A Novel Combining Method for NOMA and OMA in Cell-Free Massive MIMO System

Qiuyi Sui, Shaochuan Wu(✉), and Haoran Zhang

Harbin Institute of Technology, Harbin, China
scwu@hit.edu.cn

Abstract. Cell-Free massive MIMO (cf-mMIMO) is an incarnation of distributed massive MIMO, which provides high spectral efficieny (SE). However, when the number of users increases, the SE of cf-mMIMO will decrease. In order to increase the SE when the number of users is large, many researchers investigated cf-mMIMO with non-orthogonal multiple access (NOMA). In the existing research, cf-mMIMO with NOMA requires the number of users in each cluster to be equal, which is inflexible. This paper investigates the multiple access technology for cf-mMIMO. A more flexible scheme which uses the orthogonal multiple access (OMA) and NOMA at the same time is proposed in this paper. This scheme does not require the number of users in each cluster to be equal and it combines the advantages of both NOMA and OMA. An achievable SE of this scheme is derived. It is proved that under certain number of users and the length of the channel's coherence time conditions, using OMA and NOMA at the same time in the cf-mMIMO is able to obtain higher SE than using OMA or NOMA alone.

Keywords: Cell free massive MIMO · NOMA · OMA · Successive interference cancellation

1 Introduction

Massive MIMO is considered to be one of the key technologies in the fifth generation (5G). However, one of the limitations of it is that when the location of the users are at the edge of the cell, it will lead to low throughput, poor reliability, and increased communication delay [1, 2]. Therefore, a distributed, network-based massive MIMO, Cell-Free massive MIMO has been proposed and raised lots of concern. In cf-mMIMO, all access points (AP) serve all users simultaneously, so the problems mentioned above can be solved and higher SE can be provided [3, 4].

However, when the number of users is large, the length of the pilots increases, which will decrease the SE of cf-mMIMO. NOMA is a solution to this problem. As one of the most important key technologies in wireless communication, multiple access technology has always been a focus of research. As the number of users growing, the performance of NOMA is getting better [5, 6].

There are many researchers investigating cf-mMIMO with NOMA. The simulation results prove that when the number of users is large, the sum rate of cf-mMIMO with

Q. Guo et al. (Eds.): WiSATS 2021, LNICST 410, pp. 802–812, 2022.
https://doi.org/10.1007/978-3-030-93398-2_70

NOMA is higher than OMA. However, when the number of users is small, the performance of NOMA is worse than that of OMA due to intra-cluster interference and error propagation of imperfect SIC [7]. The number of users determine which technology provides higher sum rate. Hence, OMA should be adopted when the number of users is small, and switch to NOMA when the number of users is large. In [8] and [9], the authors discussed how to select the appropriate switching point and the specific switching method according to the actual situation and the number of users. Conjugate beamforming (CB) and normalized conjugate beamforming (NCB) is used to achieve spatial multiplexing. That is because they do not require APs to share channel state information (CSI) [3], which is very suitable for cf-mMIMO.

Although some researchers have considered the switching technique of OMA and NOMA, this switching technique cannot achieve ideal results near the switching point. Therefore, we combine NOMA and OMA, which means combining the advantages of them. Simulation results prove that this scheme provides higher SE under certain number of users and the length of the channel's coherence time conditions.

2 System, Channel and Signal Model

2.1 System and Channel Model

We let g_{mlk} denote the channel coefficient between the kth user in the lth cluster and mth AP. The channel coefficient g_{mlk} is modelled as follows:

$$g_{mlk} = \beta_{mlk}^{1/2} h_{mlk} \tag{1}$$

where h_{mlk} represents the small-scale fading, and $\beta_{mlk}^{1/2}$ represents the large-scale fading. Because each AP and user is distributed in a large area discretely, $h_{mlk}, m = 1, 2, \ldots M, k = 1, 2, \ldots, K$ is independent and identically distributed (i.i.d.).

In cf-mMIMO, the entire data transmission process is divided into three stages: uplink channel estimation, uplink transmission and downlink transmission. In uplink channel estimation stage, the users send pilots to APs, and APs obtain CSI using received pilot signals. The users do not know the accurate channel state information, but only know the statistical information of channels. As the number of APs grows sufficiently large, the underlying channels harden, which means the statistical CSI is fixed in several coherence intervals. Therefore, it is sufficient for users to decode.

Let τ be the length of the channel's coherence time, which is equal to the product of the channel's coherence time and coherence bandwidth. In this paper, τ is defined as the number of samples for each coherence interval. The coherence time depends on the users' moving speed. τ_p is the length of pilot for each user. τ_u and τ_d are the uplink and downlink transmission time respectively. Obviously, $\tau = \tau_p + \tau_u + \tau_d$. In this paper, we assume that the uplink and downlink transmission time is equal, that is $\tau = \tau_p + 2\tau_d$.

The model of cf-mMIMO with NOMA and OMA is shown in the Fig. 1. The model of cf-mMIMO with NOMA and OMA.

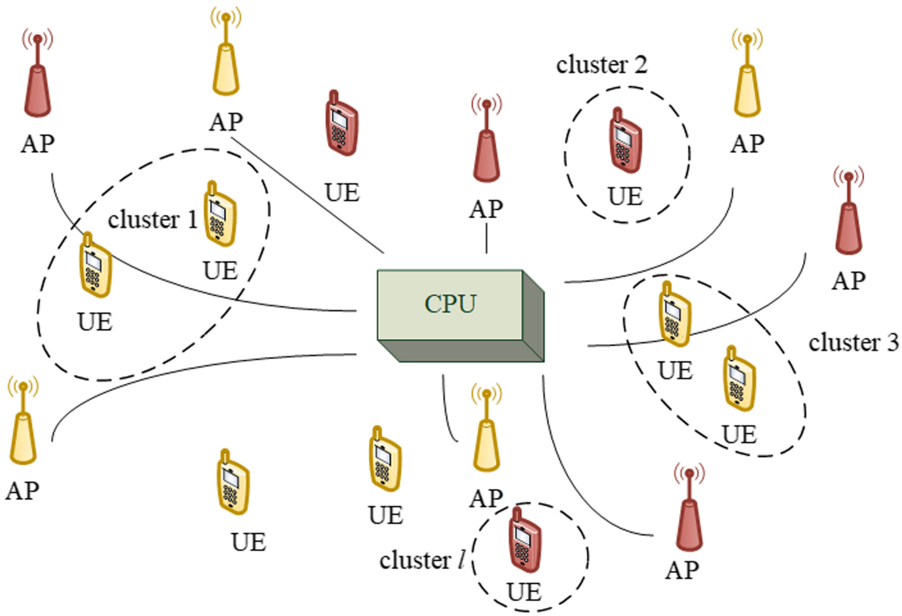

Fig. 1. The model of cf-mMIMO with NOMA and OMA. (Color figure online)

As Fig. 1 shows, all APs and user equipments (UEs) are randomly distributed. Each AP has N antennas, and each UE has 1 antenna. All APs and users are divided into 2 parts. The first part of users uses NOMA (yellow UEs). There are more than one user in each cluster, that is $k_l > 1$. All k users in one cluster use the same pilot. The set of these clusters is represented by Ω_{N}. Some APs serve them and the set of these APs is represented by Φ_{N}. The other part of users uses OMA (red UEs). There is only one user in each cluster, that is $k_l = 1$. The set of these clusters is represented by Ω_{O}. The other APs serve them and the set of these APs is represented by Φ_{O}.

Suppose that users in the system are divided into L_{total} clusters. The number of NOMA clusters is L_1, and the number of users in each cluster is K_1 ($K_1 > 1$). The number of OMA clusters is L_2, and the number of users in each cluster is K_2 ($K_2 = 1$). Then, the total number of users in the system is $K_{total} = K_1 L_1 + K_2 L_2$. There are M APs. The numbers of APs serving NOMA users and OMA users are M_1 and M_2 respectively.

In the following paper, L is used to represent the number of NOMA or OMA clusters, and K is used to represent the number of users in the cluster. When $m \in \Phi_{\mathrm{N}}$ or $l \in \Omega_{\mathrm{N}}$, there are $L = L_1$ and $K = K_1$. When $m \in \Phi_{\mathrm{O}}$ or $l \in \Omega_{\mathrm{O}}$, there are $L = L_2$ and $K = K_2$.

2.2 Signal Model

In the channel estimation stage, all users send pilots to the APs. The users in one cluster use the same pilot and the users in different clusters use orthogonal pilots. The pilot sequence allocated for lth cluster is denoted as $\boldsymbol{\varphi}_l \in \mathbb{C}^{\tau_p \times 1}$, satisfying $\|\boldsymbol{\varphi}_l\|^2 = 1$.

The received signal of mth AP can be written as

$$\mathbf{y}_m = \sqrt{\tau_p \rho_p} \sum_{l=1}^{L_{total}} \sum_{k=1}^{K} g_{mlk} \boldsymbol{\varphi}_l + \mathbf{n}_m \quad (2)$$

where ρ_p is the pilot transmit power and $\mathbf{n}_m$ is additive noise at the mth AP.

Although in [3], the uplink pilot is used to estimate the channel gain of each user, estimating a linear combination of the users' channel in the same cluster provides a better performance according to [10]. Hence, we estimate the linear combination of the users' channel.

$$f_{ml} = \sum_{k=1}^{K} g_{mlk}, \forall m, \forall l \quad (3)$$

The received signal at mth AP is projected onto $\boldsymbol{\varphi}_l$ as

$$\breve{\mathbf{y}}_{ml} = \boldsymbol{\varphi}_l^H \mathbf{y}_m = \sqrt{\tau_p \rho_p} f_{ml} + \boldsymbol{\varphi}_l^H \mathbf{n}_\mathbf{m} \quad (4)$$

The MMSE estimate of f_{ml} given $\breve{\mathbf{y}}_{ml}$ is written as

$$\hat{f}_{ml} = c_{ml} \breve{\mathbf{y}}_{ml} \quad (5)$$

where $c_{ml} = \frac{\sqrt{\tau_p \rho_p} \sum_{k'=1}^{K} \beta_{mlk'}}{\tau_p \rho_p \sum_{k'=1}^{K} \beta_{mlk'} + 1}$.

The downlink transmission uses conjugate beamforming (CB), so the transmit signal at the mth AP can be expressed as

$$x_m = \sqrt{\rho_d} \sum_{l=1}^{L} \sum_{k=1}^{K} \sqrt{\eta_{mlk}} \hat{f}_{ml}^* s_{lk} \quad (6)$$

where η_{mlk} is the power allocated to the kth UE in the lth cluster at the mth AP, s_{lk} ($\mathrm{E}\left\{|s_{lk}|^2\right\} = 1$) is the transmit symbol, and ρ_d is the downlink transmit power.

Then the normalized transmit power of the mth AP can be derived as

$$\mathrm{E}\{|x_m|^2\} = \rho_d \sum_{l=1}^{L} \sum_{k=1}^{K} \eta_{mlk} \gamma_{ml} \quad (7)$$

where $\gamma_{ml} = \mathrm{E}\left\{ |f_{ml}|^2 \right\} = \frac{\tau_p \rho_p \left(\sum_{k'=1}^{K} \beta_{mlk'} \right)}{\tau_p \rho_p \sum_{k'=1}^{K} \beta_{mlk'} + 1}$.

Each AP needs to satisfy the power constraint:

$$\sum_{l=1}^{L} \sum_{k=1}^{K} \eta_{mlk} \gamma_{ml} \le \frac{1}{N}, \forall m \tag{8}$$

Suppose that in the lth cluster, user UE-*11* is the user with the best channel, and user UE-*1K* is the user with the worst channel.

The received signal of NOMA and OMA at the kth user in the lth cluster is given by (9) and (10) respectively.

$$\begin{aligned} y_{lk.\mathrm{N}} &= \sum_{m=1}^{M} g_{mlk} x_m + n_{lk} \\ &= \sqrt{\rho_d} \sum_{m \in \Phi_\mathrm{N}} g_{mlk} \sqrt{\eta_{mlk}} \hat{f}_{ml}^* s_{lk} + \sqrt{\rho_d} \sum_{m \in \Phi_\mathrm{N}} \sum_{k'=1, k' \ne k}^{K_1} g_{mlk} \sqrt{\eta_{mlk'}} \hat{f}_{ml}^* s_{lk'} \\ &+ \sqrt{\rho_d} \sum_{m \in \Phi_\mathrm{N}} \sum_{l' \ne l}^{L_1} \sum_{k'=1}^{K_1} g_{mlk} \sqrt{\eta_{ml'k'}} \hat{f}_{ml'}^* s_{l'k'} + \sum_{m \notin \Phi_\mathrm{N}} \sum_{l' \ne l}^{L_2} \sum_{k'=1}^{K_2} g_{mlk} \sqrt{\eta_{ml'k'}} \hat{f}_{ml'}^* s_{l'k'} + n_{lk} \end{aligned} \tag{9}$$

$$\begin{aligned} y_{lk.\mathrm{O}} &= \sum_{m=1}^{M} g_{mlk} x_m + n_{lk} \\ &= \sqrt{\rho_d} \sum_{m \in \Phi_\mathrm{O}} g_{mlk} \sqrt{\eta_{mlk}} \hat{f}_{ml}^* s_{lk} + \sqrt{\rho_d} \sum_{m \in \Phi_\mathrm{O}} \sum_{k'=1, k' \ne k}^{K_2} g_{mlk} \sqrt{\eta_{mlk'}} \hat{f}_{ml}^* s_{lk'} \\ &+ \sqrt{\rho_d} \sum_{m \in \Phi_\mathrm{O}} \sum_{l' \ne l}^{L_2} \sum_{k'=1}^{K_2} g_{mlk} \sqrt{\eta_{ml'k'}} \hat{f}_{ml'}^* s_{l'k'} + \sum_{m \notin \Phi_\mathrm{O}} \sum_{l' \ne l}^{L_1} \sum_{k'=1}^{K_1} g_{mlk} \sqrt{\eta_{ml'k'}} \hat{f}_{ml'}^* s_{l'k'} + n_{lk} \end{aligned} \tag{10}$$

The first term in the above expression is the desired signal, the second term is the interference caused by other users in the cluster, the third term is the interference caused by other clusters, the fourth term is the interference caused by OMA/NOMA users, and the fifth term is noise.

The users are assumed to be ordered to satisfy [11]:

$$\mathrm{E}\{\log_2(1 + \mathrm{SINR}_{lj}^{lk})\} \ge \mathrm{E}\{\log_2(1 + \mathrm{SINR}_{lk}^{lk})\}, \forall j < k, \forall l \tag{11}$$

where SINR_{lj}^{lk} refers to the effective SINR of the jth user in the lth cluster when the jth user in lth cluster is decoding the signal intended for the kth user in the same cluster.

Finally, the achievable rate of the kth user in the lth cluster can be expressed as

$$R_{lk}^{lk,final} = \min(\mathrm{E}\{log_2(1 + \mathrm{SINR}_{lj}^{lk})\}, E\{log_2(1 + \mathrm{SINR}_{lk}^{lk})\}), \forall l, k \tag{12}$$

3 Achievable Spectral Efficieny

The signal actually used for decoding after SIC for NOMA and OMA can be written as (13) and (14) respectively.

$$\begin{aligned}
r_{lk,\mathrm{N}}^{lk} &= r_{lk} - \sqrt{\rho_d}\sum_{k''=k+1}^{K_1}\mathrm{E}\{\sum_{m\in\Phi_\mathrm{N}}\sqrt{\eta_{mlk''}}g_{mlk}^T\hat{f}_{ml}^*\}s_{lk''} \\
&= \sqrt{\rho_d}\sum_{m=1}^{M}\sum_{l'=1}^{L}\sum_{k'=1}^{K}\sqrt{\eta_{ml'k'}}g_{mlk}\hat{f}_{ml}^*s_{l'k'} + n_{lk} - \sqrt{\rho_d}\sum_{k''=k+1}^{K_1}\mathrm{E}\{\sum_{m\in\Phi_\mathrm{N}}\sqrt{\eta_{mlk''}}g_{mlk}\hat{f}_{ml}^*\}s_{lk''} \\
&= \underbrace{\sqrt{\rho_d}\mathrm{E}\{\sum_{m\in\Phi_\mathrm{N}}\sqrt{\eta_{mlk}}g_{mlk}\hat{f}_{ml}^*\}s_{lk}}_{\mathrm{DS}_{lk}} + \underbrace{\sqrt{\rho_d}(\sum_{m\in\Phi_\mathrm{N}}\sqrt{\eta_{mlk}}g_{mlk}\hat{f}_{ml}^* - \mathrm{E}\{\sum_{m\in\Phi_\mathrm{N}}\sqrt{\eta_{mlk}}g_{mlk}\hat{f}_{ml}^*\})s_{lk}}_{\mathrm{BU}_{lk}} \\
&+ \underbrace{\sum_{k'\neq k}^{k-1}\sqrt{\rho_d}\sum_{m\in\Phi_\mathrm{N}}\sqrt{\eta_{mlk'}}g_{mlk}\hat{f}_{ml}^*s_{lk'}}_{\mathrm{IUI}_{lk'}} + \underbrace{\sum_{k''=k+1}^{K_1}\sqrt{\rho_d}(\sum_{m\in\Phi_\mathrm{N}}\sqrt{\eta_{mlk''}}g_{mlk}\hat{f}_{ml}^* - \mathrm{E}\{\sum_{m\in\Phi_\mathrm{N}}\sqrt{\eta_{mlk''}}g_{mlk}\hat{f}_{ml}^*\})s_{lk''}}_{\mathrm{ISIC}_{lk''}} \\
&+ \underbrace{\sum_{l'\neq l}^{L_1}\sum_{k'=1}^{K_1}\sqrt{\rho_d}\sum_{m\in\Phi_\mathrm{N}}\sqrt{\eta_{ml'k'}}g_{mlk}\hat{f}_{ml'}^*s_{l'k'}}_{\mathrm{ICI}_{lk''}} + \underbrace{\sum_{m\notin\Phi_\mathrm{N}}\sum_{l'=1}^{L_1}\sum_{k'=1}^{K_1}\sqrt{\rho_d}\sqrt{\eta_{ml'k'}}g_{mlk}\hat{f}_{ml'}^*s_{l'k'}}_{\mathrm{OUI}_{l'k'}} + n_{lk}
\end{aligned} \tag{13}$$

$$\begin{aligned}
r_{lk,\mathrm{O}}^{lk} &= r_{lk} - \sqrt{\rho_d}\sum_{k''=k+1}^{K_2}\mathrm{E}\{\sum_{m\in\Phi_\mathrm{O}}\sqrt{\eta_{mlk''}}g_{mlk}\hat{f}_{ml}^*\}s_{lk''} \\
&= \sqrt{\rho_d}\sum_{m=1}^{M}\sum_{l'=1}^{L}\sum_{k'=1}^{K}\sqrt{\eta_{ml'k'}}g_{mlk}\hat{f}_{ml}^*s_{l'k'} + n_{lk} - \sqrt{\rho_d}\sum_{k''=k+1}^{K_2}\mathrm{E}\{\sum_{m\in\Phi_\mathrm{O}}\sqrt{\eta_{mlk''}}g_{mlk}\hat{f}_{ml}^*\}s_{lk''} \\
&= \underbrace{\sqrt{\rho_d}\mathrm{E}\{\sum_{m\in\Phi_\mathrm{O}}\sqrt{\eta_{mlk}}g_{mlk}\hat{f}_{ml}^*\}s_{lk}}_{\mathrm{DS}_{lk}} + \underbrace{\sqrt{\rho_d}(\sum_{m\in\Phi_\mathrm{O}}\sqrt{\eta_{mlk}}g_{mlk}\hat{f}_{ml}^* - \mathrm{E}\{\sum_{m\in\Phi_\mathrm{O}}\sqrt{\eta_{mlk}}g_{mlk}\hat{f}_{ml}^*\})s_{lk}}_{\mathrm{BU}_{lk}} \\
&+ \underbrace{\sum_{k'\neq k}^{k-1}\sqrt{\rho_d}\sum_{m\in\Phi_\mathrm{O}}\sqrt{\eta_{mlk'}}g_{mlk}\hat{f}_{ml}^*s_{lk'}}_{\mathrm{IUI}_{lk'}} + \underbrace{\sum_{k''=k+1}^{K_2}\sqrt{\rho_d}(\sum_{m\in\Phi_\mathrm{O}}\sqrt{\eta_{mlk''}}g_{mlk}\hat{f}_{ml}^* - \mathrm{E}\{\sum_{m\in\Phi_\mathrm{O}}\sqrt{\eta_{mlk''}}g_{mlk}\hat{f}_{ml}^*\})s_{lk''}}_{\mathrm{ISIC}_{lk''}} \\
&+ \underbrace{\sum_{l'\neq l}^{L_2}\sum_{k'=1}^{K_2}\sqrt{\rho_d}\sum_{m\in\Phi_\mathrm{O}}\sqrt{\eta_{ml'k'}}g_{mlk}\hat{f}_{ml'}^*s_{l'k'}}_{\mathrm{ICI}_{lk''}} + \underbrace{\sum_{m\notin\Phi_\mathrm{O}}\sum_{l'=1}^{L_1}\sum_{k'=1}^{K_1}\sqrt{\rho_d}\sqrt{\eta_{ml'k'}}g_{mlk}\hat{f}_{ml'}^*s_{l'k'}}_{\mathrm{NUI}_{l'k'}} + n_{lk}
\end{aligned} \tag{14}$$

In (13) and (14), DS_{lk} and BU_{lk} represent the desired signal and beamforming uncertainty for the kth UE in the lth cluster, and $\mathrm{IUI}_{lk'}$ represents the inter-user-interference imposed by the k' th UE in the lth cluster, $\mathrm{ISIC}_{lk''}$ represents the interference caused by imperfect successive interference cancellation, $\mathrm{ICI}_{l'k'}$ is the inter-cluster-interference, and $\mathrm{OUI}_{l'k'}$ and $\mathrm{NUI}_{l'k'}$ represent the OMA-user-interference and NOMA-user-interference respectively.

The SINR can be derived. The specific derivation process is the same as appendix A in [9].

$$\mathrm{SINR}_{lk,\mathrm{N}}^{lk} = \frac{N^2\left(\sum\limits_{m\in\Phi_{\mathrm{N}}}\sqrt{\eta_{mlk}}\frac{\gamma_{ml}\beta_{mlk}}{\sum\limits_{i=1}^{K_1}\beta_{mli}}\right)^2}{N^2\sum\limits_{k'=1}^{k-1}\left(\sum\limits_{m\in\Phi_{\mathrm{N}}}\sqrt{\eta_{mlk'}}\frac{\gamma_{ml}\beta_{mlk}}{\sum\limits_{i=1}^{K_1}\beta_{mli}}\right)^2 + N\sum\limits_{l'=1}^{L_1}\sum\limits_{k'=1}^{K_1}\sum\limits_{m\in\Phi_{\mathrm{N}}}\eta_{ml'k'}\beta_{mlk}\gamma_{ml'} + \frac{1}{\rho_d}} \tag{15}$$

$$\mathrm{SINR}_{lj,\mathrm{N}}^{lk} = \frac{N^2\left(\sum\limits_{m\in\Phi_{\mathrm{N}}}\sqrt{\eta_{mlk}}\frac{\gamma_{ml}\beta_{mlj}}{\sum\limits_{i=1}^{K_1}\beta_{mli}}\right)^2}{N^2\sum\limits_{k'=1}^{k-1}\left(\sum\limits_{m\in\Phi_{\mathrm{N}}}\sqrt{\eta_{mlk'}}\frac{\gamma_{ml}\beta_{mlj}}{\sum\limits_{i=1}^{K_1}\beta_{mli}}\right)^2 + N\sum\limits_{l'=1}^{L_1}\sum\limits_{k'=1}^{K_1}\sum\limits_{m\in\Phi_{\mathrm{N}}}\eta_{ml'k'}\beta_{mlj}\gamma_{ml'} + \frac{1}{\rho_d}} \tag{16}$$

$$\mathrm{SINR}_{lk,\mathrm{O}}^{lk} = \frac{N^2\left(\sum\limits_{m\in\Phi_{\mathrm{O}}}\sqrt{\eta_{mlk}}\frac{\gamma_{ml}\beta_{mlk}}{\sum\limits_{i=1}^{K_2}\beta_{mli}}\right)^2}{N^2\sum\limits_{k'=1}^{k-1}\left(\sum\limits_{m\in\Phi_{\mathrm{O}}}\sqrt{\eta_{mlk'}}\frac{\gamma_{ml}\beta_{mlk}}{\sum\limits_{i=1}^{K_2}\beta_{mli}}\right)^2 + N\sum\limits_{l'=1}^{L_2}\sum\limits_{k'=1}^{K_2}\sum\limits_{m\in\Phi_{\mathrm{O}}}\eta_{ml'k'}\beta_{mlk}\gamma_{ml'} + \frac{1}{\rho_d}} \tag{17}$$

$$\mathrm{SINR}_{lk,\mathrm{O}}^{lk} = \frac{N^2\left(\sum\limits_{m\in\Phi_{\mathrm{O}}}\sqrt{\eta_{mlk}}\frac{\gamma_{ml}\beta_{mlj}}{\sum\limits_{i=1}^{K_2}\beta_{mli}}\right)^2}{N^2\sum\limits_{k'=1}^{k-1}\left(\sum\limits_{m\in\Phi_{\mathrm{O}}}\sqrt{\eta_{mlk'}}\frac{\gamma_{ml}\beta_{mlj}}{\sum\limits_{i=1}^{K_2}\beta_{mli}}\right)^2 + N\sum\limits_{l'=1}^{L_2}\sum\limits_{k'=1}^{K_2}\sum\limits_{m\in\Phi_{\mathrm{O}}}\eta_{ml'k'}\beta_{mlj}\gamma_{ml'} + \frac{1}{\rho_d}} \tag{18}$$

According to the expressions (15) and (16), we use $\tilde{\mathbf{g}}_{lk} = [\gamma_{1l}w_{1lk}, \gamma_{2l}w_{2lk}, \ldots, \gamma_{Ml}w_{Mlk}]^T, \forall l, k$ to sort users, where $w_{mlk} = \frac{\beta_{mlk}}{\sum_{i=1}^{K}\beta_{1li}}$. After sorting, we derive $|\tilde{\mathbf{g}}_{l1}|^2 \geq |\tilde{\mathbf{g}}_{l2}|^2 \geq \ldots \geq |\tilde{\mathbf{g}}_{lK}|^2$.

The final downlink SE of single user for NOMA and OMA is derived as

$$S_{lk}^{lk,final} = \frac{1}{2}\left(1 - \frac{\tau_p}{\tau}\right)\log_2\left(1 + \mathrm{SINR}_{lk}^{lk,final}\right) \tag{19}$$

where $\mathrm{SINR}_{lk}^{lk,final} = \min\left(\mathrm{SINR}_{lj}^{lk}, \mathrm{SINR}_{lk}^{lk}\right), \forall l, k.$

4 Power Allocation

We separately allocate power to NOMA and OMA users by using max-min SINR algorithm [9].

$$\mathrm{P1}: \max_{\eta_{mlk}} \min_{k=1\ldots K, l=1\ldots L} SINR_{lk}^{lk,final}$$

$$\text{s.t.} \sum_{l=1}^{L}\sum_{k=1}^{K}\eta_{mlk}\gamma_{mlk} \leq \frac{1}{N}, \forall m,\ \eta_{mlk} \geq 0, \forall m, \forall l, \forall k \tag{20}$$

Define $\varsigma_{mlk} = \sqrt{\eta_{mlk}}$ and introduce slack variables $\upsilon_m, \tilde{\upsilon}_m, \lambda_{lk'j}, \tilde{\lambda}_{lk'k}$. We reformulate (20) as follows:

$$\mathrm{P2}: \min_{\left\{\varsigma_{mlk}, \lambda_{lk'j}, \upsilon_m\right\}} \sum_{m=1}^{M}\sum_{l'=1}^{L}\sum_{k'=1}^{K}\gamma_{ml'}\varsigma_{ml'k'}^2$$

$$\text{s.t.} \left\|\mathbf{z}_{lj}\right\| \leq \frac{N\sum_{m=1}^{M}\varsigma_{mlk}\frac{\gamma_{ml}\beta_{mlj}}{\sum_{i=1}^{K}\beta_{mli}}}{\sqrt{t}}, \forall j < k,\ \left\|\tilde{\mathbf{z}}_{lk}\right\| \leq \frac{N\sum_{m=1}^{M}\varsigma_{mlk}\frac{\gamma_{ml}\beta_{mlk}}{\sum_{i=1}^{K}\beta_{mli}}}{\sqrt{t}}$$

$$\sum_{l'\neq l}^{L}\sum_{k'=1}^{K}\gamma_{ml}\varsigma_{ml'k'}^2 + \sum_{k'=1}^{K}\gamma_{ml}\varsigma_{mlk'}^2 \leq \upsilon_m^2, \forall m, \forall j < k, 0 \leq \upsilon_m \leq \frac{1}{\sqrt{N}}, \forall m$$

$$\sum_{m=1}^{M}\varsigma_{mlk'}\frac{\gamma_{ml}\beta_{mlj}}{\sum_{i=1}^{K}\beta_{mli}} \leq \lambda_{lk'j}, 1 \leq k' \leq k-1, \forall j < k$$

$$\left(\sum_{l'\neq l}^{L}\sum_{k'=1}^{K}\gamma_{ml'}\varsigma_{ml'k'}^{2}+\sum_{k'=1}^{K}\gamma_{ml}\varsigma_{mlk'}^{2}\right)\leq\tilde{\upsilon}_{m}^{2},\forall m,0\leq\upsilon_{m}\leq\frac{1}{\sqrt{N}},\forall m$$

$$\sum_{m=1}^{M}\varsigma_{mlk'}\frac{\gamma_{ml}\beta_{mlk}}{\sum_{i=1}^{K}\beta_{mli}}\leq\tilde{\lambda}_{lk'k},1\leq k'\leq k-1$$

$$\varsigma_{mlk}\geq 0,\forall m,\forall l,\forall k,\sum_{l=1}^{L}\sum_{k=1}^{K}\varsigma_{mlk}^{2}\gamma_{ml}\leq\frac{1}{N},\forall m \tag{21}$$

where $\left\|\mathbf{z}_{lj}\right\|\triangleq\left[N\mathbf{v}_{lj,1}^{T}\quad\sqrt{N}\mathbf{v}_{lj,2}^{T}\quad\frac{1}{\sqrt{\rho_d}}\right]^{T}$, $\|\tilde{\mathbf{z}}_{lk}\|\triangleq\left[N\tilde{\mathbf{v}}_{lk,1}^{T}\quad\sqrt{N}\tilde{\mathbf{v}}_{lk,2}^{T}\quad\frac{1}{\sqrt{\rho_d}}\right]^{T}$, $\mathbf{v}_{lj,1}=\left[\lambda_{l1j}\quad\ldots\quad\lambda_{l(k-1)j}\right]^{T}$, $\mathbf{v}_{lj,2}=\left[\sqrt{\beta_{1lj}}v_1\quad\ldots\quad\sqrt{\beta_{Mlj}}v_M\right]^{T}$, $\tilde{\mathbf{v}}_{lk,1}=\left[\tilde{\lambda}_{l1k}\quad\ldots\quad\tilde{\lambda}_{l(k-1)k}\right]^{T}$, and $\tilde{\mathbf{v}}_{lk,2}=\left[\sqrt{\beta_{1lk}}\tilde{v}_1\quad\ldots\quad\sqrt{\beta_{Mlk}}\tilde{v}_M\right]^{T}$. Using bisection search algorithm shown in Table 1 can solve P2.

Table 1. Bisection search algorithm

Initialize $t_{\max}$, $t_{\min}$ and ε
Do while
Set $t=\frac{t_{\max}+t_{\min}}{2}$ and solve P2
If P2 is feasible, set $t_{\min}=t$
else, set $t_{\max}=t$
until $(t_{\max}-t_{\min})<\varepsilon$

5 Numerical Results

The simulation parameters are shown in Table 2.

Table 2. The simulation parameters for cf-mMIMO with OMA and NOMA

Simulation parameters	Parameters configuration
Carrier frequency	1.9 GHz
Bandwidth	20 MHz
Noise figure	9 dB
Downlink transmit power	200 mW
Pilot transmit power	100 mW
The number of antenna on each AP	10
The standard deviation of shadow fading	8 dB

There are 40 APs. 20 APs are randomly selected to serve NOMA users and the other 20 APs serve OMA users. There are 20 users in total. 10 users are randomly selected to use NOMA and the other 10 users use OMA. The result is shown in Fig. 2.

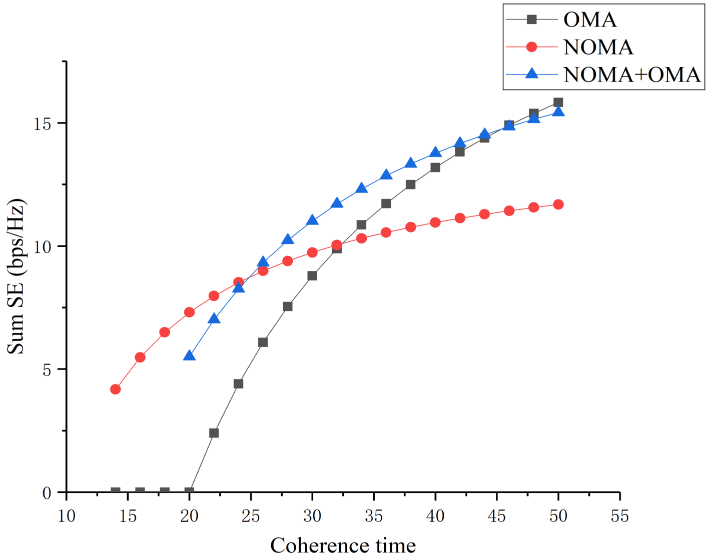

Fig. 2. Comparison of SE among NOMA, OMA and NOMA combined with OMA.

As shown in Fig. 2, using OMA and NOMA at the same time in the cf-mMIMO is able to obtain higher sum SE than using OMA or NOMA alone under certain number of users and the length of the channel's coherence time conditions.

The advantage of OMA is that there is no interference caused by the users in the same cluster and imperfect SIC. Therefore, the sum rate of OMA is higher than NOMA. However, OMA requires a large number of orthogonal pilots, which means the length of the pilot sequences is large. It takes a long time to send the pilots, resulting in a shorter time that can be used for uplink and downlink data transmission, which will seriously affect SE when the coherence time is short. The advantage of NOMA is that fewer orthogonal pilots are required, which means the length of pilot sequences can be shorter. However, there is interference caused by the users in the same cluster and imperfect SIC. When the coherence time is long, the SE of NOMA is lower than OMA. Additionally, in cf-mMIMO with NOMA scheme, the number of users in each cluster must be equal (usually 2), which is inflexible.

Using OMA and NOMA at the same time is a compromise. When the coherence time is not too long or too short, cf-mMIMO with OMA and NOMA combine the advantages of OMA and NOMA. Some users use NOMA to shorten the length of pilot sequences, and the other users use OMA to get a higher sum rate. In addition, it does not require the number of users in each cluster to be equal. However, when the coherence time is too short, the length of pilots is greater than that of NOMA, make the SE of it lower than NOMA. And when the coherence time is too long, the length of pilots has little effect, but the interference caused by NOMA users decreases the sum rate and SE. Therefore, cf-mMIMO with OMA achieves the best performance.

6 Conclusion

This paper studies the multiple access technology in cf-mMIMO, and proposes a cf-mMIMO with NOMA and OMA scheme. Unlike cf-mMIMO with NOMA, this scheme does not require the number of users in each cluster to be equal. The simulation results prove that under certain number of users and the length of the channel's coherence time conditions, using OMA and NOMA at the same time in the cf-mMIMO is able to obtain higher sum SE than using OMA or NOMA alone.

Acknowledgements. This research is supported by the National Key R&D Program of China (Under Grant: 2018YFC0806803) and the National Science Foundation of China (Under Grant: 61671173).

References

1. Xu, Y., Yue, G., Mao, S.: User grouping for massive MIMO in FDD systems: new design methods and analysis. IEEE Trans. Wirel. Commun. **14**(12), 6827–6842 (2014)
2. Bashar, M., Ngo, H.Q., Burr, A.G., Maryopi, D., Cumanan, K., Larsson, E.G.: On the performance of backhaul constrained cell-free massive MIMO with linear receivers. In: 52nd Asilomar Conference on Signals, Systems, and Computers, pp. 624–628 (2018)
3. Ngo, H.Q., Ashikhmin, A., Yang, H., Larsson, E.G., Marzetta, T.L.: Cell-free massive MIMO versus small cells. IEEE Trans. Wirel. Commun. **16**(3), 1834–1850 (2017)
4. Nguyen, L.D., Duong, T.Q., Ngo, H.Q., Tourki, K.: Energy efficiency in cell-free massive MIMO with zero-forcing precoding design. IEEE Commun. Lett. **21**(8), 1871–1874 (2017)
5. Xiang, L., Ng, D.W.K., Ge, X., Ding, Z., Wong, V.W.S., Schober, R.: Cache-aided non-orthogonal multiple access: the two-user case. IEEE J. Sel. Top. Signal Process. **13**(3), 436–451 (2019)
6. Saito, Y., Kishiyama, Y., Benjebbour, A., Nakamura, T., Li, A., Higuchi, K.: Non-orthogonal multiple access (NOMA) for cellular future radio access. In: IEEE 77th Vehicular Technology Conference (VTC Spring), pp. 1–5 (2013)
7. Li, Y., Aruma Baduge, G.A.: NOMA-aided cell-free massive MIMO systems. IEEE Commun. Lett. **7**(6), 950–953 (2018)
8. Bashar, M., Cumanan, K., Burr, A.G., Ngo, H.Q., Hanzo, L., Xiao, P.: NOMA/OMA mode selection-based cell-free massive MIMO. In: International Conference on Communications (ICC), pp. 1–6 (2019)
9. Bashar, M., Cumanan, K., Burr, A.G., Ngo, H.Q., Hanzo, L., Xiao, P.: On the performance of Cell-Free massive MIMO relying on adaptive NOMA/OMA mode-switching. IEEE Trans. Commun. **2**(68), 792–810 (2020)
10. Cheng, H.V., Björnson, E., Larsson, E.G.: Performance analysis of NOMA in training-based multiuser MIMO systems. IEEE Trans. Wireless Commun. **17**(1), 372–385 (2018)
11. Hanif, M.F., Ding, Z., Ratnarajah, T., Karagiannidis, G.K.: A minorization-maximization method for optimizing sum rate in the downlink of non-orthogonal multiple access systems. IEEE Trans. Signal Process. **64**(1), 76–88 (2016)