ is divided into two subproblems, $\mathcal{P}_1$ and $\mathcal{P}_2$, which are subsequently added to the queue $\mathcal{Q}$. Within the branch and bound process, non-integer LP relaxation solutions act as upper bounds, whereas integer solutions serve as lower bounds. Nodes may be pruned if the upper bound is less than the current lower bound. Finally, in Step 20, the algorithm returns the optimal x^* and minimum cost y^*. The algorithm begins by solving the LP relaxation of a single subproblem derived from the original problem. During the execution, new subproblems are generated through branching, with solutions to these subproblems potentially producing two or more child subproblems. This process can be visualized as a tree structure, where nodes represent the associated subproblems, and edges denote the parent-child relationships between them.

VI. Simulation Results

In this paper, we compare the cost between FD-RAN and other reference network architectures: massive MIMO networks (mMIMO), small cell networks (Small Cell), and cell-free networks (Cell-free) [4]. We assume that both Small Cell

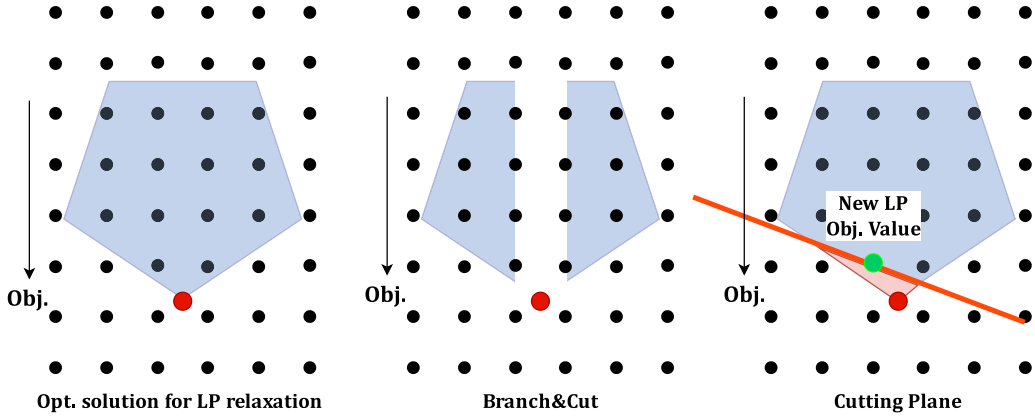

Fig. 3. The relationship among LP optimal solution, cutting plane, and branch&cut algorithm. The red point represents the optimal solution of the LP-relaxed problem. The red line indicates the cutting plane added to the original problem, which eliminates the fractional optimal solution of the LP and part of the region that does not contain any feasible integer solutions.

Algorithm 3: Branch&Cut Based Network Deployment Algorithm.

Input: Original problem $\mathcal{P}_0$
Output: Optimal solution $x^* \in X_{\text{MILP}}$, objective function value y^*, or $X_{\text{MILP}} = \emptyset$, $y^* = +\infty$
1: Define a queue of subproblems to be solved $Q \triangleq \{\mathcal{P}_0\}$, $y^* := +\infty$, $x^* = \text{null}$
2: **while** $Q \neq \emptyset$ **do**
3: Select a subproblem $\mathcal{P}$ from the queue Q, and delete it from the queue $Q = Q \backslash \{\mathcal{P}\}$
4: Solve linear programming $y(\mathcal{P}) = \min\{c^T x : x \in \mathcal{P}\}$, and obtain the optimal solution $\bar{x}(\mathcal{P})$
5: **if** $\bar{x}(\mathcal{P})$ is not a feasible solution **then**
6: Go back to step 2
7: **end if**
8: **if** $y(\mathcal{P}) \geq y^*$ **then**
9: Go back to step 2
10: **end if**
11: **if** $\bar{x}(\mathcal{P})$ is an integer **then**
12: $y* = y(\mathcal{P})$, $x^* = \bar{x}(\mathcal{P})$
13: **end if**
14: **Cut**: Use a heuristic algorithm to obtain a set of cutting planes, $Dx \geq e$, where $D \in \mathbb{R}^{k \times n}$, $e \in \mathbb{R}^k$.
15: **if** if there are cutting planes **then**
16: Add corresponding constraints to the problem $\mathcal{P} := \mathcal{P} \cup \{Dx \geq e\}$, go back to step 4
17: **end if**
18: **Branch**: Divide $\mathcal{P}$ into two subproblems $\mathcal{P}_1$ and $\mathcal{P}_2$, and add them to the queue $\mathcal{Q}$, go back to step 3
19: **end while**
20: **return** Optimal value x^* and y^*

and Cell-free networks consist of fully functional micro base stations, with the distinction that Cell-free networks utilize cooperative transmission and reception across multiple base stations. In addition, we assume that mMIMO networks are deployed with macro base stations. However, in FD-RAN, we assume that the downlink deployment comprises only macro base stations with downlink functionality, while the uplink deployment consists of solely micro base stations with uplink functionality.

Table I shows all cost parameters for mMIMO, Small Cell, FD-RAN, and Cell-free. All equipment costs are based on actual data from domestic operators and public sources. However, since FD-RAN has not been deployed, we estimate costs based on simplified functionality and reduced expenses. APs and UBSs bear similarities to WiFi nodes and are capable of being installed within interior environments. Consequently, the expenses associated with renting APs and UBSs are considerably lower compared to traditional towers. Furthermore, in the absence of established benchmarks for network operation and maintenance expenses, we have used a conservative estimate of 1000 per month as a representative value for all networks over a 20-year operational period. Additionally, as prices vary significantly by region and time, it is reasonable to suppose that they exhibit a range of around 50%. Intuitively, since the device cost and power consumption of UBS are much lower than other base stations, this is the source of cost advantage, especially in scenarios where uplink service dominates.

We assume a user transmission power of 23 dBm. Additionally, we assume that the transmission power for mMIMO macro base stations (MacBS) and DBS is set to 46 dBm, while the power for Small Cell micro base stations (MicBS) and Cell-free access points (AP) is 39 dBm. In order to examine the correlation between network cost and the number of VSPs, we make the assumption that the geographical region encompasses a range of 1000 to 15,000 VSPs. We establish a baseline of 10,000 VSPs for comparative analysis of other performance indicators. It is assumed that the guaranteed rate for users conforms to a normal distribution with a mean of 1 Mb/s and a variance of 0.1, $w_k^i \sim \mathcal{N}(1, 0.1)$. Furthermore, it should be noted that the simulation assumes a default uplink-downlink traffic ratio (U/D) of 5. The 3GPP UMa channel model is employed as the fundamental simulation scenario for analytical purposes [39].

TABLE I
COST PARAMETERS FOR DIFFERENT NETWORK ARCHITECTURES ($)

Cost Components		FD-RAN	mMIMO	Small Cell	Cell-free
RAN	AAU	•	13596	5313	5313
	BBU	•	8813	5051	
	DBS/UBS	5313/1300	•	•	•
	C-BBU	10000	•	•	10000
Fronthaul	Fiber	1000	1000	1000	1000
	Router	2000	2000	2000	2000
Install	DBS/MacBS	1000	1000	•	•
	UBS/MicBS	200	•	200	200
	Fiber	500	500	500	500
Power	CN Room	2.0/h	2.0/h	2.0/h	2.0/h
	Fronthaul Fiber	0.1/h	0.1/h	0.1/h	0.1/h
	DBS/Mac-BS	0.6/h	1.0/h	•	•
	UBS/AP	0.1/h	•	0.4/h	0.4/h
Rental	CN Room	400/mon.	400/mon.	400/mon.	400/mon.
	Tower	200/mon.	200/mon.	•	•
	Building	100/mon.	•	100/mon.	100/mon.
O&M	all devices	1000/mon.	1000/mon.	1000/mon.	1000/mon.

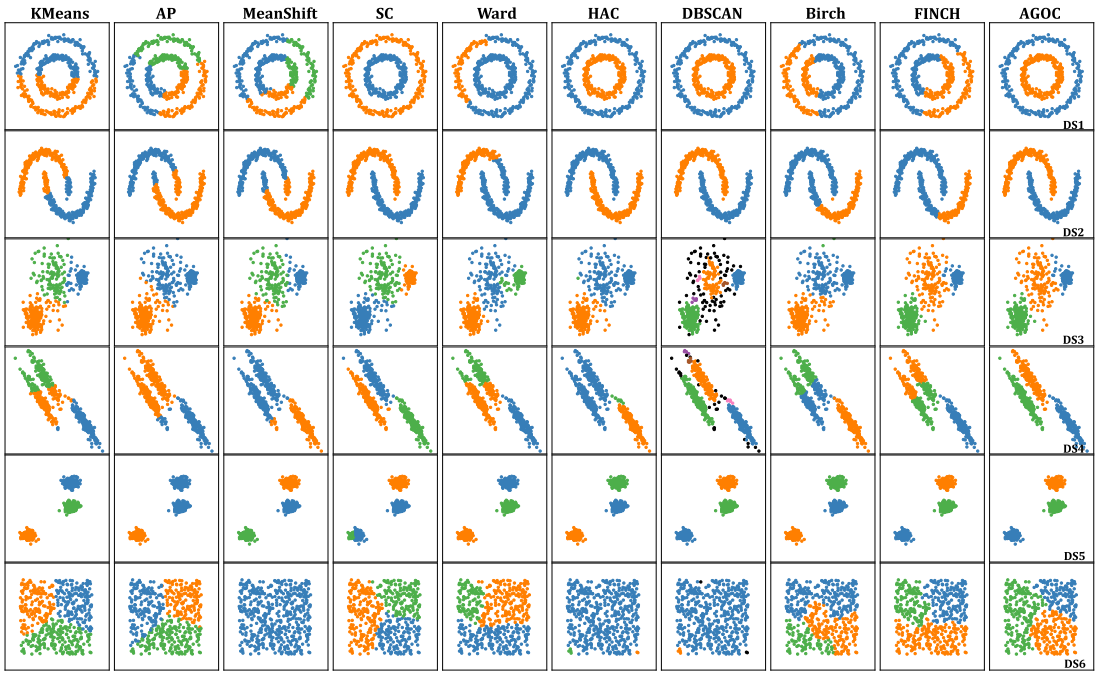

Fig. 4. Comparison of clustering results between AGOC and other benchmark algorithms.

A. Performance Analysis of Different Clustering Algorithms

Since channel clustering quality affects VSP partitioning and subsequent personalized service guarantees, we first compare the AGOC algorithm with other representative algorithms:

- Automatically estimate the number of clusters. Given specific input hyperparameters or threshold settings, this category includes Affinity Propagation (AP), Density-Based Spatial Clustering of Applications with Noise (DBSCAN), etc. [41].
- Require the specification of the number of clusters. This category includes KMeans, Birch, Spectral Clustering (SC), Hierarchical Agglomerative Clustering (HAC), etc. [41].
- Cluster without input parameters. First Integer Neighbor indices can produce a Clustering Hierarchy (FINCH), etc. [44].

To clearly demonstrate the characteristics of different algorithms, in Fig. 4, we employ six classic datasets (DS1∼DS6) to reveal the clustering quality of the AGOC algorithm [46].

TABLE II
PERFORMANCE COMPARISON OF VARIOUS CLUSTERING ALGORITHMS

Metric	HAC	KMeans	KMedoids	FINCH	Birch	AGOC
RI	0.2513	0.5495	0.5929	0.4136	0.7337	**0.7350**
MI	0.6020	0.7563	0.7639	0.7100	0.8392	**0.8663**
VM	0.6445	0.7915	0.7984	0.7525	0.8624	**0.8853**
FM	0.4263	0.5951	0.6289	0.5127	0.7519	**0.7590**
CHI	31.8292	68.7088	100.1830	68.0111	127.7275	**163.9629**

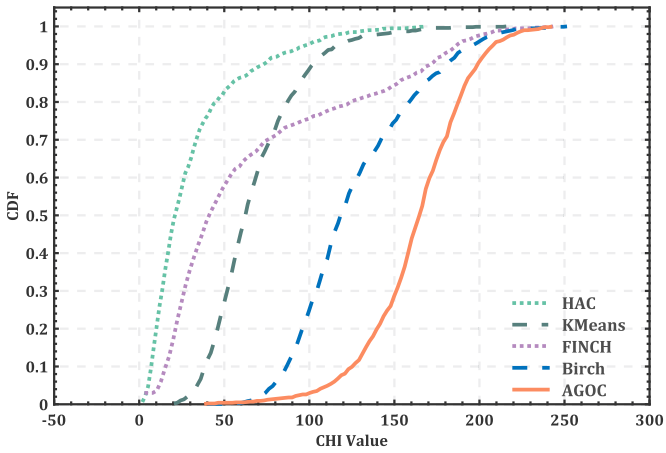

Fig. 5. Comparison of the CHI index between AGOC and other benchmark algorithms.

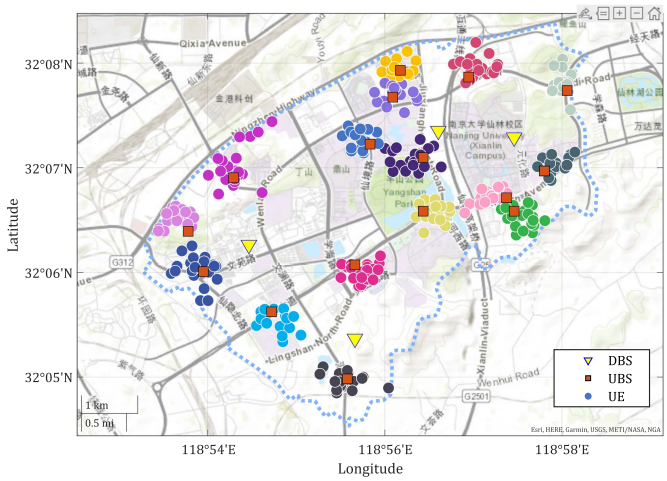

Fig. 6. Display of AGOC algorithm clustering results under the FD-RAN architecture, showing decoupled deployment results for uplink and downlink networks.

Firstly, it is observed that the AGOC algorithm yields good clustering results across almost all datasets. For DS1 (annular data), DS2 (curved data), and DS4 (linear data), algorithms such as KMeans struggle to distinguish these data sets completely. Even the FINCH algorithm, which is similar to AGOC, fails to completely separate them. Only AGOC demonstrates the ability to accurately differentiate them. The primary reason is that AGOC uses AG to characterize similarity between clusters, with a focus on the division of clustering boundaries. Furthermore, for the isolated points in DS4, it is noteworthy that none of the other reference methods can accurately cluster them, except AGOC. The main reason lies in the proposed isolated point processing mechanism of AGOC. At the initial stage of clustering, isolated points are avoided, preventing erroneous clustering. This successfully reduces the impact of isolated points on the overall clustering process, thereby avoiding undesirable outcomes in both KMeans and Ward clustering algorithms, as shown in Fig. 4.

To assess the efficacy of AGOC in addressing channel clustering issues, a dataset including 1000 users was generated. These individuals were then categorized into 20 distinct clusters. Various algorithms were employed for channel clustering and subsequent performance assessment. The performance metrics include the following: Rand Index (RI), Mutual Information (MI), V-Measure (VM), Fowlkes-Mallows (FM) score, Calinski-Harabasz Index (CHI), and Silhouette Coefficient Index (SCI) [47]. Fig. 5 illustrates the CDF of the CHI index for clustering. The AGOC algorithm has a significantly higher

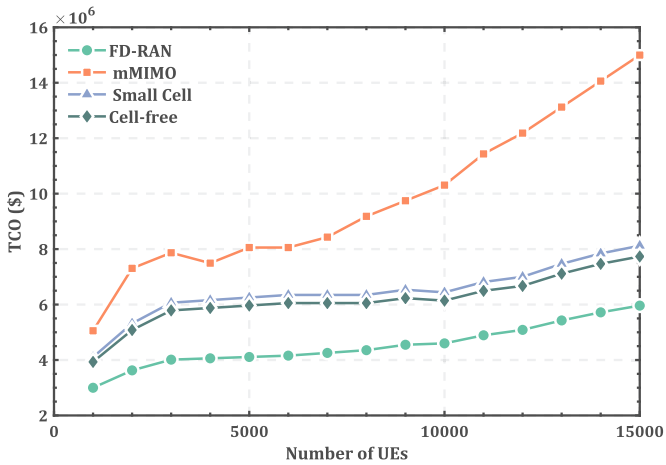

Fig. 7. Variation of TCO with the number of users for different network architectures.

CHI index compared to the reference methods. Additionally, the AGOC algorithm exhibits a CHI below 50 in fewer than 1% of cases, surpassing other hierarchical clustering techniques like HAC and FINCH.

Table II compares the performance of the AGOC algorithm with other previously mentioned algorithms. Based on the data shown in the table, AGOC outperforms several benchmark approaches for channel clustering, as indicated by the five metrics.

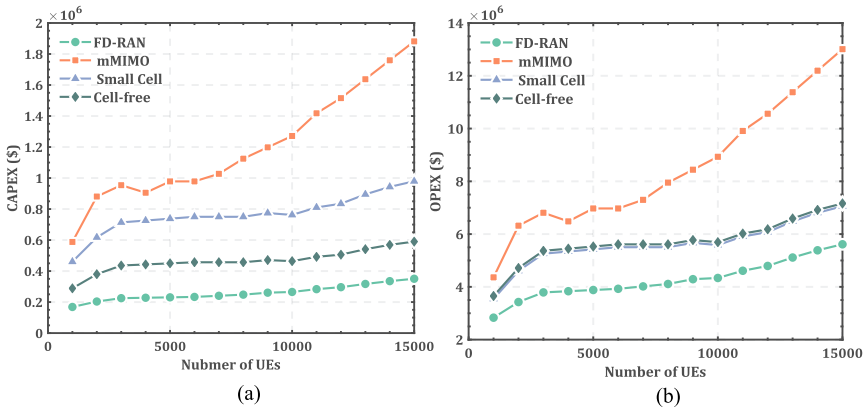

Fig. 8. Variation of CAPEX and OPEX with the number of users. (a) CAPEX. (b) OPEX.

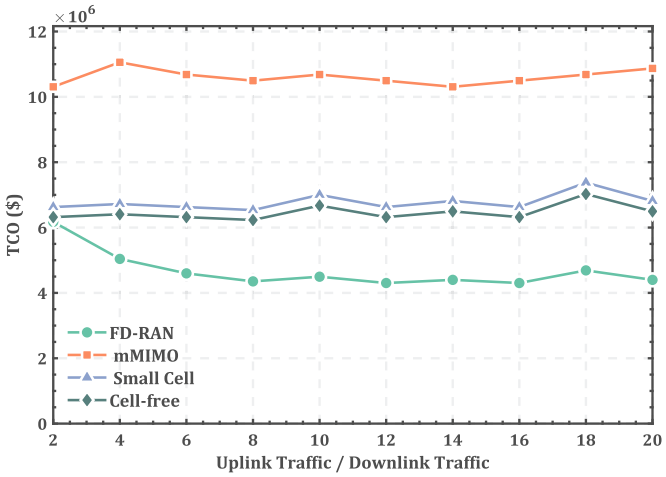

Fig. 9. Costs of FD-RAN for different U/D traffic ratios.

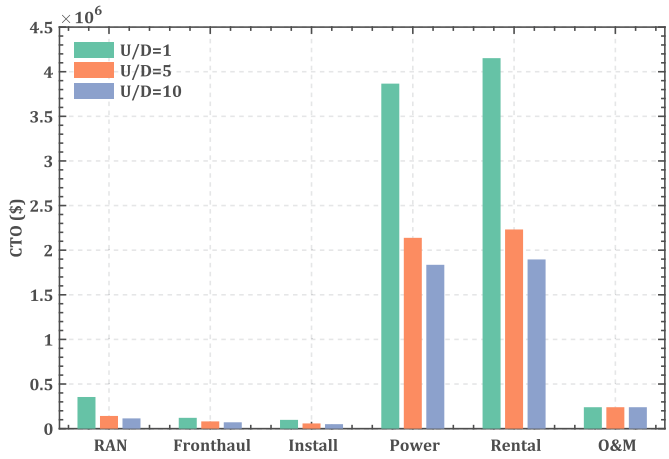

Fig. 10. Costs of FD-RAN for different cost factors.

Table II reveals that HAC exhibits suboptimal performance, with its various indicators significantly lower than those of other algorithms. This phenomenon might be attributed to the fact that channel data does not conform to a spherical space. HAC uses average distance as a benchmark, but the manifold of channel data lacks Euclidean space characteristics. The performance of the KMeans algorithm also confirms this point. In addition, compared to FINCH, the AGOC algorithm also demonstrates a comprehensive advantage in channel clustering results. This is because AGOC incorporates the AG distance index and outlier point processing mechanism, significantly enhancing the algorithm's performance.

Fig. 6 illustrates the clustering outcomes achieved by the proposed AGOC algorithm. In the figure, users belonging to the same cluster are represented with the same color. It can be observed that, regardless of whether the user distribution is linear or clustered, the proposed AGOC algorithm consistently achieves satisfactory user clustering results. In addition, by looking at the distribution of base stations, it can be seen that the proposed BCND algorithm can efficiently cover user demand hotspots: UBSs are always deployed within user aggregation areas, while DBS deployment locations can achieve more comprehensive coverage to meet a broader range of business needs at a lower cost. Although this scenario pertains to a specific area, the algorithm framework is scalable to other scenarios.

B. *Cost Analysis of Different Network Architectures*

Fig. 7 shows the variation of TCO across different network architectures as the number of users increases. Firstly, it is evident that mMIMO incurs significantly higher expenses compared to other networks in fulfilling users' service demands, primarily due to its elevated equipment costs and power consumption. FD-RAN incurs the lowest cost among all networks, primarily because it employs a smaller number of uplink network devices compared to other networks. Additionally, the power consumption of UBSs is much lower than that of full-function base stations. Furthermore, Small Cell and Cell-free networks incur similar costs because they both utilize full-function base stations, resulting in comparable power consumption and rental costs, and thereby similar overall network costs. It is noteworthy that when the number of users exceeds 10,000 for a given network, the cost growth curve exhibits an almost linear pattern. This is related to the cost model and also reflects another point: when network

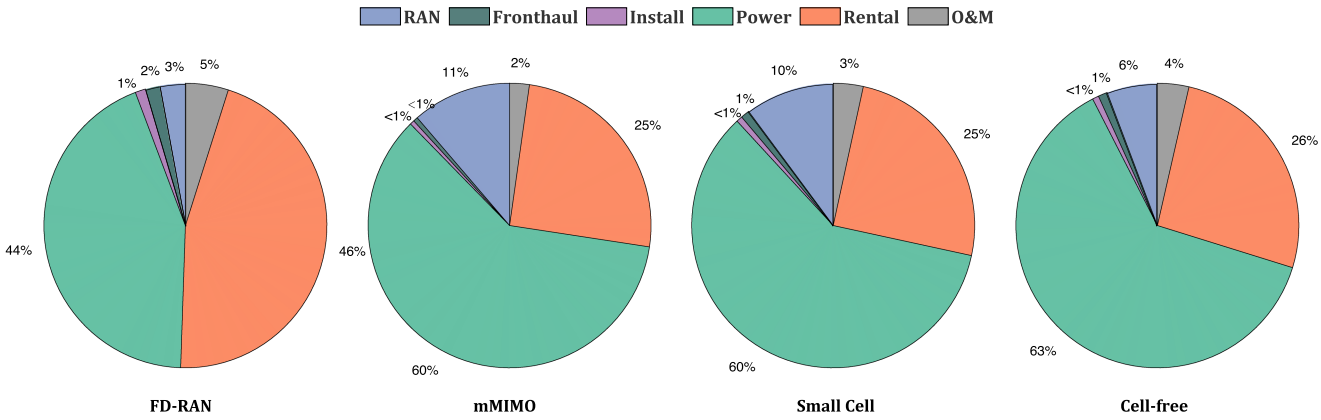

Fig. 11. Proportions of various cost factors for different network architectures.

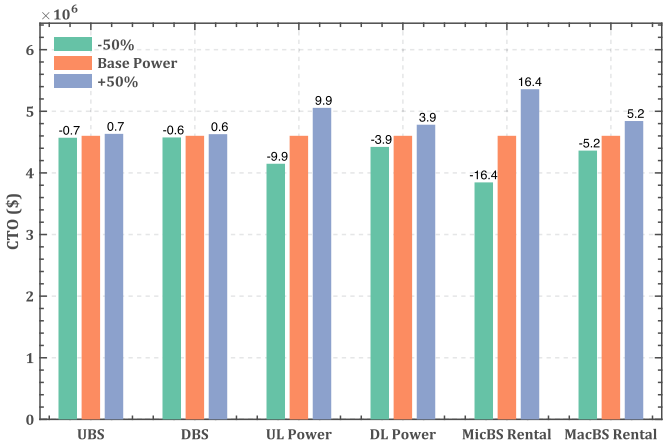

Fig. 12. Impact of price parameters on TCO.

demand is low, deployment decisions must meet the service requirements of all users across a wide area, putting the network in a coverage-limited mode. As service demand gradually increases, the network must deploy additional resources to meet higher service requirements, transitioning to a service-limited mode.

Fig. 8 illustrates how CAPEX and OPEX change with an increasing number of service users. They exhibit the same trend but differ in order of magnitude, indicating that OPEX is the primary factor affecting network cost. For CAPEX, FD-RAN's CAPEX is only 60% of Cell-free's on average, and Cell-free's cost is also much lower than Small Cell's. In addition, it can be observed that once the number of users exceeds 10,000, network CAPEX grows almost linearly. OPEX follows a similar trend. The difference lies in the fact that Small Cell and Cell-free have nearly identical OPEX due to their similar OPEX models, with their primary difference being the BBU cost. Given the magnitude difference between CAPEX and OPEX, optimizing network TCO is primarily dependent on OPEX, with energy consumption and rental costs comprising the largest proportion of OPEX. Hence, optimizing the network's deployment strategy and reducing power consumption are crucial for minimizing TCO.

Fig. 9 depicts the system cost when the uplink traffic volume is fixed and the uplink-downlink traffic ratio varies. As the uplink-downlink traffic ratio increases, the cost of the FD-RAN network gradually decreases until it stabilizes at a certain level. In contrast, the other network architectures are almost unaffected by the change in the uplink-downlink traffic ratio. In the simulation scenario presented, the equilibrium uplink-downlink ratio is approximately 8, though this threshold may vary in different scenarios. Additionally, at a ratio of 18, all networks experience a slight cost increase due to inherent simulation randomness. In the simulation, the FD-RAN's cost savings compared to mMIMO, Small Cell, and Cell-free architectures are 59%, 36%, and 33%, respectively, in the optimal case. Thus, Fig. 9 demonstrates that FD-RAN offers a notable cost benefit in commercial scenarios where uplink traffic is dominant.

C. Cost Analysis of Different Network Cost Factors

Fig. 10 shows the costs of FD-RAN for different uplink-downlink traffic ratios. It should be noted that the uplink traffic volume remains fixed when altering the U/D ratio. Hence, altering the traffic ratio results in a decrease in the amount of downlink transmission. The costs of each network in Fig. 9 decrease as the U/D increases. Nevertheless, this downward trend flattens out because DBSs must meet coverage requirements, preventing further reductions in the number of base stations. Furthermore, it is evident that network power consumption and site rental are the primary determinants influencing network expenditure.

Fig. 11 shows the proportions of different costs in different network architectures in the form of pie charts. Firstly, concerning network power consumption, FD-RAN accounts for only 44%, the lowest among all networks. In contrast, mMIMO's energy consumption ratio is nearly 60%, even if Cell-free and Small Cell also exceed 60%. This indicates that FD-RAN exhibits lower energy consumption compared to other network schemes. Secondly, site rental costs constitute a significant portion for each network architecture, reaching 46% in FD-RAN, indicating that FD-RAN increases site rental costs while decoupling deployment. Finally, the wireless cost ratio

also varies significantly in different network architectures. Since mMIMO is deployed in the form of macro base stations, the cost of a single base station is considerably elevated, resulting in a comparatively higher total cost. However, the use of decoupled low-cost UBSs and DBSs in FD-RAN results in a reduced equipment cost ratio. The most crucial factors determining network cost are power consumption, deployment style, and wireless congestion.

Fig. 12 demonstrates the impact of price parameters on network cost. It can be observed that the price of UBS/DBS has a relatively low impact on total network cost. Specifically, a 50% increase in the uplink/DBS price leads to a mere 0.7% and 0.6% rise in costs, respectively. In contrast, network power consumption and base station rent have a more substantial influence on network expenditure. When the power consumption of UBS grows by 50%, there will be a corresponding increase of 9.9% in TCO. For MicBS rental, a 50% increase in rental cost results in a corresponding 16.4% increase in network cost. Moreover, a 50% increase in DBS power consumption results in a 3.9% rise in cost, while a 50% increase in MacBS rent leads to a 5.2% rise in TCO. Hence, in forthcoming networks, reducing base station power consumption and simplifying deployment forms will be imperative to mitigate the total expenses associated with network operations.

VII. Conclusion and Future Works

In this paper, we address the challenge of low-cost network deployment for personalized services of massive terminals in FD-RAN. We propose an FD-RAN deployment cost model, TECM, and formulate a mathematical model for network decoupled deployment. To cope with the complexity of massive terminals and high-precision deployment, we introduce the AGOC algorithm, which transforms the significant 0-1 programming problem into a mixed integer program through user clustering. Leveraging FD-RAN's uplink-downlink decoupling, we decompose the original problem into separate subproblems for uplink and downlink network deployment. For the decoupled problem, we further propose the BCND algorithm to solve the two decoupled deployment subproblems separately. Simulation results indicate that FD-RAN can effectively reduce network deployment costs while ensuring user service quality for differentiated service demands, thus alleviating the profitability challenges faced by operators.

Given the dynamic nature of wireless service demands, fixed base stations are limited in terms of cost and flexibility. The development of drone communication technology offers a new avenue for dynamic network deployment and systematic resource allocation. In future work, we will explore the dynamic cooperative networking mode between drones and FD-RAN.

References

[1] H. Tataria, M. Shafi, A. F. Molisch, M. Dohler, H. Sjöland, and F. Tufvesson, "6G wireless systems: Vision, requirements, challenges, insights, and opportunities," *Proc. IEEE*, vol. 109, no. 7, pp. 1166–1199, Jul. 2021.

[2] Y.-J. Liu, H. Du, D. Niyato, G. Feng, J. Kang, and Z. Xiong, "Slicing4meta: An intelligent integration architecture with multi-dimensional network resources for metaverse-as-a-service in web 3.0," *IEEE Commun. Mag.*, vol. 61, no. 8, pp. 20–26, Aug. 2023.

[3] Q. Yu et al., "A fully-decoupled RAN architecture for 6G inspired by neurotransmission," *J. Commun. Inf. Netw.*, vol. 4, no. 4, pp. 15–23, Dec. 2019.

[4] J. Zhao et al., "Fully-decoupled radio access networks: A resilient uplink base stations cooperative reception framework," *IEEE Trans. Wireless Commun.*, vol. 22, no. 8, pp. 5096–5110, Aug. 2023.

[5] L. Jiao, K. Yu, J. Chen, T. Liu, H. Zhou, and L. Cai, "Performance analysis of uplink/downlink decoupled access in cellular-V2X networks," *IEEE Trans. Mobile Comput.*, vol. 23, no. 5, pp. 5616–5630, May 2024.

[6] H. Lu et al., "Eye accommodation-inspired neuro-metasurface focusing," *Nature Commun.*, vol. 14, no. 1, Jun. 2023, Art. no. 3301.

[7] W. Zhang et al., "Optimizing federated learning in distributed industrial IoT: A multi-agent approach," *IEEE J. Sel. Areas Commun.*, vol. 39, no. 12, pp. 3688–3703, Dec. 2021.

[8] D. Yang et al., "DetFed: Dynamic resource scheduling for deterministic federated learning over time-sensitive networks," *IEEE Trans. Mobile Comput.*, vol. 23, no. 5, pp. 5162–5178, May 2024.

[9] H. Du et al., "Exploring attention-aware network resource allocation for customized metaverse services," *IEEE Netw.*, vol. 37, no. 6, pp. 166–175, Nov. 2023.

[10] B. Mao, F. Tang, Y. Kawamoto, and N. Kato, "AI models for green communications towards 6G," *IEEE Commun. Surveys Tuts.*, vol. 24, no. 1, pp. 210–247, Firstquarter 2022.

[11] T. Qasim et al., "An ant colony optimization based approach for minimum cost coverage on 3-D grid in wireless sensor networks," *IEEE Commun. Lett.*, vol. 22, no. 6, pp. 1140–1143, Jun. 2018.

[12] D. Pliatsios, P. Sarigiannidis, I. D. Moscholios, and A. Tsiakalos, "Cost-efficient remote radio head deployment in 5G networks under minimum capacity requirements," in *Proc. IEEE 2019 Panhellenic Conf. Electron. Telecommun.*, Volos, Greece, Nov. 2019, pp. 1–4.

[13] M. Dong, M. Cho, K. Lee, S. Yoon, and T. Kim, "Cost-optimal deployment of millimeter-wave base stations under outage requirement," *IEEE Trans. Wireless Commun.*, vol. 21, no. 12, pp. 10544–10559, Dec. 2022.

[14] J. Zhao, H. Zhou, B. Qian, Y. Xu, K. Yu, and X. S. Shen, "Joint reception and parallel base stations selection in uplink fully-decoupled RAN for 6G," in *Proc. IEEE 2021 13th Int. Conf. Wireless Commun. Signal Process.*, Changsha, China, Oct. 2021, pp. 1–5.

[15] M. K. Mishra and A. Trivedi, "Spectral efficiency and deployment cost efficiency analysis of mmW/UHF-based cellular network," *IEEE Trans. Veh. Technol.*, vol. 68, no. 7, pp. 6565–6577, Jul. 2019.

[16] W. Xie, N.-T. Mao, and K. Rundberget, "Cost comparisons of backhaul transport technologies for 5G fixed wireless access," in *Proc. IEEE 2018 5G World Forum*, Silicon Valley, CA, USA, Jul. 2018, pp. 159–163.

[17] G. Gemmi, L. Cerdà-Alabern, and L. Maccari, "A realistic open-data-based cost model for wireless backhaul networks in rural areas," in *Proc. IEEE 2022 18th Int. Conf. Netw. Service Manage.*, Oct. 2022, pp. 55–63.

[18] G. Gemmi, R. Lo Cigno, and L. Maccari, "On cost-effective, reliable coverage for LoS communications in urban areas," *IEEE Trans. Netw. Service Manag.*, vol. 19, no. 3, pp. 2767–2779, Sep. 2022.

[19] S. S. Jaffer, A. Hussain, M. A. Qureshi, J. Mirza, and K. K. Qureshi, "A low cost PON-FSO based fronthaul solution for 5G CRAN architecture," *Opt. Fiber Technol.*, vol. 63, May 2021, Art. no. 102500.

[20] F. Yaghoubi et al., "A techno-economic framework for 5G transport networks," *IEEE Wireless Commun.*, vol. 25, no. 5, pp. 56–63, Oct. 2018.

[21] P. Avella, A. Calamita, and L. Palagi, "A compact formulation for the base station deployment problem in wireless networks," *Networks*, vol. 82, no. 1, pp. 52–67, 2023.

[22] M. Dong, T. Kim, J. Wu, and E. W. M. Wong, "Cost-efficient millimeter wave base station deployment in Manhattan-type geometry," *IEEE Access*, vol. 7, pp. 149959–149970, 2019.

[23] G. Liu, Y. Huang, Z. Chen, L. Liu, Q. Wang, and N. Li, "5G deployment: Standalone vs. non-standalone from the operator perspective," *IEEE Commun. Mag.*, vol. 58, no. 11, pp. 83–89, Nov. 2020.

[24] X. Liu, "A deployment strategy for multiple types of requirements in wireless sensor networks," *IEEE Trans. Cybern.*, vol. 45, no. 10, pp. 2364–2376, Oct. 2015.

[25] G. Prasad, D. Mishra, and A. Hossain, "Joint optimization framework for operational cost minimization in green coverage-constrained wireless networks," *IEEE Trans. Green Commun. Netw.*, vol. 2, no. 3, pp. 693–706, Sep. 2018.
[26] C. Huang, H. Huang, and A. V. Savkin, "Deployment of UAV base stations for wireless communication coverage," in *Autonomous Navigation and Deployment of UAVs for Communication, Surveillance and Delivery*. Hoboken, NJ, USA: Wiley Press, 2023, pp. 11–56.
[27] C.-C. Lin and J.-W. Yang, "Cost-efficient deployment of fog computing systems at logistics centers in industry 4.0," *IEEE Trans. Ind. Informat.*, vol. 14, no. 10, pp. 4603–4611, Oct. 2018.
[28] E. J. Oughton, K. Katsaros, F. Entezami, D. Kaleshi, and J. Crowcroft, "An open-source techno-economic assessment framework for 5G deployment," *IEEE Access*, vol. 7, pp. 155930–155940, 2019.
[29] C. Bouras, S. Kokkalis, A. Kollia, and A. Papazois, "Techno-economic comparison of MIMO and DAS cost models in 5G networks," *Wireless Netw.*, vol. 26, no. 1, pp. 1–15, Jan. 2020.
[30] M. A. D. Souza et al., "A techno-economic framework for installing broadband networks in rural and remote areas," *IEEE Access*, vol. 9, pp. 58421–58447, 2021.
[31] L. Furtado, A. Fernandes, A. Ohashi, F. Farias, A. Cavalcante, and J. Costa, "Cell-free massive MIMO deployments: Fronthaul topology options and techno-economic aspects," in *Proc. IEEE 2022 16th Eur. Conf. Antennas Propag.*, Madrid, Spain, Mar. 2022, pp. 1–5.
[32] E. J. Oughton and W. Lehr, "Surveying 5G techno-economic research to inform the evaluation of 6G wireless technologies," *IEEE Access*, vol. 10, pp. 25237–25257, 2022.
[33] C. Bouras, P. Ntarzanos, and A. Papazois, "Cost modeling for SDN/NFV based mobile 5G networks," in *Proc. IEEE 2016 8th Int. Congr. Ultra Modern Telecommun. Control Syst. Workshops*, Lisbon, Portugal, Oct. 2016, pp. 56–61.
[34] C. Bouras, S. Kokkalis, A. Kollia, and A. Papazois, "Techno-economic analysis of MIMO & DAS in 5G," in *Proc. IEEE 2018 11th IFIP Wireless Mobile Netw. Conf.*, Prague, Czech Republic, Sep. 2018, pp. 1–8.
[35] Y. Chen, S. Zhang, and S. Xu, "Characterizing energy efficiency and deployment efficiency relations for green architecture design," in *Proc. IEEE 2010 Int. Conf. Commun. Workshops*, Cape Town, South Africa, May 2010, pp. 1–5.
[36] Y. Chen, S. Zhang, S. Xu, and G. Y. Li, "Fundamental trade-offs on green wireless networks," *IEEE Commun. Mag.*, vol. 49, no. 6, pp. 30–37, Jun. 2011.
[37] W. Yu, L. Musavian, and Q. Ni, "Weighted tradeoff between effective capacity and energy efficiency," in *Proc. IEEE 2015 Int. Conf. Commun.*, London, U.K., Jun. 2015, pp. 238–243.
[38] K. Yu et al., "Fully-decoupled radio access networks: A flexible downlink multi-connectivity and dynamic resource cooperation framework," *IEEE Trans. Wireless Commun.*, vol. 22, no. 6, pp. 4202–4214, Jun. 2023.
[39] 3rd Generation Partnership Project (3GPP), "Study on channel model for frequencies from 0.5 to 100 GHz," 3GPP, Tech. Specification (TS) 38.901, Mar. 2022.
[40] E. Chen, M. Tao, and Y.-F. Liu, "Joint base station clustering and beamforming for non-orthogonal multicast and unicast transmission with backhaul constraints," *IEEE Trans. Wireless Commun.*, vol. 17, no. 9, pp. 6265–6279, Sep. 2018.
[41] A. Saxena et al., "A review of clustering techniques and developments," *Neurocomputing*, vol. 267, pp. 664–681, 2017.
[42] F. Murtagh and P. Contreras, "Algorithms for hierarchical clustering: An overview," *Wiley Interdiscipl. Rev.: Data Mining Knowl. Discov.*, vol. 2, no. 1, pp. 86–97, 2012.
[43] F. H. Kuwil, Ü. Atila, R. Abu-Issa, and F. Murtagh, "A novel data clustering algorithm based on gravity center methodology," *Expert Syst. Appl.*, vol. 156, 2020, Art. no. 113435.
[44] S. Sarfraz, V. Sharma, and R. Stiefelhagen, "Efficient parameter-free clustering using first neighbor relations," in *Proc. IEEE/CVF 2019 Conf. Comput. Vis. Pattern Recognit.*, Long Beach, CA, USA, Jun. 2019, pp. 8926–8935.
[45] M.-F. Balcan, A. Blum, and S. Vempala, "A discriminative framework for clustering via similarity functions," in *Proc. 40th Annu. ACM Symp. Theory Comput.*, 2008, pp. 671–680.
[46] L. Buitinck et al., "API design for machine learning software: Experiences from the scikit-learn project," in *Proc. ECML PKDD Workshop: Lang. Data Mining Mach. Learn.*, 2013, pp. 108–122.
[47] J. A. Hartigan, "Statistical theory in clustering," *J. Classification*, vol. 2, pp. 63–76, 1985.

Jiwei Zhao (Member, IEEE) received the M.S. degrees in communication and information system from Xidian University, Xi'an, China, in 2015, and the Ph.D. degree in information and communication engineering from Nanjing University, Nanjing, China, in 2022. He is currently a Postdoctoral Fellow with Zhejiang University, Hangzhou, China. His research interests include resource management in B5G/6G networks, metasurface, and machine learning applications for wireless communication. Dr. Zhao was the recipient of Best Paper Award from IEEE WSCP 2021 and PIERS 2023.

Jiacheng Chen (Member, IEEE) received the Ph.D. degree in information and communications engineering from Shanghai Jiao Tong University, Shanghai, China, in 2018. From 2015 to 2016, he was a Visiting Scholar with BBCR Group, University of Waterloo, Waterloo, ON, Canada. He is currently an Assistant Researcher with Peng Cheng Laboratory, Shenzhen, China. His research focuses on fully-decoupled radio access network technologies. From 2021 to 2024, he was the Guest Editor ofIEEE INTERNET OF THINGS JOURNAL, *Journal of Communications and Information Networks* (JCIN), and Workshop Co-Chair of IEEE/CIC ICCC. He was the recipient of the JCIN Best Paper Award in 2016 and Chinese Institute of Electronics (CIE) Outstanding Scientific Paper in the Field of Electronic Information in 2020.

Bo Cheng (Member, IEEE) received the M.S. degree in information and communication engineering from the South China University of Technology, Guangzhou, China, in 2013. He is currently working toward the Ph.D. degree with the School of Electronic Science and Engineering, Nanjing University, Nanjing, China. He is also the Project General Manager of the Innovation Base Operations Center, China United Network Communications Corporation Guangzhou Branch, Guangzhou. His research focuses on real-time PUE optimization algorithms and their applications in IDC data centers based on AI.

Bo Qian (Member, IEEE) received the B.S. and M.S. degrees in statistics from Sichuan University, Chengdu, China, in 2015 and 2018, respectively, and the Ph.D. degree in information and communication engineering from Nanjing University, Nanjing, China, in 2022. From 2022 to 2024, he was a Postdoctoral Fellow with Peng Cheng Laboratory, Shenzhen, China. He is currently a Project Researcher with the National Institute of Informatics, Tokyo, Japan. His research interests include resource management in B5G/6G networks, VANET, network economics, blockchain, and game theory. Dr. Qian was the recipient of Best Paper Award from IEEE VTC2020-Fall.

Yunting Xu (Student Member, IEEE) received the B.S. degree in communication engineering from Nanjing University, Nanjing, China, in 2017. He is currently working toward the Ph.D. degree with the School of Electronic Science and Engineering, Nanjing University. His research interests include wireless network access, dynamic resource management, and personalized wireless service in the field of next generation wireless networks.

Haibo Zhou (Senior Member, IEEE) received the Ph.D. degree in information and communication engineering from Shanghai Jiao Tong University, Shanghai, China, in 2014. From 2014 to 2017, he was a Postdoctoral Fellow with the Broadband Communications Research Group, Department of Electrical and Computer Engineering, University of Waterloo, Waterloo, ON, Canada. He is currently a Full Professor with the School of Electronic Science and Engineering, Nanjing University, Nanjing, China. His research interests include resource management and protocol design in B5G/6G networks, vehicular ad hoc networks, and space-air-ground integrated networks. He was the recipient of the 2019 IEEE ComSoc Asia-Pacific Outstanding Young Researcher Award, 2023 IEEE ComSoc WTC Outstanding Young Researcher Award, 2023-2024 IEEE ComSoc Distinguished Lecturer, and 2023-2025 IEcEE VTS Distinguished Lecturer. He was the Track/Symposium Co-Chair for IEEE/CIC ICCC 2019, IEEE VTC-Fall 2020, IEEE VTC-Fall 2021, WCSP 2022, IEEE GLOBECOM 2022, IEEE ICC 2024, and IEEE GLOBECOM 2024. He is an Associate Editor for the IEEE TRANSACTIONS ON WIRELESS COMMUNICATIONS, IEEE INTERNET of Things Journal, IEEE NETWORK MAGAZINE, and IEEE JOURNAL OF COMMUNICATIONS AND INFORMATION NETWORKS.