

Received October 9, 2021, accepted October 20, 2021, date of publication November 2, 2021, date of current version November 9, 2021.

Digital Object Identifier 10.1109/ACCESS.2021.3124816

Cell-Free Massive MIMO System With an Adaptive Switching Algorithm Between Cooperative NOMA, Non-Cooperative NOMA, and OMA Modes

REZA SAYYARI^{ID}, JAFAR POURROSTAM^{ID}, AND MIR JAVAD MUSEVI NIYA^{ID}, (Member, IEEE)

Faculty of Electrical and Computer Engineering, University of Tabriz, Tabriz 5166616471, Iran

Corresponding author: Jafar Pourrostam (j.pourrostam@tabrizu.ac.ir)

ABSTRACT Non-orthogonal multiple access (NOMA) aided cell-free massive multiple-input multiple-output (CF mMIMO) systems have gained great attention over the past few years to be employed in the sixth generation of the wireless network. With distributed antennas over a large area, CF mMIMO systems have much better cell coverage and higher achievable sum-rate compared to the conventional cellular and small-cell networks. This paper will propose a novel cooperative NOMA-aided CF mMIMO system. The proposed scheme will increase the achievable sum-rate and improve the system reliability by creating additional data streams between users within the same NOMA cluster. Also, we will suggest an adaptive switching algorithm between non-cooperative NOMA, cooperative NOMA, and orthogonal multiple access (OMA) modes in order to maximize the power efficiency and achievable sum-rate. We will derive closed-form expressions for the achievable downlink data rates of the mentioned modes. Lastly, we will offer a special user clustering algorithm that needs no complex optimization and ensures the existence of reliable cooperative links between the users. The simulation results will prove the superiority of the proposed scheme over the conventional NOMA-based and OMA-based CF mMIMO systems.

INDEX TERMS Cell-free massive MIMO, cooperative NOMA, non-orthogonal multiple access (NOMA), user grouping, 6G.

I. INTRODUCTION

The fifth generation of the wireless network (5G) was full of novel and enabling technologies, like the use of millimeter-wave bands and massive multiple-input multiple-output (mMIMO) technique. However, it has some serious drawbacks, such as uneven cell coverage and a trade-off between battery life and data rate. The sixth generation of the wireless network (6G) is planning to remove all of those barriers that exist in 5G. Many papers are presented to address the key indicators and introduce the enabling technologies for 6G [1]–[5]. New technologies such as virtual and augmented reality (VAR), real-time remote health care, autonomous transportation, and intellectual industry systems need ultra-reliable high-speed connections. For holographic teleportation, we will need a data rate up to 5 Tbit/s [1]. Remote surgery needs an ultra-reliable link with extremely

The associate editor coordinating the review of this manuscript and approving it for publication was Jiayi Zhang^{ID}.

TABLE 1. 5G and 6G parameter comparison.

Parameter	5G	6G
Peak data rate (Tbit/s)	0.02	>1
Latency (ms)	1	<1
Connection density (million devices/km ²)	1	10
Mobility (km/h)	500	>1000
Experienced spectrum efficiency (b/s/Hz)	0.3	3
End-to-end reliability (%)	99.999	99.99999

low latency [6] to guarantee patients' lives. An autonomous transportation system needs a lot of connected sensors [7] with above 1 Tbit/s links at a speed of up to 1000 km/h in order to ensure passengers' safety. For handling those demands, 6G has a long way. It should manage massive connectivity with ultra-reliable connections and low latency at a data rate above 1 Tbit/s. Table 1 shows some of the key performance indicators (KPIs) for 6G and compares it to 5G [1]–[3].

Achieving the mentioned aims with the conventional cellular and small-cell network topologies is pretty challenging. CF mMIMO [8] is one of the most advantageous network topologies for 6G. In CF mMIMO network, a large number of single-antenna access points (APs) are distributed over a large area instead of placing many antennas on a single base station. The APs employ conjugate beamforming to transmit all users' data over the same time-frequency resources. CF mMIMO networks have many benefits, such as removing cell boundaries and a low distance between the users and APs. Terahertz frequency band (between 100 GHz and 10 THz) is one of the promising candidates for 6G. However, these kinds of ultra-high frequency waves have a very high path loss. CF mMIMO is the best network scheme for matching the terahertz communication due to the low distance between users and APs. In [9], the authors proposed normalized-conjugate beamforming, which has a better achievable data rate than the conventional conjugate beamforming. A user-centric approach, in which only a limited number of APs transmit each user's data, is introduced in [10]. Reference [11] studied the benefits and challenges of the CF mMIMO system. Reference [12] offered an iterative robust minimum mean-square error precoder in order to mitigate the interference caused by imperfect channel state information. Reference [13] analyzed achievable data rate of the CF mMIMO systems considering the limitations of back-haul connection. [14] proposed a power control algorithm for the pilot transmission in order to minimize the interference caused by non-orthogonal pilot signals. Reference [15] studied the positive effects of downlink pilot transmission in CF mMIMO systems. The authors in [16] have studied channel hardening in CF mMIMO systems. Reference [17] proposed a generalized benders decomposition based algorithm to solve the power and user grouping problems. In [18], the authors proposed an efficient pilot assignment scheme using a weighted graphic framework to mitigate the effect of pilot contamination.

In CF mMIMO systems, uplink pilot transmission is necessary for conjugate beamforming. Users transmit data and pilot signals using the time-division duplexing (TDD) technique in the same coherence time. However, wireless channels have limited coherence time, and hence, the users have a limited number of orthogonal pilot sequences, which is not enough for massive connectivity. NOMA [19] is one of the leading multiple access techniques that transmits different users' data over the same time-frequency resources but with different power levels. At the receiver, each user employs successive interference cancelation (SIC) to decode its data. Reference [20] proposed NOMA-aided CF mMIMO system in which users within the same NOMA cluster employ the same orthogonal pilot sequence. This approach can considerably solve the orthogonal pilot sequence shortage problem, and the network can support more connected users. However, due to the errors like intra-cluster interference, pilot contamination, and imperfect SIC in the NOMA-based approach, OMA-aided CF mMIMO has better performance

than the NOMA-aided one when there are only a few connected users. In [21], the authors proposed an adaptive switching mode between NOMA and OMA to maximize the achievable sum-rate. Reference [22] investigated the uplink performance of a NOMA-aided CF mMIMO system and offered a novel total spectral efficiency maximization algorithm. In [23], the authors suggested an optimal backhaul combining process, which maximizes the uplink signal-to-interference-plus-noise ratio (SINR). Reference [24] studied the achievable rate of NOMA-aided CF mMIMO underlaid below a primary mMIMO network and proposed a sub-optimal user-clustering method. Reference [25] analyzed the impact of the type of precoders used at the APs. A closed-form expression for the achievable downlink data rate of a NOMA-aided CF mMIMO system over spatially correlated Rician fading channel is derived in [26]. Reference [27] modeled the location of users and APs as homogeneous Poisson point processes (PPPs). Reference [28] analyzed the impact of hardware mismatches on the performance of NOMA-aided CF mMIMO systems. Reference [29] studied the performance of NOMA-aided CF mMIMO system for future communication networks under stochastic AP and user locations.

In each NOMA cluster, users with better channel conditions (stronger users) get a lower fraction of power than users with weaker channels (weaker users). Stronger users need to perform SIC in order to decode their data. Firstly, they have to decode the weaker users' data and then subtract it from the received signal. In cooperative NOMA [30], stronger users re-transmit the decoded data of prior users (weaker users). Those additional links between users within the same NOMA cluster can considerably guarantee the connectivity of all users, even if their direct link connections are unreliable.

In this paper, we will investigate the performance of a cooperative NOMA-aided CF mMIMO network. The addition of cooperative links will increase the achievable sum-rate and improve the system reliability. However, some of those re-transmissions are unnecessary because the weaker user might have an acceptable direct link connection. Hence, the re-transmission in that cases will degrade the energy efficiency of the system. We will propose an adaptive switching algorithm between non-cooperative NOMA, cooperative NOMA, and OMA modes in order to solve the previous problem and maximize the system performance. We will derive closed-form expressions for the achievable downlink data rates of non-cooperative NOMA-aided, cooperative NOMA-aided, and OMA-aided CF mMIMO systems. One of the main challenges in NOMA-based systems is user clustering. Most of the previous works, such as [23] and [31], used complicated algorithms for solving the user clustering problem. We will offer a simple user clustering technique that guarantees the existence of reliable channels between users within the same NOMA cluster. Lastly, we will use computer simulations to prove the performance of the proposed method and compare it to the conventional ones.

A. NOTATIONS

Matrices and vectors are defined with upper-case boldface and lower-case boldface letters. An Italic letter denotes a scalar variable. $(\cdot)^T$, $(\cdot)^H$, and $(\cdot)^*$ mean the transpose, conjugate transpose, and conjugate of (\cdot) , respectively. $a \sim CN(\mu, \sigma)$ indicates a complex normal random variable with a mean of μ and covariance of σ . $E\{\cdot\}$ and $|\cdot|$ represent expectation and absolute operators. Finally, $a.b^T$ indicates the inner product of vectors a and b with an equal length of n , which defines as $\sum_{i=1}^n a_i b_i$.

II. SYSTEM MODEL

Figure 1 depicts a NOMA-aided CF mMIMO network in which M single-antenna APs are serving $2L$ single-antenna users simultaneously, where L is the total number of NOMA clusters. Since having more than two users in each NOMA cluster is not practical and causes so much complexity in SIC, we assume there are only two users in each NOMA cluster. All APs are connected to the central processing unit (CPU) via a perfect backhaul connection. h_{mlk} indicates the channel between the k^{th} user within the l^{th} cluster and the m^{th} AP, and $g_{lk'k}$ indicates the channel between user k and user k' within the l^{th} cluster. h_{mlk} can be modeled as

$$h_{mlk} = \sqrt{\beta_{mlk}} \alpha_{mlk}, \quad (1)$$

where β_{mlk} and α_{mlk} are the large-scale and small-scale fading coefficients, respectively. The large-scale fading changes about 40 times slower than the small-scale fading [32]. Hence, we assumed large-scale fading values have already been estimated, and are known to the users and APs. $\alpha_{mlk} \sim CN(0, 1)$ is considered to be a circularly-symmetric complex normal random variable with zero mean and unit covariance, which is independent for different mlk . Thus, $h_{mlk} \sim CN(0, \beta_{mlk})$ is a complex normal random variable with zero mean and covariance of β_{mlk} . Also, users within the same NOMA cluster are arranged according to their channel conditions:

$$\sum_{m=1}^M |h_{ml1}|^2 \geq \sum_{m=1}^M |h_{ml2}|^2; \quad 1 \leq l \leq L. \quad (2)$$

The location of users and APs is modeled with homogeneous PPP [27]. The following equation shows the probability of having V APs or users with a density of λ within an area of B [33]:

$$\text{probability}[Z(B) = V] = \frac{(\lambda B)^V e^{-\lambda B}}{V!}. \quad (3)$$

If λ_u is the density of users and λ_a is the density of APs, then $\lambda_a \geq \frac{\lambda_u}{K}$. K is the total number of users within the same NOMA cluster, which is equal to 2 for the considered NOMA-aided CF mMIMO network and 1 for the OMA-aided one.

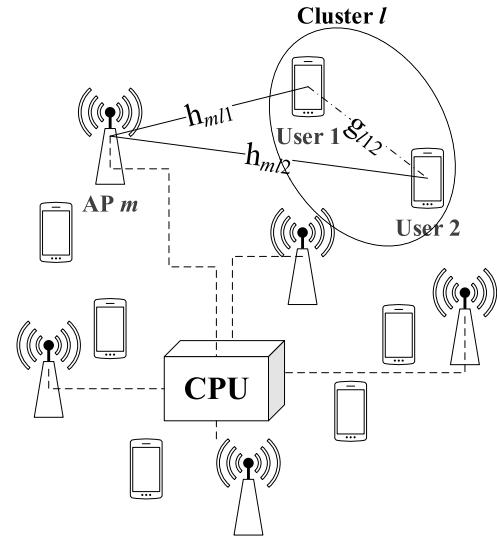


FIGURE 1. NOMA-aided cell-free massive MIMO network with two users in each cluster.

A. CHANNEL ESTIMATION

In the studied NOMA-aided CF mMIMO network, APs perform downlink channel estimation using the received pilot signals from the users. Users rely on channel statistics as a real channel and do not perform any extra channel estimation. They utilize the TDD technique to transmit data and pilot signals in the same coherence time. Due to the limited channel coherence interval τ_c (in samples), a limited number of orthogonal pilot sequences is available ($\tau_c \geq \tau_{upp}$, where τ_{upp} is the length of the pilot sequence). In order to overcome the mentioned problem, we use the same pilot sequence for the users within the same NOMA cluster [34]. The received pilot signal at the m^{th} AP will be as follows:

$$r_m^{upp} = \sqrt{\tau_{upp} P_{upp}} \sum_{l=1}^L \sum_{k=1}^2 h_{mlk} \varphi_l + n_m, \quad (4)$$

where $n_m \sim CN(0_{\tau_{upp} \times 1}, 1_{\tau_{upp} \times 1})$ is the additive white gaussian noise (AWGN) vector at the m^{th} AP, and P_{upp} is the normalized uplink pilot signal-to-noise ratio (SNR). φ_l is the assigned pilot sequence to the l^{th} cluster, which satisfies the following conditions:

$$\varphi_l \varphi_i^H = \begin{cases} 0, & l \neq i \\ 1, & l = i. \end{cases} \quad (5)$$

The m^{th} AP multiplies the received pilot signal with φ_l^H in order to estimate the channel:

$$\begin{aligned} \tilde{r}_{ml}^{upp} &= r_m^{upp} \varphi_l^H \\ &= \sqrt{\tau_{upp} P_{upp}} \sum_{k=1}^2 h_{mlk} + \underbrace{n_m \varphi_l^H}_{\tilde{n}_{ml}}. \end{aligned} \quad (6)$$

The users within the same NOMA cluster are using the same pilot sequence. Hence, the APs can only estimate a linear combination of their channels. Using (5), \tilde{r}_{ml}^{upp} becomes

a sufficient statistic. Therefore, the minimum mean square error (MMSE) estimation of h_{mlk} can be written as [35]

$$\begin{aligned}\hat{h}_{mlk} &= \frac{E\left\{\tilde{r}_{ml}^{upp*} h_{mlk}\right\}}{E\left\{\left|\tilde{r}_{ml}^{upp}\right|^2\right\}} \tilde{r}_{ml}^{upp} \\ &= \frac{\sqrt{\tau_{upp} P_{upp}} \beta_{mlk}}{1 + \tau_{upp} P_{upp} \sum_{k=1}^K \beta_{mlk}} \tilde{r}_{ml}^{upp}.\end{aligned}\quad (7)$$

Since \tilde{r}_{ml}^{upp} is a complex normal random variable with zero mean and covariance of $(1 + \tau_{upp} P_{upp} \sum_{k=1}^2 \beta_{mlk})$, \hat{h}_{mlk} can be written as a complex normal random variable with zero mean and covariance of $\left(\frac{\tau_{upp} P_{upp} \beta_{mlk}^2}{1 + \tau_{upp} P_{upp} \sum_{k=1}^2 \beta_{mlk}}\right)$. Now \hat{h}_{mlk} can be re-written as

$$\hat{h}_{mlk} = \sqrt{\gamma_{mlk}} b_{mlk}, \quad (8)$$

where $b_{mlk} \sim CN(0, 1)$, and γ_{mlk} is as follows:

$$\begin{aligned}\gamma_{mlk} &= E\left\{\left|\hat{h}_{mlk}\right|^2\right\} \\ &= \frac{\tau_{upp} P_{upp} \beta_{mlk}^2}{1 + \tau_{upp} P_{upp} \sum_{k=1}^2 \beta_{mlk}}.\end{aligned}\quad (9)$$

B. SIGNAL MODEL

Using the estimated channel state information (CSI) in (8), APs employ normalized-conjugate beamforming [9] to transmit users' data. The beamforming vector can be defined as

$$c_{lk} = \left[\frac{\hat{h}_{1lk}^*}{|\hat{h}_{1lk}|}, \frac{\hat{h}_{2lk}^*}{|\hat{h}_{2lk}|}, \dots, \frac{\hat{h}_{Mlk}^*}{|\hat{h}_{Mlk}|} \right]^T. \quad (10)$$

β_{mlk} , τ_{upp} , and P_{upp} are positive real numbers. Using (8) and (10), the beamforming vector can be re-written as

$$c_{lk} = c_l = \left[\frac{b_{1l}^*}{|b_{1l}|}, \frac{b_{2l}^*}{|b_{2l}|}, \dots, \frac{b_{Ml}^*}{|b_{Ml}|} \right]^T. \quad (11)$$

Consider P_{mlk} as the allocated power to the k^{th} user within the l^{th} cluster from the m^{th} AP. We can define the power allocation matrix as follows:

$$P_{lk} = \begin{bmatrix} P_{1lk} & 0 & \dots & 0 \\ 0 & P_{2lk} & & \cdot \\ \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & 0 \\ 0 & \cdot & \cdot & 0 & P_{Mlk} \end{bmatrix}. \quad (12)$$

Via multiplying the power allocation matrix by the beamforming vector, the following power control vector defines:

$$\begin{aligned}\tilde{h}_{lk} &= P_{lk} c_l \\ &= \left[\sqrt{P_{1lk}} \frac{b_{1l}^*}{|b_{1l}|}, \sqrt{P_{2lk}} \frac{b_{2l}^*}{|b_{2l}|}, \dots, \sqrt{P_{Mlk}} \frac{b_{Ml}^*}{|b_{Ml}|} \right]^T.\end{aligned}\quad (13)$$

Using normalized-conjugate beamforming, the transmitted signal from the m^{th} AP to all users within L clusters is given by

$$t_m = \sqrt{P_{dl}} \sum_{l=1}^L \sqrt{P_{ml}} \frac{b_{ml}^*}{|b_{ml}|} \bar{q}_l, \quad (14)$$

where P_{dl} is the normalized downlink SNR, $P_{ml} = \sum_{k=1}^2 P_{mlk}$ is the total allocated power to the l^{th} cluster from the m^{th} AP, and \bar{q}_l is the superposition coded signal of all users within the l^{th} cluster in which $E\{\left|\bar{q}_l\right|^2\} = 1$. Short-term power constraint [36] prevents the transmit power from exceeding a predetermined value in order to avoid interference and environmental problems. If the predetermined value is P_{dl} , then the short-term power constraint can be written as

$$E\left\{\left|t_m\right|^2\right\} \leq P_{dl}. \quad (15)$$

Using 14 and 15, the short-term power constraint defines as

$$\sum_{l=1}^L P_{ml} = \sum_{l=1}^L \sum_{k=1}^K P_{mlk} \leq 1; \quad \text{for all } m. \quad (16)$$

An efficient selection of P_{mlk} is a max-min power control problem, which is beyond the scope of this paper.

III. ACHIEVABLE DATA RATE ANALYSIS

A. NON-COOPERATIVE NOMA AIDED CELL-FREE MASSIVE MIMO

In a non-cooperative NOMA-aided CF mMIMO network, the received signal at the k^{th} user within the l^{th} cluster will be as follows:

$$r_{l,k}^{dl,NOMA} = \sqrt{P_{dl}} \sum_{l'=1}^L \sum_{k'=1}^2 h_{lk} \tilde{h}_{l'k'}^T q_{l'k'} + n_{lk}, \quad (17)$$

where $n_{lk} \sim CN(0, 1)$ indicates the AWGN noise at the k^{th} user within the l^{th} cluster, $h_{lk} = [h_{1lk}, h_{2lk}, \dots, h_{Mlk}]^T$ is the channel coefficient vector, and q_{lk} is the message signal for the k^{th} user within the l^{th} cluster in which $E\{\left|q_{lk}\right|^2\} = 1$.

The second user does not need to perform SIC in order to decode its message. It simply takes the first user's data as noise. We can re-write the received signal at the second user as

$$\begin{aligned}r_{l,2}^{dl,NOMA} &= \underbrace{\sqrt{P_{dl}} E\left\{h_{l2} \tilde{h}_{l2}^T\right\}}_{(DS)_{l2}} q_{l2} \\ &\quad + \underbrace{\sqrt{P_{dl}} \left(h_{l2} \tilde{h}_{l2}^T - E\left\{h_{l2} \tilde{h}_{l2}^T\right\}\right)}_{(BU)_{l2}} q_{l2} \\ &\quad + \underbrace{\sqrt{P_{dl}} h_{l2} \tilde{h}_{l1}^T}_{(ICI)_{l1}} q_{l1}\end{aligned}$$

$$+ \sum_{l' \neq l}^L \sum_{k'''=1}^2 \underbrace{\sqrt{P_{dl}} h_{l2} \tilde{h}_{l'k''}^T q_{l'k''}}_{(IUI)_{l'k''}} + n_{l2}, \quad (18)$$

where $(DS)_{lk}$ indicates the desired signal for the k^{th} user within the l^{th} cluster and $(BU)_{lk}$ denotes beamforming uncertainty for the k^{th} user within the l^{th} . $(ICI)_{l1}$ represents intra-cluster interference caused by the first user within the same cluster. $(IUI)_{l'k''}$ expresses inter-user interference caused by the users within the different clusters.

The first user has to perform SIC to get its data. In this regard, it should firstly decode the second user's data and then subtract it from the received signal. After the SIC process, we can express the received signal of the first user within the l^{th} cluster as follows:

$$\begin{aligned} r_{l1}^{dl,NOMA} = & \underbrace{\sqrt{P_{dl}} E \left\{ h_{l1} \tilde{h}_{l1}^T \right\} q_{l1}}_{(DS)_{l1}} \\ & + \underbrace{\sqrt{P_{dl}} \left(h_{l1} \tilde{h}_{l1}^T - E \left\{ h_{l1} \tilde{h}_{l1}^T \right\} \right) q_{l1}}_{(BU)_{l1}} \\ & + \underbrace{\sqrt{P_{dl}} \left(h_{l1} \tilde{h}_{l2}^T - E \left\{ h_{l1} \tilde{h}_{l2}^T \right\} \right) q_{l2}}_{(ISIC)_{l2}} \\ & + \sum_{l' \neq l}^L \sum_{k'''=1}^2 \underbrace{\sqrt{P_{dl}} h_{l1} \tilde{h}_{l'k''}^T q_{l'k''}}_{(IUI)_{l'k''}} + n_{l1}, \quad (19) \end{aligned}$$

where $(ISIC)_{l2}$ denotes the interference caused by imperfect SIC of the second user's data.

Now we can define the signal-to-interference-plus-noise ratios (SINRs) of the first and second users within the l^{th} cluster as (20), shown at the bottom of the next page.

q_{l1} , q_{l2} , and BU are independent of each other [8]. Hence, BU, ICI, ISIC, and IUI are uncorrelated with DS. Considering uncorrelated noise as the worst-case [37], the achievable downlink data rate of the k^{th} user within the l^{th} cluster can be written as follows:

$$R_{lk}^{dl,NOMA} = BW^{direct} \rho \log_2 \left(1 + (SINR)_{lk,k}^{dl,NOMA} \right), \quad (21)$$

where BW^{direct} is the allocated bandwidth to the direct link transmission and ρ is a pre-log factor, which is inserted to model the effect of pilot and data transmission in the same coherence time. ρ can be defined as

$$\rho = 1 - \frac{\tau_{upp}}{\tau_c}. \quad (22)$$

The downlink achievable sum-rate for the described non-cooperative NOMA-aided CF mMIMO network is given by

$$R_{sum}^{dl,NOMA} = \sum_{l=1}^L \sum_{k=1}^2 R_{lk}^{dl,NOMA}. \quad (23)$$

Theorem 1: In a non-cooperative NOMA-aided CF mMIMO network with two users in each NOMA cluster, the achievable downlink data rates can be written as follows:

$$\begin{aligned} \tilde{R}_{l1}^{dl,NOMA} = & BW^{direct} \rho \\ & \times \log_2 \left(1 + \frac{P_{dl} \frac{\pi}{4} \left(\sum_{m=1}^M \sqrt{P_{ml1} \gamma_{ml1}} \right)^2}{N_{1,1} + N_{3,1} + N_{4,1} + 1} \right), \end{aligned} \quad (24a)$$

$$\begin{aligned} \tilde{R}_{l2}^{dl,NOMA} = & BW^{direct} \rho \\ & \times \log_2 \left(1 + \frac{P_{dl} \frac{\pi}{4} \left(\sum_{m=1}^M \sqrt{P_{ml2} \gamma_{ml2}} \right)^2}{N_{1,2} + N_{2,2} + N_{4,2} + 1} \right), \end{aligned} \quad (24b)$$

where

$$N_{1,k} = P_{dl} \sum_{m=1}^M P_{mlk} \beta_{mlk} - P_{dl} \frac{\pi}{4} \sum_{m=1}^M P_{mlk} \gamma_{mlk}, \quad (25a)$$

$$N_{2,k} = P_{dl} \sum_{m=1}^M P_{ml1} \beta_{mlk}, \quad (25b)$$

$$N_{3,k} = P_{dl} \sum_{m=1}^M P_{ml2} \beta_{mlk} - P_{dl} \frac{\pi}{4} \sum_{m=1}^M P_{ml2} \gamma_{mlk}, \quad (25c)$$

$$N_{4,k} = \sum_{l' \neq l}^L \sum_{k'''=1}^2 P_{dl} \sum_{m=1}^M P_{ml'k''} \beta_{mlk}. \quad (25d)$$

Proof: See Appendix A.

B. COOPERATIVE NOMA AIDED CELL-FREE MASSIVE MIMO

In the proposed cooperative NOMA-aided CF mMIMO network, the first user within each cluster utilizes short-range communication to re-transmit the second user's data after decoding it for the SIC process. Using short-range communication avoids the interference between the direct and cooperative links because their data are being transmitted over different time-frequency resources. Table 2 shows some of the possible technologies and network types to be used in the cooperative data transmission. Considering an ultra-dense network for 6G, the existence of a line of sight (LOS) between the users within the same cluster is a reasonable assumption. Therefore, visible light communication, which needs LOS for connection, and terahertz communication, which has high path-loss, can be the possible candidates to be employed in cooperative links. WiFi and Bluetooth are the other available technologies to be used in cooperative data transmission. The addition of cooperative links helps users who are facing severe fading to receive their data from different channels. The received signal for the second user within the l^{th} cluster

TABLE 2. Some of the possible technologies and network types to be used in the cooperative link [38].

Network Type	Frequency Band	Transmission Range	Achievable Rate
WiFi	2.4GHz, 5GHz	120m	11-600Mb/s
Bluetooth	2.4GHz	100m	~1Mb/s
Visible Light	430-790THz	<100m (LOS)	~500Mb/s
THz wave	0.3-3THz	<1km	1Tb/s

from the cooperative link will be as follows:

$$r_{l2}^{cl} = \sqrt{P_{cl}} g_{l12} s_{l2} + n_{l2}, \quad (26)$$

where P_{cl} is the normalized cooperative link SNR and s_{l2} is the second user's decoded data by the first user within the l^{th} cluster in which $E\{|s_{lk}|^2\} = 1$. Using (1) and (26), the cooperative link SINR of the second user can be defined as

$$(\text{SINR})_{l2}^{cl} = E \left\{ \left| \sqrt{P_{cl}} g_{l12} \right|^2 \right\} = P_{cl} \hat{\beta}_{l12}, \quad (27)$$

where $\hat{\beta}_{l12}$ indicates the large-scale fading coefficient between the first and second users within the l^{th} cluster.

Also, the SINR of the second user's data for the first user within the l^{th} cluster can be defined as follows:

$$(\text{SINR})_{l1,2}^{dl,NOMA} = \frac{\left| \sqrt{P_{dl}} E \left\{ h_{l1} \tilde{h}_{l2}^T \right\} \right|^2}{\sum_{i=1}^3 \dot{N}_i + 1}, \quad (28)$$

where

$$\dot{N}_1 = E \left\{ \left| \sqrt{P_{dl}} \left(h_{l1} \tilde{h}_{l2}^T - E \left\{ h_{l1} \tilde{h}_{l2}^T \right\} \right) \right|^2 \right\}, \quad (29a)$$

$$\dot{N}_2 = E \left\{ \left| \sqrt{P_{dl}} h_{l1} \tilde{h}_{l1}^T \right|^2 \right\}, \quad (29b)$$

$$\dot{N}_3 = \sum_{l' \neq l}^L \sum_{k'=1}^2 E \left\{ \left| \sqrt{P_{dl}} h_{l1} \tilde{h}_{l'k'}^T \right|^2 \right\}. \quad (29c)$$

Theorem 2: $(\text{SINR})_{l1,2}^{dl,NOMA}$ can be written as

$$(\text{SINR})_{l1,2}^{dl,NOMA} = \frac{P_{dl} \frac{\pi}{4} \left(\sum_{m=1}^M \sqrt{P_{ml2} \gamma_{ml1}} \right)^2}{\sum_{i=1}^3 \dot{N}_i + 1}, \quad (30)$$

where

$$\ddot{N}_1 = P_{dl} \sum_{m=1}^M P_{ml2} \beta_{ml1} - P_{dl} \frac{\pi}{4} \sum_{m=1}^M P_{ml2} \gamma_{ml1}, \quad (31a)$$

$$\ddot{N}_2 = P_{dl} \sum_{m=1}^M P_{ml1} \beta_{ml1}, \quad (31b)$$

$$\ddot{N}_3 = \sum_{l' \neq l}^L \sum_{k''=1}^2 P_{dl} \sum_{m=1}^M P_{ml'k''} \beta_{ml1}. \quad (31c)$$

Proof: Similar to Appendix A.

Now we can define the achievable data rate for the cooperative link within the l^{th} cluster as (32), shown at the bottom of the next page,

where BW^{cl} indicates the allocated bandwidth to the cooperative link transmission. Finally, the downlink achievable sum-rate of the proposed cooperative NOMA-aided CF mMIMO network can be written as follows:

$$R_{\text{sum}}^{dl,C-NOMA} = \sum_{l=1}^L \left(R_{l1}^{dl,NOMA} + \max \left(R_{l2}^{dl,NOMA}, R_{l2}^{dl,cl} \right) \right). \quad (33)$$

C. OMA AIDED CELL-FREE MASSIVE MIMO

OMA-aided CF mMIMO is a special type of non-cooperative NOMA-aided CF mMIMO network in which there is only one user in each NOMA cluster. In this case, the received signal at the u^{th} user, which is equal to the l^{th} cluster in the NOMA, can be written as follows:

$$\begin{aligned} r_u^{dl,OMA} &= \sqrt{P_{dl}} \sum_{u'=1}^U h_u \tilde{h}_{u'}^T q_{u'} + n_u \\ &= \underbrace{\sqrt{P_{dl}} E \left\{ h_u \tilde{h}_u^T \right\} q_u}_{(DS)_u} \end{aligned}$$

$$(\text{SINR})_{l1,1}^{dl,NOMA} = \frac{|DS_{l1}|^2}{E \left\{ |BU_{l1}|^2 \right\} + E \left\{ |ISIC_{l2}|^2 \right\} + \sum_{l' \neq l}^L \sum_{k''=1}^2 E \left\{ |IUI_{l'k''}|^2 \right\} + 1} \quad (20a)$$

$$(\text{SINR})_{l2,2}^{dl,NOMA} = \frac{|DS_{l2}|^2}{E \left\{ |BU_{l2}|^2 \right\} + E \left\{ |IC_{l1}|^2 \right\} + \sum_{l' \neq l}^L \sum_{k''=1}^2 E \left\{ |IUI_{l'k''}|^2 \right\} + 1} \quad (20b)$$

$$\begin{aligned}
& + \underbrace{\sqrt{P_{dl}} \left(h_u \tilde{h}_u^T - E \left\{ h_u \tilde{h}_u^T \right\} \right) q_u}_{(BU)_u} \\
& + \sum_{u' \neq u}^U \underbrace{\sqrt{P_{dl}} h_u \tilde{h}_u^T q_{u'}}_{(IUI)_{uu'}} + n_u,
\end{aligned} \quad (34)$$

where $h_u = h_{l1}$, $\tilde{h}_u = \tilde{h}_{l1}$, $q_u = q_{l1}$, $U = L$, and $n_u \sim CN(0, 1)$ indicates the AWGN noise at the u^{th} user.

The SINR of the u^{th} user can be written as

$$\begin{aligned}
(SINR)_u^{dl, OMA} &= \frac{|(DS)_u|^2}{E \left\{ |(BU)_u|^2 \right\} + \sum_{u' \neq u}^U E \left\{ |(IUI)_{uu'}|^2 \right\} + 1}. \quad (35)
\end{aligned}$$

Now we can define the achievable downlink data rate of the u^{th} user as

$$R_u^{dl, OMA} = BW^{direct} \rho \log_2 \left(1 + (SINR)_u^{dl, OMA} \right). \quad (36)$$

Ultimately, the achievable sum-rate of an OMA-aided CF mMIMO network can be defined as

$$R_{sum}^{dl, OMA} = \sum_{u=1}^U R_u^{dl, OMA}. \quad (37)$$

Theorem 3: In an OMA-aided CF mMIMO network, the achievable downlink data rate of the u^{th} user can be written as follows:

$$\begin{aligned}
\tilde{R}_u^{dl, OMA} &= BW^{direct} \rho \\
&\times \log_2 \left(1 + \frac{P_{dl} \frac{\pi}{4} \left(\sum_{m=1}^M \sqrt{P_{mu} \gamma_{mu}} \right)^2}{\sum_{i=1}^2 \ddot{N}_i + 1} \right), \quad (38)
\end{aligned}$$

where

$$\ddot{N}_1 = P_{dl} \sum_{m=1}^M P_{mu} \beta_{mu} - P_{dl} \frac{\pi}{4} \sum_{m=1}^M P_{mu} \gamma_{mu}, \quad (39a)$$

$$\ddot{N}_2 = \sum_{u' \neq u}^U P_{dl} \sum_{m=1}^M P_{mu'} \beta_{mu}, \quad (39b)$$

where $\gamma_{mu} = \gamma_{ml1}$, and $\beta_{mu} = \beta_{ml1}$ is the large-scale fading coefficient between the m^{th} AP and the u^{th} user.

Proof: Similar to Appendix A.

IV. PERFORMANCE MAXIMIZATION

A. ADOPTIVE SWITCHING ALGORITHM

Due to the lack of errors caused by imperfect SIC and intra-cluster interference, OMA-aided CF mMIMO has better performance than the NOMA-aided one when there are few

Algorithm 1 The Proposed Adaptive Switching Algorithm for the k^{th} User Within the l^{th} Cluster

```

Calculate  $R_{sum}^{dl, OMA}$  and  $R_{sum}^{dl, NOMA}$  ;
if  $R_{sum}^{dl, OMA} > R_{sum}^{dl, NOMA}$  ;
then
| Operate in OMA mode;
else
| Switch to NOMA mode;
for  $k = 2$  do
| if the SNR of the second user >  $P_{Threshold}$  ;
| then
| | Operate in non-cooperative NOMA mode ;
| else
| | Request re-transmission from the first user
| | within the  $l^{th}$  cluster ;
for  $k = 1$  do
| if any re-transmission request is received from
| the  $l^{th}$  cluster ;
| then
| | Re-transmit the second user's data using
| | short-range communication
| else
| | Operate in non-cooperative NOMA mode

```

users connected to the network. On the other hand, when there are more connected users, NOMA-aided CF mMIMO has better performance than the OMA-aided one because of using a single orthogonal pilot sequence for all users within the same NOMA cluster. In cooperative NOMA-aided CF mMIMO, users with better channel conditions re-transmit the weaker ones' data. This re-transmission increases the system reliability and improves the system performance. However, blind re-transmission reduces the energy efficiency of the system. In this section, we will offer an adaptive switching algorithm between OMA, non-cooperative NOMA, and cooperative NOMA modes to solve the mentioned problems. The proposed algorithm maximizes the achievable sum-rate and energy efficiency of the system.

In the proposed algorithm, the system initially operates in OMA mode. If the achievable sum-rate of the NOMA mode exceeds the achievable sum-rate of the OMA mode, the system switches to NOMA mode. In other words, while $R_{sum}^{dl, OMA}$ is greater than $R_{sum}^{dl, NOMA}$, the system works in OMA mode. Otherwise, it switches to NOMA mode. The optimal point for switching between OMA and NOMA modes depends on the channel coherence time. This relation will be more clear in section V. When the system is operating in NOMA mode, if the SNR of the second user is below

$$R_{l2}^{dl, cl} = \min \left(BW^{direct} \rho \log_2 \left(1 + (SINR)_{l1,2}^{dl, NOMA} \right), BW^{cl} \log_2 \left(1 + (SINR)_{l2}^{cl} \right) \right) \quad (32)$$

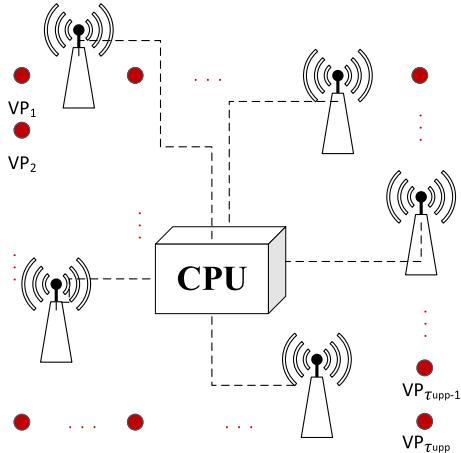


FIGURE 2. Distribution of the VPs on the network.

a predetermined threshold, the user requests re-transmission from the first user within the same NOMA cluster. This process prevents unnecessary data transmissions and improves the energy efficiency of users. Algorithm 1 illustrates the explained procedure. Note that switching between OMA and NOMA modes depends only on the number of connected users and channel coherence time. The variation of fading coefficients will not alter it because those changes increase or decrease the sum-rate in both modes. On the other hand, switching between cooperative and non-cooperative NOMA modes in each cluster depends only on the second user's SNR and its channel fading coefficient.

B. USER GROUPING

User grouping is one of the main challenges in NOMA-based systems. This problem is even more challenging in NOMA-aided CF mMIMO networks because users within the same cluster share the same pilot sequence. More distance between the users within the same cluster will result in high channel estimation error and weak beamforming. Unlike conventional cellular networks, users within the same NOMA cluster do not need to have a noticeable channel gain difference because there is no cell-edge user in cell-free networks. In recent papers, this problem has mostly been solved via complicated optimization problems. In this part, we will propose a novel user clustering algorithm using the cooperative links between the users. The proposed method does not need any complex optimization. Additionally, it guarantees the existence of reliable channels among the users within the same NOMA cluster. Hence, this method can considerably help the system to reach the ultra-reliability goal.

Consider there are τ_{upp} available orthogonal pilot sequences. First, the system specifies τ_{upp} virtual points (VPs) on the area, which are equally distributed (as depicted in figure 2). VPs are imaginary points used to help the system evenly divide the region into τ_{upp} clusters. The position of VPs is fixed, and their distance from each AP is clear. Hence, the system can easily calculate the large-scale fading

TABLE 3. Simulation parameters.

Parameter	Value
f	1.9 GHz
BW^{direct}	20 MHz
BW^{cl}	5 MHz
\tilde{P}_{dl}	200 mW
$\tilde{P}_{upp}, \tilde{P}_{cl}$	100 mW
Noise figure	9 dB
h_{AP}	15 meter
h_U	1.65 meter
d_0, d_1	10 meter, 50 meter

coefficients between the VPs and APs. Then, it finds τ_{upp} users whose large-scale fading coefficients have the minimum Euclidean distance with the VPs' large-scale fading coefficients:

$$\min_u \left(\sum_{m=1}^M |\beta_{mu} - \tilde{\beta}_{mp}| \right); \quad \text{for } p = \{1, \dots, \tau_{upp}\}, \quad (40)$$

where $\tilde{\beta}_{mp}$ is the large-scale fading coefficient between the m^{th} AP and p^{th} VP. Note that β_{mu} values have already been estimated and are known to the users and APs.

Now each of the non-selected users broadcasts a pilot sequence using short-range communication. The primary users select the user of the strongest received signal to be in the same NOMA cluster. Note that being a primary user does not necessarily mean being the first user in the NOMA cluster.

V. SIMULATION RESULTS

In this section, we will analyze the performance of the proposed scheme using computer simulations. For modeling the large-scale fading, a three-slope path loss model [39] has been employed in which $\beta_{mlk} = (PL)_{mlk}$. $(PL)_{mlk}$ denotes path loss between the m^{th} AP and k^{th} user within the l^{th} cluster and defines as

$$(PL)_{mlk} = \begin{cases} -L - 35\log_{10}(d_{mlk}), & \text{if } d_{mlk} > d_1 \\ -L - 15\log_{10}(d_1) - 20\log_{10}(d_{mlk}), & \text{if } d_0 < d_{mlk} < d_1 \\ -L - 15\log_{10}(d_1) - 20\log_{10}(d_0), & \text{if } d_{mlk} \leq d_0; \\ \end{cases}$$

$$L = 46.3 + 33.9\log_{10}(f) - 13.82\log_{10}(h_{AP}) - (1.1\log_{10}(f) - 0.7)h_U + (1.56\log_{10}(f) - 0.8), \quad (41)$$

where h_{AP} and h_U are the AP's and user's antenna height, f (MHz) represents the carrier frequency, and d_{mlk} indicates the distance between the m^{th} AP and k^{th} user within the l^{th} cluster.

We assumed $M = 200$ uniformly distributed single-antenna APs are transmitting users' data within an area of size $1 \times 1 \text{ km}^2$. Figure 3 shows the topology diagram of the

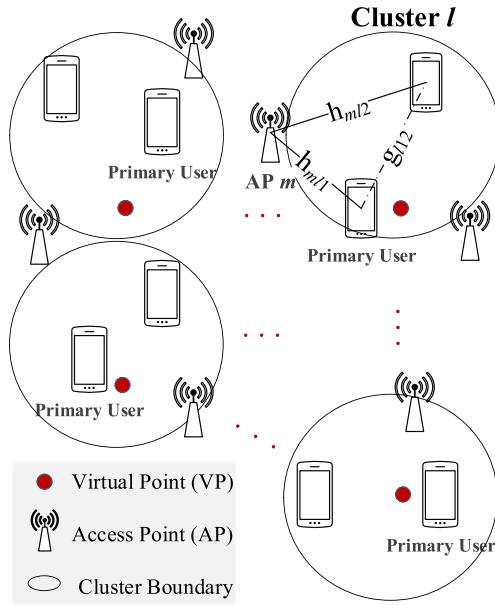


FIGURE 3. Topological diagram of the simulated network.

simulated network. VPs are equally distributed over the area. The users are clustered according to the user grouping method discussed in section IV-B. As can be seen, the selection of primary users with the help of VPs has helped the network to have evenly distributed clusters. The APs employ conjugate beamforming to transmit all users' data over the same time-frequency resources. In each NOMA cluster, if the SNR of the weaker user is below a predetermined value, it asks the stronger user to retransmit the data after decoding it for SIC. The system model and channel estimation used in this simulation have been modeled in section II. Channel coherence internal (τ_c) is set to 196 symbols unless otherwise specified. Table 3 summarises the used parameters in this simulation. \tilde{P}_{dl} , \tilde{P}_{upp} and \tilde{P}_{cl} represent the downlink data transmission, pilot transmission and cooperative link data transmission power values, respectively. For having a fair comparison among the methods, we do not consider any max-min power control algorithm. Hence, equal power allocation is used in the simulations. The power allocation ratio inside the NOMA clusters is 3:7, which means that the first user gets 30% and the second user gets 70% of the total transmit power in each NOMA cluster.

Figure 4 shows the achievable sum-rate versus the number of users for the proposed cooperative NOMA-aided CF mMIMO system and compares it with the conventional OMA and NOMA aided CF mMIMO systems. Also, it presents the performance of a CF mMIMO system with the suggested adaptive switching algorithm. As can be seen, NOMA-based CF mMIMO systems can support two times more users than the OMA-based one because of sharing the same pilot sequence between the users within the same NOMA cluster.

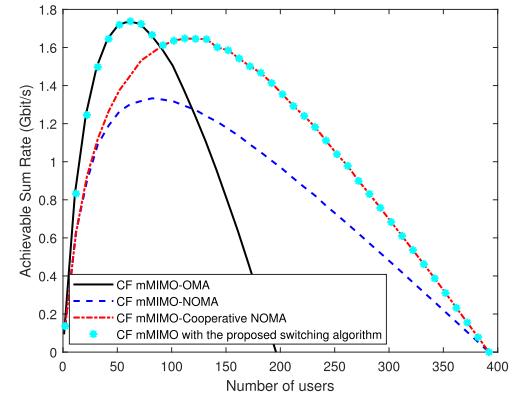


FIGURE 4. Achievable sum-rate versus the number of users. For NOMA-based systems, there are two users in each cluster and $M = 200$ APs are transmitting users' data.

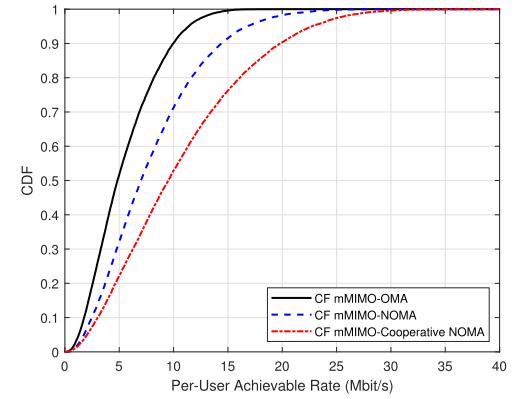


FIGURE 5. CDFs of the downlink per-user achievable data rate for three CF mMIMO systems with OMA, NOMA, and cooperative-NOMA multiple access schemes ($M = 200$, $\tau_c = 196$, total number of users=150).

On the other hand, when there are a small number of users connected to the network, the OMA-aided system has better performance than the NOMA-aided one due to the lack of errors caused by imperfect SIC and intra-cluster interference. The proposed cooperative NOMA-aided CF mMIMO system has a much higher sum-rate compared to the non-cooperative one. For the studied system, the optimal point for switching between OMA and non-cooperative NOMA modes is when the number of users is 118. Also, the optimal point for switching between OMA and cooperative NOMA modes is when the number of users is 90. By employing the recommended adaptive switching algorithm, the system can always reach the highest possible sum-rate.

Figure 5 and figure 6 show cumulative distribution functions (CDFs) of the per-user downlink achievable data rate for the OMA, non-cooperative NOMA, and cooperative NOMA aided CF mMIMO systems. In figure 5, there are 150 users connected to the network. As can be seen, the proposed cooperative NOMA-aided CF mMIMO system has the best performance compared to the conventional OMA and NOMA aided CF mMIMO systems. The 90%-likely per-user achievable rate for the cooperative NOMA-aided

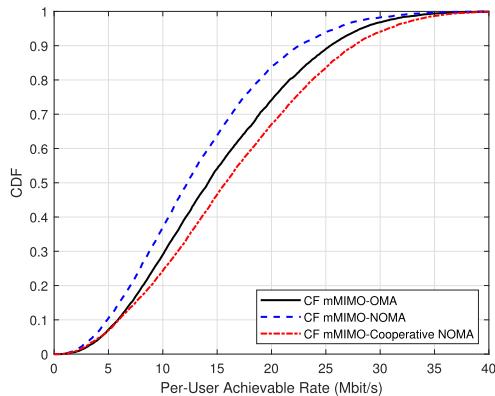


FIGURE 6. CDFs of the downlink per-user achievable data rate for three CF mMIMO systems with OMA, NOMA, and cooperative-NOMA multiple access schemes ($M = 200$, $\tau_c = 196$, total number of users=100).

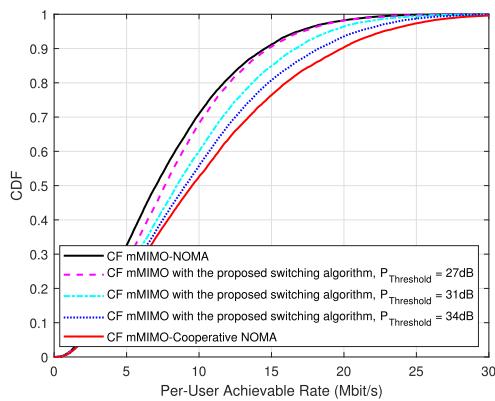


FIGURE 7. CDFs of the downlink per-user achievable data rate for CF mMIMO systems with the proposed switching algorithm and different threshold values ($M = 200$, $\tau_c = 196$, total number of users = 150).

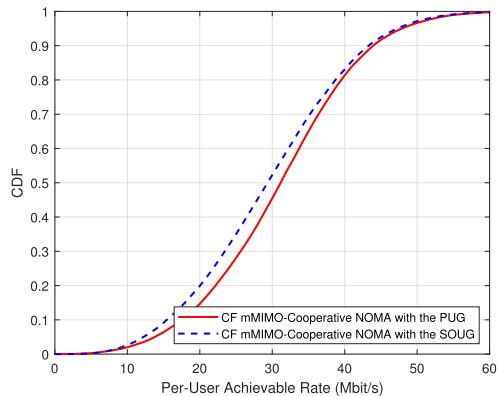


FIGURE 8. CDFs of the downlink per-user achievable data rate for cooperative NOMA aided CF mMIMO systems with the proposed user grouping (PUG) method and the mentioned suboptimal user grouping (SOUQ) one ($M = 200$, $\tau_c = 196$, total number of users=40).

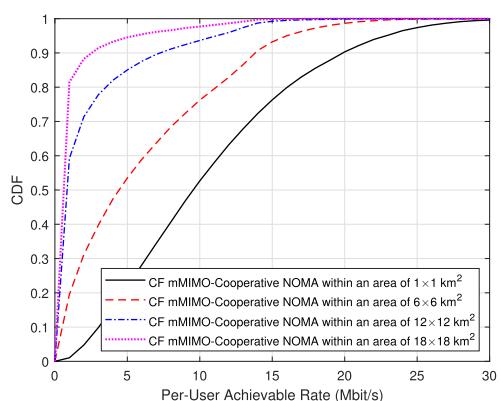


FIGURE 9. CDFs of the downlink per-user achievable data rate for cooperative NOMA-aided CF mMIMO systems within different areas ($M = 200$, $\tau_c = 196$, total number of users = 150).

CF mMIMO system is 20 Mbit/s while the 90%-likely per-user achievable rate for the conventional OMA and NOMA aided CF mMIMO systems are 10 Mbit/s and 15 Mbit/s, respectively. Figure 6 illustrates CDFs of the per-user downlink achievable rate for the compared systems when there are 100 users connected to the network. Due to the lack of errors like imperfect SIC, the OMA mode has better performance than the non-cooperative NOMA mode. However, the proposed cooperative NOMA mode still has the best performance. The 90%-likely per-user achievable rate for the cooperative NOMA-aided CF mMIMO is 27.5 Mbit/s, which is about 2.2 Mbit/s and 5 Mbit/s more than the OMA and non-cooperative NOMA modes.

Figure 7 demonstrates CDFs of the downlink per-user achievable data rate for CF mMIMO systems with the proposed switching algorithm and different threshold values. As can be seen, by increasing the threshold value, the system acts as a cooperative NOMA-aided one because the second user frequently asks for re-transmission. On the other hand, by decreasing the threshold value, the system functions as a non-cooperative NOMA-aided one because the second user's received SNR is usually higher than the threshold value.

Optimal user grouping is a challenging problem in NOMA-based systems. According to [40], for solving such a problem, the system must search over $\frac{(4L)!}{2^{2L}}$ possibilities, which is impractical in a network with a large number of users and we are unable to simulate it. For solving this problem, we simplified the optimization and used a suboptimal method. In this method, instead of maximizing the network sum rate by grouping users, we select two users that have the maximum sum rate when grouped together. Note that being the optimal pair for maximizing the paired user's data rate is not always equal to being the optimal pair for maximizing the network sum rate. Hence, this suboptimal user grouping method will be less effective than the optimal one. Figure 8 compares our proposed user grouping method with the mention suboptimal one. As can be seen, our proposed user grouping method performs better with less complexity than the mentioned suboptimal method. This indicates that our proposed user clustering algorithm is closer to the optimal.

Figure 9 shows CDFs of the downlink per-user achievable data rate for the proposed cooperative NOMA-aided CF mMIMO system with 200 APs and 150 users within different

areas. As can be seen, by increasing the area size, the achievable rate decreases because the large-scale fading between users and APs increases. Also, due to the enlargement of clusters, the distance between users within the same cluster increases, which leads to lower data rates in cooperative links.

VI. CONCLUSION

Cell-free massive MIMO systems are one of the most prominent candidates to be employed in the sixth generation of the wireless network. However, due to the limited channel coherence interval, they mostly suffer from the orthogonal pilot sequence shortage problem. By employing NOMA in CF mMIMO systems, we can assign a single orthogonal pilot sequence to all users within the same NOMA cluster. This approach can considerably dwindle the orthogonal pilot sequence shortage problem. However, when there are a small number of users connected to the network, the OMA-aided CF mMIMO systems outperform the NOMA-aided ones in the term of achievable sum-rate due to the lack of errors caused by imperfect SIC and intra-cluster interference. In this paper, we proposed a novel CF mMIMO scheme based on the cooperative NOMA technique. The suggested method can significantly increase the achievable sum-rate and improve the system reliability by creating extra data streams between users within the same NOMA cluster. Also, we derived closed-form expressions for the achievable downlink data rates of the OMA, non-cooperative NOMA, and cooperative NOMA aided Cf massive MIMO systems. We suggested an adaptive switching algorithm between OMA, non-cooperative NOMA, and cooperative NOMA modes to maximize the efficiency and performance of the system. Finally, we offered a special user clustering procedure, which needs no complex optimization and guarantees the existence of reliable links between the users within the same cluster. Simulation results prove that our proposed method has a much better performance than the previous ones and can be one of the leading candidates for the next generation of wireless networks.

APPENDIX

In this appendix, we will derive closed-form expressions for $|DS_{lk}|^2$, $E\{|BU_{lk}|^2\}$, $E\{|ICI_{lk'}|^2\}$, $E\{|ISIC_{lk''}|^2\}$, and $E\{|IUI_{l'k'''}|^2\}$ in order to prove theorem 1.

The power of the desired signal for the k^{th} user within the l^{th} cluster is given by

$$\begin{aligned} |DS_{lk}|^2 &= P_{dl} \left| E \left\{ h_{lk} \tilde{h}_{lk}^T \right\} \right|^2 \\ &= P_{dl} \left| E \left\{ \sum_{m=1}^M \sqrt{P_{mlk}} h_{mlk} \frac{b_{ml}^*}{|b_{ml}|} \right\} \right|^2. \end{aligned} \quad (42)$$

Let's define $e_{mlk} = h_{mlk} - \hat{h}_{mlk}$ as the channel estimation error, where \hat{h}_{mlk} and e_{mlk} are independent. Using (8) we have

$$|DS_{lk}|^2 = P_{dl} \left| \sum_{m=1}^M \sqrt{P_{mlk}} \right|^2 \left| \left(\sqrt{\gamma_{mlk}} b_{ml} + e_{mlk} \right) \frac{b_{ml}^*}{|b_{ml}|} \right|^2. \quad (43)$$

The magnitude of $b_{ml} \sim CN(0, 1)$ has Rayleigh distribution. Hence, we have

$$|DS_{lk}|^2 = P_{dl} \frac{\pi}{4} \left(\sum_{m=1}^M \sqrt{P_{mlk} \gamma_{mlk}} \right)^2. \quad (44)$$

Next, we can write $E\{|BU_{lk}|^2\}$ as

$$\begin{aligned} E\{|BU_{lk}|^2\} &= P_{dl} var \left\{ h_{lk} \tilde{h}_{lk}^T \right\} \\ &= P_{dl} E \left\{ \left| h_{lk} \tilde{h}_{lk}^T - E \left\{ h_{lk} \tilde{h}_{lk}^T \right\} \right|^2 \right\} \\ &= P_{dl} \sum_{m=1}^M P_{mlk} E \underbrace{\left\{ \left| h_{mlk} \frac{b_{ml}^*}{|b_{ml}|} \right|^2 \right\}}_{(I)} \\ &\quad - P_{dl} \sum_{m=1}^M P_{mlk} \left| E \left\{ h_{mlk} \frac{b_{ml}^*}{|b_{ml}|} \right\} \right|^2, \end{aligned} \quad (45)$$

where (I) is

$$\begin{aligned} E \left\{ \left| h_{mlk} \frac{b_{ml}^*}{|b_{ml}|} \right|^2 \right\} &= E \left\{ \left(h_{mlk} \frac{b_{ml}^*}{|b_{ml}|} \right) \left(h_{mlk}^* \frac{b_{ml}}{|b_{ml}|} \right) \right\} \\ &= E \left\{ |h_{mlk}|^2 \right\} = \beta_{mlk}. \end{aligned} \quad (46)$$

Using 44, 45, and 46, we have

$$\begin{aligned} E\{|BU_{lk}|^2\} &= P_{dl} \sum_{m=1}^M P_{mlk} \beta_{mlk} \\ &\quad - P_{dl} \frac{\pi}{4} \sum_{m=1}^M P_{mlk} \gamma_{mlk}. \end{aligned} \quad (47)$$

Now let's write the intra-cluster interference from the k'^{th} user to the k^{th} user within the l^{th} cluster as follows:

$$\begin{aligned} E\{|ICI_{lk'}|^2\} &= P_{dl} E \left\{ \left| h_{lk} \tilde{h}_{lk'}^T \right|^2 \right\} \\ &= P_{dl} E \left\{ \left| \sum_{m=1}^M \sqrt{P_{mlk'}} h_{mlk} \frac{b_{ml}^*}{|b_{ml}|} \right|^2 \right\} \\ &= P_{dl} E \left\{ \left(\sum_{m=1}^M \sqrt{P_{mlk'}} h_{mlk} \frac{b_{ml}^*}{|b_{ml}|} \right) \right. \\ &\quad \left. \left(\sum_{m'=1}^M \sqrt{P_{m'l'}} h_{m'l'}^* \frac{b_{m'l}}{|b_{m'l}|} \right) \right\} \\ &= P_{dl} \sum_{m=1}^M P_{mlk'} E \left\{ |h_{mlk}|^2 \right\} \\ &= P_{dl} \sum_{m=1}^M P_{mlk'} \beta_{mlk}. \end{aligned} \quad (48)$$

Later, $E \{ |ISIC_{lk''}|^2 \}$ can be written as

$$\begin{aligned}
E \{ |ISIC_{lk''}|^2 \} &= P_{dl} var \left\{ h_{lk} \tilde{h}_{lk''}^T \right\} \\
&= P_{dl} E \left\{ \left| h_{lk} \tilde{h}_{lk''}^T - E \left\{ h_{lk} \tilde{h}_{lk''}^T \right\} \right|^2 \right\} \\
&= P_{dl} \sum_{m=1}^M P_{mlk''} E \left\{ \left| h_{mlk} \frac{b_{ml}^*}{|b_{ml}|} \right|^2 \right\} \\
&\quad - P_{dl} \sum_{m=1}^M P_{mlk''} \left| E \left\{ h_{mlk} \frac{b_{ml}^*}{|b_{ml}|} \right\} \right|^2 \\
&= P_{dl} \sum_{m=1}^M P_{mlk''} \beta_{mlk} \\
&\quad - P_{dl} \frac{\pi}{4} \sum_{m=1}^M P_{mlk''} \gamma_{mlk}. \tag{49}
\end{aligned}$$

Finally, we can calculate $E \{ |IUI_{l'k''}|^2 \}$ as follows:

$$\begin{aligned}
E \{ |IUI_{l'k''}|^2 \} &= P_{dl} E \left\{ \left| h_{lk} \tilde{h}_{lk}^T \right|^2 \right\} \\
&= P_{dl} E \left\{ \left| \sum_{m=1}^M \sqrt{P_{ml'k''}} h_{mlk} \frac{b_{ml}^*}{|b_{ml}|} \right|^2 \right\} \\
&= P_{dl} \sum_{m=1}^M P_{ml'k''} \beta_{mlk}. \tag{50}
\end{aligned}$$

REFERENCES

- [1] I. F. Akyildiz, A. Kak, and S. Nie, “6G and beyond: The future of wireless communications systems,” *IEEE Access*, vol. 8, pp. 133995–134030, 2020.
- [2] L. U. Khan, I. Yaqoob, M. Imran, Z. Han, and C. S. Hong, “6G wireless systems: A vision, architectural elements, and future directions,” *IEEE Access*, vol. 8, pp. 147029–147044, 2020.
- [3] A. Dogra, R. K. Jha, and S. Jain, “A survey on beyond 5G network with the advent of 6G: Architecture and emerging technologies,” *IEEE Access*, vol. 9, pp. 67512–67547, 2021.
- [4] C. D. Alwis, A. Kalla, Q.-V. Pham, P. Kumar, K. Dev, W.-J. Hwang, and M. Liyanage, “Survey on 6G frontiers: Trends, applications, requirements, technologies and future research,” *IEEE Open J. Commun. Soc.*, vol. 2, pp. 836–886, 2021.
- [5] W. Jiang, B. Han, M. A. Habibi, and H. D. Schotten, “The road towards 6G: A comprehensive survey,” *IEEE Open J. Commun. Soc.*, vol. 2, pp. 334–366, 2021.
- [6] S. Mohapatra, S. Mohanty, and S. Mohanty, “Smart healthcare: An approach for ubiquitous healthcare management using IoT,” in *Big Data Analytics for Intelligent Healthcare Management*, N. Dey, H. Das, B. Naik, and H. S. Behera, Eds. New York, NY, USA: Academic, 2019, ch. 7, pp. 175–196.
- [7] J. Choi, V. Va, N. Gonzalez-Prelcic, R. Daniels, C. R. Bhat, and R. W. Heath, Jr., “Millimeter-wave vehicular communication to support massive automotive sensing,” *IEEE Commun. Mag.*, vol. 54, no. 12, pp. 160–167, Dec. 2016.
- [8] H. Q. Ngo, A. Ashikhmin, H. Yang, E. G. Larsson, and T. L. Marzetta, “Cell-free massive MIMO versus small cells,” *IEEE Trans. Wireless Commun.*, vol. 16, no. 3, pp. 1834–1850, Mar. 2017.
- [9] G. Interdonato, H. Q. Ngo, E. G. Larsson, and P. Frenger, “On the performance of cell-free massive MIMO with short-term power constraints,” in *Proc. IEEE Int. Workshop Comput. Aided Model. Design Commun. Links Netw. (CAMAD)*, Oct. 2016, pp. 225–230.
- [10] S. Buzzi and C. D’Andrea, “Cell-free massive MIMO: User-centric approach,” *IEEE Wireless Commun. Lett.*, vol. 6, no. 6, pp. 706–709, Dec. 2017.
- [11] J. Zhang, S. Chen, Y. Lin, J. Zheng, B. Ai, and L. Hanzo, “Cell-free massive MIMO: A new next-generation paradigm,” *IEEE Access*, vol. 7, pp. 99878–99888, 2019.
- [12] V. M. T. Palhares, A. R. Flores, and R. C. de Lamare, “Robust MMSE precoding and power allocation for cell-free massive MIMO systems,” *IEEE Trans. Veh. Technol.*, vol. 70, no. 5, pp. 5115–5120, May 2021.
- [13] M. Bashar, K. Cumanan, A. G. Burr, H. Q. Ngo, and M. Debbah, “Cell-free massive MIMO with limited backhaul,” in *Proc. IEEE Int. Conf. Commun. (ICC)*, May 2018, pp. 1–7.
- [14] T. C. Mai, H. Q. Ngo, M. Egan, and T. Q. Duong, “Pilot power control for cell-free massive MIMO,” *IEEE Trans. Veh. Technol.*, vol. 67, no. 11, pp. 11264–11268, Nov. 2018.
- [15] G. Interdonato, H. Q. Ngo, E. G. Larsson, and P. Frenger, “How much do downlink pilots improve cell-free massive MIMO?” in *Proc. IEEE Global Commun. Conf. (GLOBECOM)*, Dec. 2016, pp. 1–7.
- [16] Z. Chen and E. Björnson, “Channel hardening and favorable propagation in cell-free massive MIMO with stochastic geometry,” *IEEE Trans. Commun.*, vol. 66, no. 11, pp. 5205–5219, Nov. 2018.
- [17] F. Guo, H. Lu, and Z. Gu, “Joint power and user grouping optimization in cell-free massive MIMO systems,” *IEEE Trans. Wireless Commun.*, early access, Aug. 4, 2021, doi: [10.1109/TWC.2021.3100573](https://doi.org/10.1109/TWC.2021.3100573).
- [18] W. Zeng, Y. He, B. Li, and S. Wang, “Pilot assignment for cell free massive MIMO systems using a weighted graphic framework,” *IEEE Trans. Veh. Technol.*, vol. 70, no. 6, pp. 6190–6194, Jun. 2021.
- [19] S. M. R. Islam, N. Avazov, O. A. Dobre, and K.-S. Kwak, “Power-domain non-orthogonal multiple access (NOMA) in 5G systems: Potentials and challenges,” *IEEE Commun. Surveys Tuts.*, vol. 19, no. 2, pp. 721–742, 2nd Quart., 2017.
- [20] Y. Li and G. A. A. Badige, “NOMA-aided cell-free massive MIMO systems,” *IEEE Wireless Commun. Lett.*, vol. 7, no. 6, pp. 950–953, Dec. 2018.
- [21] M. Bashar, K. Cumanan, A. G. Burr, H. Q. Ngo, L. Hanzo, and P. Xiao, “On the performance of cell-free massive MIMO relying on adaptive NOMA/OMA mode-switching,” *IEEE Trans. Commun.*, vol. 68, no. 2, pp. 792–810, Feb. 2020.
- [22] Y. Zhang, H. Cao, M. Zhou, and L. Yang, “Spectral efficiency maximization for uplink cell-free massive MIMO-NOMA networks,” in *Proc. IEEE Int. Conf. Commun. Workshops (ICC Workshops)*, May 2019, pp. 1–6.
- [23] T. K. Nguyen, H. H. Nguyen, and H. D. Tuan, “Max-min QoS power control in generalized cell-free massive MIMO-NOMA with optimal backhaul combining,” *IEEE Trans. Veh. Technol.*, vol. 69, no. 10, pp. 10949–10964, Oct. 2020.
- [24] F. Rezaei, A. R. Heidarpour, C. Tellambura, and A. Tadaion, “Underlaid spectrum sharing for cell-free massive MIMO-NOMA,” *IEEE Commun. Lett.*, vol. 24, no. 4, pp. 907–911, Apr. 2020.
- [25] F. Rezaei, C. Tellambura, A. A. Tadaion, and A. R. Heidarpour, “Rate analysis of cell-free massive MIMO-NOMA with three linear precoders,” *IEEE Trans. Commun.*, vol. 68, no. 6, pp. 3480–3494, Jun. 2020.
- [26] J. Zhang, J. Fan, B. Ai, and D. W. K. Ng, “NOMA-based cell-free massive MIMO over spatially correlated Rician fading channels,” in *Proc. IEEE Int. Conf. Commun.*, Jun. 2020, pp. 1–6.
- [27] S. Kusaladharma, W. P. Zhu, W. Ajib, and G. Amarasuriya, “Achievable rate analysis of NOMA in cell-free massive MIMO: A stochastic geometry approach,” in *Proc. IEEE Int. Conf. Commun. (ICC)*, May 2019, pp. 1–6.
- [28] A. A. Ohashi, D. B. D. Costa, A. L. P. Fernandes, W. Monteiro, R. Failache, A. M. Cavalcante, and J. C. W. A. Costa, “Cell-free massive MIMO-NOMA systems with imperfect SIC and non-reciprocal channels,” *IEEE Wireless Commun. Lett.*, vol. 10, no. 6, pp. 1329–1333, Jun. 2021.
- [29] S. Kusaladharma, W.-P. Zhu, W. Ajib, and G. A. A. Badige, “Achievable rate characterization of NOMA-aided cell-free massive MIMO with imperfect successive interference cancellation,” *IEEE Trans. Commun.*, vol. 69, no. 5, pp. 3054–3066, May 2021.
- [30] Z. Ding, M. Peng, and H. V. Poor, “Cooperative non-orthogonal multiple access in 5G systems,” *IEEE Commun. Lett.*, vol. 19, no. 8, pp. 1462–1465, Aug. 2015.
- [31] S. Solaiman, L. Nassef, and E. Fadel, “User clustering and optimized power allocation for D2D communications at mmWave underlaying MIMO-NOMA cellular networks,” *IEEE Access*, vol. 9, pp. 57726–57742, 2021.
- [32] A. Ashikhmin, L. Li, and T. L. Marzetta, “Interference reduction in multi-cell massive MIMO systems with large-scale fading precoding,” *IEEE Trans. Inf. Theory*, vol. 64, no. 9, pp. 6340–6361, Sep. 2018.

- [33] J. F. C. Kingman, "Stochastic models for random sets of points," in *Poisson Processes* (Oxford Studies in Probability), vol. 3. New York, NY, USA: Clarendon Press, 1993, pp. 1–10.
- [34] H. V. Cheng, E. Björnson, and E. G. Larsson, "Performance analysis of NOMA in training-based multiuser MIMO systems," *IEEE Trans. Wireless Commun.*, vol. 17, no. 1, pp. 372–385, Jan. 2018.
- [35] K. V. Prasad, "Fundamentals of statistical signal processing: Estimation theory," *Control Eng. Pract.*, vol. 2, no. 4, p. 728, Aug. 1994.
- [36] M. Khoshnevisan and J. N. Laneman, "Power allocation in multi-antenna wireless systems subject to simultaneous power constraints," *IEEE Trans. Commun.*, vol. 60, no. 12, pp. 3855–3864, Dec. 2012.
- [37] M. Médard, "The effect upon channel capacity in wireless communications of perfect and imperfect knowledge of the channel," *IEEE Trans. Inf. Theory*, vol. 46, no. 3, pp. 933–946, May 2000.
- [38] S. Zhang, J. Liu, H. Guo, M. Qi, and N. Kato, "Envisioning device-to-device communications in 6G," *IEEE Netw.*, vol. 34, no. 3, pp. 86–91, Jun. 2020.
- [39] A. Tang, J. Sun, and K. Gong, "Mobile propagation loss with a low base station antenna for NLOS street microcells in urban area," in *Proc. IEEE VTS 53rd Vehicular Technol. Conf.*, vol. 1, May 2001, pp. 333–336.
- [40] J.-M. Kang and I.-M. Kim, "Optimal user grouping for downlink NOMA," *IEEE Wireless Commun. Lett.*, vol. 7, no. 5, pp. 724–727, Oct. 2018.



REZA SAYYARI was born in Tabriz, Iran, in 1997. He received the B.Sc. degree in electrical engineering from the Islamic Azad University of Tabriz, Tabriz, in 2019. He is currently pursuing the M.Sc. degree with the Department of Electrical and Computer Engineering, University of Tabriz, Iran. His research interests include wireless communications and signal processing.



JAFAR POURROSTAM received the B.Sc. degree in electronic engineering from the Amirkabir University of Technology, Tehran, Iran, in 2000, the M.Sc. degree in communication systems engineering from the University of Tehran, Iran, in 2003, and the Ph.D. degree in electrical engineering from Michigan Technological University, USA, in 2007. He was with the Research and Development Department, PCTEL, a leading global provider of wireless technology solutions, from 2008 to 2013. He is currently an Assistant Professor at the Department of Electrical and Computer Engineering, University of Tabriz, Iran. His research interests include mobile communication systems and signal processing.



MIR JAVAD MUSEVI NIYA (Member, IEEE) was born in Ahar, Iran. He received the B.Sc. degree from the University of Tehran, the M.Sc. degree from the Sharif University of Technology, and the Ph.D. degree in communications from the University of Tabriz. Since 2006, he has been with the Faculty of Electrical and Computer Engineering, University of Tabriz. His current research interests include wireless communication systems, multimedia networks, and signal processing for communication systems.

• • •