On the Performance of Cell-Free Massive MIMO Relying on Adaptive NOMA/OMA Mode-Switching

On the Performance of Cell-Free Massive MIMO Relying on Adaptive NOMA/OMA Mode-Switching

Manijeh Bashar, *Student Member, IEEE*, Kanapathippillai Cumanan, *Senior Member, IEEE*, Alister G. Burr, *Senior Member, IEEE*, Hien Quoc Ngo, *Member, IEEE*, Lajos Hanzo, *Fellow, IEEE*, and Pei Xiao, *Senior Member, IEEE*

***Abstract*—The downlink (DL) of a non-orthogonal-multiple-access (NOMA)-based cell-free massive multiple-input multiple-output (MIMO) system is analyzed, where the channel state information (CSI) is estimated using pilots. It is assumed that the users are grouped into multiple clusters. The same pilot sequences are assigned to the users within the same clusters whereas the pilots allocated to all clusters are mutually orthogonal. First, a user's bandwidth efficiency (BE) is derived based on his/her channel statistics under the assumption of employing successive interference cancellation (SIC) at the users' end with no DL training. Next, the classic max-min optimization framework is invoked for maximizing the minimum BE of a user under per-access point (AP) power constraints. The max-min user BE of NOMA-based cell-free massive MIMO is compared to that of its orthogonal multiple-access (OMA) counter part, where all users employ orthogonal pilots. Finally, our numerical results are presented and an operating mode switching scheme is proposed based on the average per-user BE of the system, where the mode set is given by Mode = { OMA, NOMA }. Our numerical results confirm that the switching point between the NOMA and OMA modes depends both on the length of the channel's coherence time and on the total number of users.**

Keywords: **Cell-free massive MIMO, convex optimization, max-min bandwidth efficiency, NOMA.**

I. Introduction

Non-orthogonal-multiple-access (NOMA) is capable of significantly increasing the bandwidth efficiency (BE), hence contributing to the improvements required in fifth generation (5G) networks over the fourth generation (4G) [2]–[5]. By contrast, in orthogonal-multiple-access (OMA), orthogonal resources such as time, frequency, spreading codes and pilots are assigned to different users for avoiding inter-user interference at a low complexity [6]. However, NOMA relying on power domain multiplexing and successive interference cancellation (SIC) at the receivers is designed to support multiple users in the same resources [7]–[9].

On the other hand, massive multiple-input multiple-output (MIMO) is a key technology for next-generation systems because of the improvement in BE it provides [10]–[12]. However, the low throughput of cell-edge users remains a limitation in realistic multi-cell massive MIMO systems [13], [14]. In cell-free massive MIMO, on the other hand, distributed access points (APs) are connected to a central processing unit (CPU) and jointly serve distributed users [15]–[19]. This approach reaps many of the benefits of cloud radio access network (C-RAN) such as a low pathloss as well as distributed signal processing. Cell-free massive MIMO and C-RAN may be viewed as scalable versions of the network-MIMO concept, or coordinated multipoint processing (CoMP) [20]–[24], both of which depart from the concept of conventional cells. In [25]–[28], the authors show that exploiting optimal uniform quantization and wireless microwave links with capacity 100 Mbits/s, the performance of limited-backhaul cell-free massive MIMO system closely approaches the performance of cell-free massive MIMO with perfect backhaul links. Cell-free massive MIMO may find its way into next-generation networks [29]. In [30], Marzetta characterized the performance of massive MIMO systems in the context of time division duplexing (TDD), which has widely inspired the community [31], which is capable of outperforming frequency division duplexing (FDD). Since the channel's coherence time is short, the users have to resort to non-orthogonal pilot sequences [15]. The analysis in [15] shows that pilot contamination tends to significantly reduce the performance of cell-free massive MIMO. On the other hand, the short coherence time does not allow the system to support a large number of users in cell-free massive MIMO, as demonstrated in [32].

A practical combination of the aforementioned techniques is training-based massive MIMO aided NOMA [6], [32]. In [6], Cheng *et al.* investigate the effect of NOMA in collocated massive MIMO systems, where the users are paired based on their distance from the base station. However, in this paper we use a different system model and beamforming design. Moreover, Cheng *et al.* [6] do not investigate any power allocation scheme.

M. Bashar, P. Xiao are with Institute for Communication Systems, the Home of the 5G Innovation Centre, University of Surrey, GU2 7XH, U.K., e-mail: {m.bashar,p.xiao@surrey}@surrey.ac.uk. and mb1465@york.ac.uk. K. Cumanan and A. G. Burr are with the Department of Electronic Engineering, University of York, Heslington, York, YO10 5DD, U.K., e-mail: {kanapathippillai.cumanan, alister.burr}@york.ac.uk. H. Q. Ngo is with the School of Electronics, Electrical Engineering and Computer Science, Queen's University Belfast, Belfast, BT7 1NN, U.K., e-mail: hien.ngo@qub.ac.uk. L. Hanzo is with the School of ECS, University of Southampton, SO17 1TU, U.K., e-mail: lh@ecs.soton.ac.uk.

The work of K. Cumanan and A. G. Burr was supported by H2020-MSCA-RISE-2015 under grant number 690750.

The work of H. Q. Ngo was supported by the UK Research and Innovation Future Leaders Fellowships under Grant MR/S017666/1.

The work of P. Xiao was supported by the U.K. Engineering and Physical Sciences Research Council under Grant EP/P008402/2. The authors also would like to acknowledge the support of the 5G Innovation Centre, University of Surrey, U.K. (http://www.surrey.ac.uk/5gic) for this work.

L. Hanzo would like to acknowledge the financial support of the Engineering and Physical Sciences Research Council projects EP/Noo4558/1, EP/PO34284/1, COALESCE, of the Royal Society's Global Challenges Research Fund Grant as well as of the European Research Council's Advanced Fellow Grant QuantCom.

Parts of this work was presented at the IEEE ICC 2019 [1].

The authors of [15], [29] indicate that conjugate beamforming does not need any channel state information (CSI) sharing between the APs. Hence, it is an attractive beamforming scheme for cell-free massive MIMO. Note that similar to [15], [33], we consider the per-AP power constraints. Moreover, in [34], [35], the authors investigate two different kinds of conjugate beamforming schemes operating under per-antenna power constraints: conjugate beamforming subject to long-term power constraints and normalized conjugate beamforming subject to short-term power constraints. Note that the average is taken over the codewords for long-term power constraints and channel fading coefficients, whereas the average is taken over the codewords in short-term power constraints for short-term power constraints [34], [35]. However, it is not clear, which form of conjugate beamforming and power constraint should be exploited in NOMA-based cell-free massive MIMO to achieve the best performance. Hence, it is important to investigate the effect of short-term and long-term per-antenna power constraints on the system performance [34], [35]. The max-min rate performance of OMA cell-free massive MIMO, relying on conjugate beamforming subject to long-term average power constraints is studied in [15], whereas the authors of [35] consider OMA cell-free massive MIMO, with normalized conjugate beamforming subject to short-term average power constraints. Note that the authors in [35] do not consider any optimization problem. Explicitly, we evaluate the performance of both conjugate beamforming and normalized conjugate beamforming in the NOMA-based cell-free massive MIMO system considered. The signal-to-interference-plus-noise ratio (SINR) formulas are derived for both beamforming designs.

In most practical scenarios, the OMA vs NOMA operating mode, is fixed. However, we achieve the best performance by adaptively switching between the Modes = {OMA, NOMA}. A mode switching scheme was proposed in [6], [36] for massive MIMO systems having collocated antennas. This mode switching is performed based on the statistics of the channel. In this paper, a downlink (DL) max-min BE problem is investigated. Fairness is defined as the equal opportunity for users to exploit the resources, which is an important aspect in next-generation systems. In [15], the authors have investigated the max–min SINR optimization problem in a cell-free massive MIMO system. Additionally, Li *et al.* [32] investigate the performance of NOMA-based cell-free massive MIMO system using the normalized conjugate beamforming proposed by Interdonato *et al.* [35]. Moreover, we conceive beneficial clustering schemes for NOMA-based cell-free massive MIMO systems. Explicitly, our clustering schemes are different from those of [37], [38], since no CSI is available at the APs and at the CPU. The numerical results reveal that with increasing the number of users in the cell-free massive MIMO system results in NOMA outperforms OMA. A similar conclusion has been made in [39] for collocated massive MIMO. Our numerical results will demonstrate that switching between different operating modes is capable of improving the overall system performance.

Our new contributions and results are summarized as follows:

1. The closed-form SINR of the NOMA-based cell-free massive MIMO system is derived using both conjugate beamforming and normalized conjugate beamforming, taking into account the effects of pilot contamination and imperfect SIC.
2. A max-min fairness power control problem is formulated which maximizes the smallest of all user BEs under per-AP power constraints. Then a bisection scheme is conceived for optimally solving the optimization problem. The power minimization problem of conjugate beamforming is solved using second order cone programming (SOCP), whereas for the normalized conjugate beamforming standard semidefinite programming (SDP) is utilized.
3. The complexity analysis of proposed schemes is presented.
4. A mode switching technique is proposed based on the average BE, where we define the mode set as Mode = {OMA, NOMA}. Then simulation results are provided for validating the superiority of NOMA over OMA in terms of the max-min BE of cell-free massive MIMO. Finally, the performance of the proposed clustering schemes is quantified.

Outline: The rest of the paper is organized as follows. Section II describes the system model, while Section III provides our performance analysis. The proposed max-min BE is presented in Section IV. Followed by our complexity analysis in Section V. Finally, Section VI provides numerical results, while Section VII concludes the paper.

Notation: The following notations are adopted in the rest of the paper. Uppercase and lowercase boldface letters are used for matrices and vectors, respectively. The notation $\mathbb{E}\{\cdot\}$ denotes expectation. $|\cdot|$ stands for absolute value. The norm (or Euclidean distance) of vector $\mathbf{x}$ is represented by $||\mathbf{x}||$. The operations $()^H$, $()^T$ and $()^*$ denote the conjugate transpose, transpose and conjugate, respectively. In addition, $x \sim \mathcal{CN}(0, \sigma^2)$ represents a zero-mean circularly symmetric complex Gaussian random variable with variance σ^2. Moreover, diag[$\mathbf{x}$] refers to a diagonal matrix, whose diagonal elements are the elements of vector $\mathbf{x}$. Finally, Tr [$\mathbf{X}$] refers to the trace of matrix $\mathbf{X}$, whereas rank[$\mathbf{X}$] represents the rank of matrix $\mathbf{X}$.

II. SYSTEM MODEL

We consider DL transmission in a NOMA-based cell-free massive MIMO system with M APs and K_{tot} single-antenna users randomly distributed in a large area. The users are grouped into L clusters, supporting K users. Hence, we have $K_{\text{tot}} = KL$. Furthermore, each AP has N DL transmit antennas. The channel coefficient vector $\mathbf{g}_{mlk} \in \mathbb{C}^{N\times 1}$, between the mth AP and the kth user in the lth cluster, is modeled as $\mathbf{g}_{mlk} = \sqrt{\beta_{mlk}}\mathbf{h}_{mlk}$, where β_{mlk} denotes the large-scale fading. The elements of $\mathbf{h}_{mlk}$ are independent and identically distributed (i.i.d.) $\mathcal{CN}(0, 1)$ random variables, and represent the small-scale fading [15].

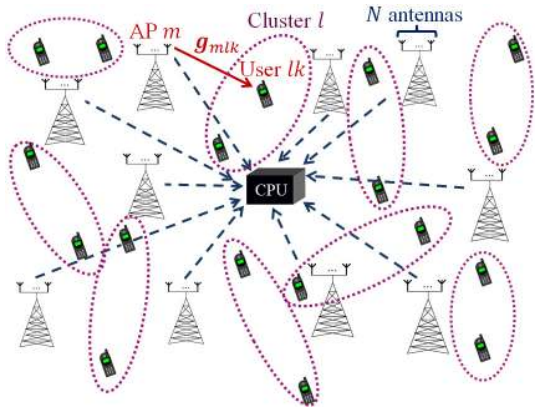


Figure 1. A cell-free massive MIMO system with K_{tot} single-antenna users and M APs. Each AP is equipped with N antennas. The solid lines denote the DL channels and the dashed lines present the backhaul links from the APs to the CPU. The users are grouped into L clusters which are shown by dotted lines. Each cluster includes K users, and it is assumed $K = 2$ here.

A. Uplink Channel Estimation

All pilot sequences transmitted by all the K_{tot} users in the channel estimation phase are collected in a matrix $\mathbf{\Phi} = [\boldsymbol{\phi}_{11} \cdots \boldsymbol{\phi}_{K1} \cdots \boldsymbol{\phi}_{1L} \cdots \boldsymbol{\phi}_{KL}] \in \mathbb{C}^{\tau_p \times K_{\text{tot}}}$, where $||\boldsymbol{\phi}_{lk}||^2 = 1$, τ_p is the length of the pilot sequence for each user and the klth column, $\boldsymbol{\phi}_{kl}$, represents the pilot sequence used for the kth user in the lth cluster. Moreover, it is assumed that the users in the same cluster employ the same pilot sequences ($\boldsymbol{\phi}_{kl} = \hat{\boldsymbol{\phi}}_l$, $\forall k$), whereas orthogonal pilots are assigned to different clusters. Note that in [15], the authors exploit uplink pilots to estimate the channel of users. However, the authors of [6] claim that estimating a linear combination of the users' channel in the same cluster provides a better performance. Hence, an alternative technique of employing the uplink pilots in NOMA-based massive MIMO systme is to estimate the following linear combination [6]:

$$\mathbf{f}_{ml} = \sum_{k=1}^{K} \mathbf{g}_{mlk}, \forall l. \tag{1}$$

After performing a de-spreading operation, the minimum mean square error (MMSE) estimate of the linear combination $\mathbf{f}_{ml}$ is given by

$$\hat{\mathbf{f}}_{ml} = c_{ml} \left(\sqrt{\tau_p \rho_p} \sum_{k=1}^{K} \mathbf{g}_{mlk} + \mathbf{W}_{p,m} \hat{\boldsymbol{\phi}}_l \right), \tag{2}$$

where $\mathbf{W}_{p,m} \in \mathbb{C}^{M \times K}$ denotes the noise sequence at the mth AP whose elements are i.i.d. $\mathcal{CN}(0,1)$ and ρ_p represents the normalized signal-to-noise ratio (SNR) of each pilot sequence (which we define in Section VI). Hence, there is no pilot contamination between clusters. Moreover, the users in the same cluster employ the same pilot sequences, resulting $\boldsymbol{\phi}_{lk}^H \boldsymbol{\phi}_{lk'} = 1$. Additionally, c_{ml} is given by

$$c_{ml} = \frac{\sqrt{\tau_p \rho_p} \sum_{k'=1}^{K} \beta_{mlk'}}{\tau_p \rho_p \sum_{k'=1}^{K} \beta_{mlk'} + 1}. \tag{3}$$

Note that as in [15], we assume that the large-scale fading, β_{mlk}, is known both at the CPU and at the users. The estimated channels in (2) are then used by the APs for determining both the receiver filter coefficients and the power allocations.

B. Downlink Transmission with Conjugate Beamforming

In this subsection, we consider the DL data transmission relying on conjugate beamforming [15]. The signal transmitted from the mth AP is represented by

$$\mathbf{x}_m = \sqrt{\rho_d} \sum_{l=1}^{L} \sum_{k=1}^{K} \sqrt{\eta_{mlk}} \hat{\mathbf{f}}_{ml}^* s_{lk}, \tag{4}$$

where s_{lk} ($\mathbb{E}\{|s_{lk}|^2\} = 1$) and η_{mlk} denote the transmitted symbol and the transmit power at the mth AP, respectively. Furthermore, ρ_d represents the maximum normalized transmit power (normalized by the noise power N_0, defined in Section VI) at the APs. Hence, the normalized transmit power is given by

$$\mathbb{E}\{||\mathbf{x}_m||^2\} = \rho_d N \sum_{l=1}^{L} \sum_{k=1}^{K} \eta_{mlk} \gamma_{ml}, \tag{5}$$

where

$$\gamma_{ml} = \mathbb{E}\left\{ \left\| \left[\hat{\mathbf{f}}_{ml} \right]_n \right\|^2 \right\} = \frac{\tau_p \rho_p (\sum_{k'=1}^{K} \beta_{mlk'})^2}{\tau_p \rho_p \sum_{k'=1}^{K} \beta_{mlk'} + 1}. \tag{6}$$

Moreover, note that $\gamma_{ml} = \sqrt{\tau_p \rho_p} (\sum_{k'=1}^{K} \beta_{mlk'}) c_{ml}$. The power elements η_{mlk} are designed to satisfy the following power constraints:

$$\sum_{l=1}^{L} \sum_{k=1}^{K} \eta_{mlk} \gamma_{ml} \leq \frac{1}{N}, \forall m. \tag{7}$$

The power constraints in (7) are referred to long-term power constraints, since the expectation is taken both over codewords and the channel fading coefficients [35].

C. Downlink Transmission with Normalized Conjugate Beamforming

Next, we consider the DL data transmission relying on normalized conjugate beamforming [35]. Explicitly, the beamforming vectors are the conjugate of the estimated channels and they are normalized by their Euclidean norm. The signal transmitted from the mth AP is represented by

$$\mathbf{x}_m = \sqrt{\rho_d} \sum_{l=1}^{L} \sum_{k=1}^{K} \sqrt{\mu_{mlk}} \frac{\hat{\mathbf{f}}_{ml}^*}{||\hat{\mathbf{f}}_{ml}||} s_{lk}, \tag{8}$$

where μ_{mlk} is the DL transmit power at the mth AP. The normalized transmit power is given by

$$\mathbb{E}\{||\mathbf{x}_m||^2\} = \rho_d \sum_{l=1}^{L} \sum_{k=1}^{K} \mu_{mlk}. \tag{9}$$

Following the same strategy as in [35], the power elements μ_{mlk} are designed to satisfy the following short-term power constraints:

$$\sum_{l=1}^{L} \sum_{k=1}^{K} \mu_{mlk} \leq 1, \forall m. \tag{10}$$

The power constraints in (10) are referred as short-term power constraints, since the expectation is taken only over the codewords [35].

D. Received Signal

The signal received at the kth user in the lth cluster is given by

$$r_{lk}^{\mathrm{B}} = \sum_{m=1}^{M} \mathbf{g}_{mlk}^{T} \mathbf{x}_m + n_{lk}, \tag{11}$$

where $n_{lk} \sim \mathcal{CN}(0,1)$ is the noise at the kth user in the lth cluster, and the superscript B refers either to CB or NCB for the cases of conjugate and normalized conjugate beamforming techniques, respectively.

III. Performance Analysis

In this section, we derive the DL BE for the NOMA-based cell-free massive MIMO system without using any DL training by following a similar approach to [15]. In particular, it is assumed that no instantaneous CSI is available at the users, which is a reasonable assumption in massive MIMO systems thanks to the channel hardening phenomenon. Therefore, the users exploit the channel statistics instead of the instantaneous CSI to perform SIC. In the following subsections, we first highlight the basic concepts of NOMA in cell-free massive MIMO. Next, the attainable throughput is derived for the both conjugate and for the normalized conjugate beamforming techniques.

A. NOMA without Downlink Training

Again, we assume that the users exploit the channel statistics to decode data. Since there is no DL training, users rely on the average of the effective channel gain as an estimate of the channel gain [6]. We assume that in the lth cluster, "user-$l1$" is the strongest user (and hence can decode the signals intended to other users by using SIC technique) whereas "user-lK" is the weakest user and he can decode only his signal and not the other users' signals. In other words, NOMA is employed only within each cluster and not between the clusters. When the instantaneous CSI is available at the users, the users in each cluster can be ordered based on their effective channel gain. However, in the absence of DL training, we sort the users based on their channel statistics, which is discussed details in Sections III-B and III-C. In order to successfully implement the SIC at the stronger users to decode the weaker user signals, the following necessary NOMA condition should be satisfied [6], [40]:

$$\mathbb{E}\left\{\log_2\left(1+\mathrm{SINR}_{lj}^{lk}\right)\right\} \geq \mathbb{E}\left\{\log_2\left(1+\mathrm{SINR}_{lk}^{lk}\right)\right\}, \forall j<k, \forall l, \tag{12}$$

where SINR_{lj}^{lk} refers to the effective SINR of user j in cluster l when user j in cluster l is decoding the signal intended for user k in the same cluster l. Based on this necessary condition, the achievable rate of the kth user in the lth cluster can be written as

$$\begin{aligned} &\mathrm{R}_{lk}^{lk,\mathrm{final}} \\ &= \min\left(\mathbb{E}\left\{\log_2\left(1+\mathrm{SINR}_{lj}^{lk}\right)\right\}, \mathbb{E}\left\{\log_2\left(1+\mathrm{SINR}_{lk}^{lk}\right)\right\}\right), \forall l,k, \end{aligned} \tag{13}$$

where $\mathrm{R}_{lk}^{lk,\mathrm{final}}$ is the achievable rate of user k in cluster l. Several clustering schemes are investigated in Section VI-B.

B. Conjugate Beamforming

In this subsection, we derive the achievable SINR of the user with the aid of conjugate beamforming defined in (4). We use NOMA, in which the user having a higher received power detects its signal first, which is then demodulated and the corresponding signal is subtracted from the composite received signal, hence leaving behind the uninterfered signal of the lower-power user. Given the fact that only the statistics of the channels are available at the users' ends and exploiting the analysis in [6], the signal received by the kth user in the lth cluster is given by

$$\begin{aligned} r_{lk}^{lk,\mathrm{CB}} &= r_{lk}^{\mathrm{CB}} - \sqrt{\rho_d}\sum\nolimits_{k''=k+1}^{K} \mathbb{E}\left\{\sum\nolimits_{m=1}^{M}\sqrt{\eta_{mlk''}}\mathbf{g}_{mlk}^{T}\hat{\mathbf{f}}_{ml}^{*}\right\} s_{lk''} \\ &= \sqrt{\rho_d}\sum_{m=1}^{M}\sum_{l'=1}^{L}\sum_{k'=1}^{K}\sqrt{\eta_{ml'k'}}\mathbf{g}_{mlk}^{T}\hat{\mathbf{f}}_{ml'}^{*}s_{l'k'} + n_{lk} \\ &- \sqrt{\rho_d}\sum_{k''=k+1}^{K}\mathbb{E}\left\{\sum_{m=1}^{M}\sqrt{\eta_{mlk''}}\mathbf{g}_{mlk}^{T}\hat{\mathbf{f}}_{ml}^{*}\right\}s_{lk''} \\ &= \underbrace{\sqrt{\rho_d}\,\mathbb{E}\left\{\sum_{m=1}^{M}\sqrt{\eta_{mlk}}\mathbf{g}_{mlk}^{T}\hat{\mathbf{f}}_{ml}^{*}\right\}s_{lk}}_{\mathrm{DS}_{lk}} \\ &+ \underbrace{\sqrt{\rho_d}\left(\sum_{m=1}^{M}\sqrt{\eta_{mlk}}\mathbf{g}_{mlk}^{T}\hat{\mathbf{f}}_{ml}^{*} - \mathbb{E}\left\{\sum_{m=1}^{M}\sqrt{\eta_{mlk}}\mathbf{g}_{mlk}^{T}\hat{\mathbf{f}}_{ml}^{*}\right\}\right)s_{lk}}_{\mathrm{BU}_{lk}} \\ &+ \sum_{k'\neq k}^{k-1}\underbrace{\sqrt{\rho_d}\sum_{m=1}^{M}\sqrt{\eta_{mlk'}}\mathbf{g}_{mlk}^{T}\hat{\mathbf{f}}_{ml}^{*}}_{\mathrm{IUI}_{lk'}}s_{lk'} \\ &+ \sum_{k''=k+1}^{K}\underbrace{\sqrt{\rho_d}\left(\sum_{m=1}^{M}\sqrt{\eta_{mlk''}}\mathbf{g}_{mlk}^{T}\hat{\mathbf{f}}_{ml}^{*} - \mathbb{E}\left\{\sum_{m=1}^{M}\sqrt{\eta_{mlk''}}\mathbf{g}_{mlk}^{T}\hat{\mathbf{f}}_{ml}^{*}\right\}\right)}_{\mathrm{ISIC}_{lk''}}s_{lk''} \\ &+ \sum_{l'\neq l}^{L}\sum_{k'=1}^{K}\underbrace{\sqrt{\rho_d}\sum_{m=1}^{M}\sqrt{\eta_{ml'k'}}\mathbf{g}_{mlk}^{T}\hat{\mathbf{f}}_{ml'}^{*}}_{\mathrm{ICI}_{l'k'}}s_{l'k'} + n_{lk}, \end{aligned} \tag{14}$$

where DS_{lk} and BU_{lk} denote the desired signal (DS) and beamforming uncertainty (BU) for the kth user in the lth cluster, respectively, and $\mathrm{IUI}_{lk'}$ represents the inter-user-interference (IUI) caused by the k'th user in the lth cluster. Additionally, $\mathrm{ISIC}_{lk''}$ accounts for the interference imposed by the k''th to the kth user in the lth cluster due to imperfect SIC (ISIC), and $\mathrm{ICI}_{l'k'}$ is the inter-cluster-interference (ICI) caused by the users in clusters $l' \neq l$. Moreover, the superscript CB in (14) refers to conjugate beamforming.

Proposition 1. *The terms DS_{lk}, BU_{lk}, $IUI_{lk'}$, $ISIC_{lk''}$, and $ICI_{l'k'}$ are mutually uncorrelated.*

The achievable rate can be defined as $\mathrm{R}_{lk}^{\mathrm{CB}} = \log_2(1+\mathrm{SINR}_{lk}^{\mathrm{CB}})$. Based on the analysis in [12] and exploiting Proposition 1, $\mathrm{SINR}_{lk}^{\mathrm{CB}}$ is given by (15). The closed-form expression for the achievable DL rate of the kth user in the lth cluster is given in the following theorems:

$$\mathrm{SINR}_{lk} = \frac{|\mathrm{DS}_{lk}|^2}{\mathbb{E}\{|\mathrm{BU}_{lk}|^2\} + \sum_{k'=1}^{k-1} \mathbb{E}\left\{|\mathrm{IUI}_{lk'}|^2\right\} + \sum_{k''=k+1}^{K} \mathbb{E}\left\{|\mathrm{ISIC}_{lk''}|^2\right\} + \sum_{l'\neq l}^{L} \sum_{k'=1}^{K} \mathbb{E}\left\{|\mathrm{ICI}_{l'k'}|^2\right\} + 1}. \tag{15}$$

$$\mathrm{SINR}_{lk}^{lk,\mathrm{CB}} = \frac{N^2 \left(\sum_{m=1}^{M} \sqrt{\eta_{mlk}} \frac{\gamma_{ml}\beta_{mlk}}{\sum_{i=1}^{K} \beta_{mli}} \right)^2}{N^2 \sum_{k'=1}^{k-1} \left(\sum_{m=1}^{M} \sqrt{\eta_{mlk'}} \frac{\gamma_{ml}\beta_{mlk}}{\sum_{i=1}^{K} \beta_{mli}} \right)^2 + N \sum_{l'=1}^{L} \sum_{k'=1}^{K} \sum_{m=1}^{M} \eta_{ml'k'}\beta_{mlk}\gamma_{ml'} + \frac{1}{\rho_d}}. \tag{16}$$

$$\mathrm{SINR}_{lj}^{lk,\mathrm{CB}} = \frac{N^2 \left(\sum_{m=1}^{M} \sqrt{\eta_{mlk}} \frac{\gamma_{ml}\beta_{mlj}}{\sum_{i=1}^{K} \beta_{mli}} \right)^2}{N^2 \sum_{k'=1}^{k-1} \left(\sum_{m=1}^{M} \sqrt{\eta_{mlk'}} \frac{\gamma_{ml}\beta_{mlj}}{\sum_{i=1}^{K} \beta_{mli}} \right)^2 + N \sum_{l'=1}^{L} \sum_{k'=1}^{K} \sum_{m=1}^{M} \eta_{ml'k'}\beta_{mlj}\gamma_{ml'} + \frac{1}{\rho_d}}. \tag{17}$$

Theorem 1. *Having the channel statistics at the users and employing conjugate beamforming at the APs, the closed-form expression for the achievable DL rate of the kth user in the lth cluster is given by* $R_{lk}^{CB} = \log_2(1 + SINR_{lk}^{CB})$*, where the* $SINR_{lk}^{CB}$ *is given by (16).*

Proof: Please refer to Appendix A. ∎

Theorem 2. *Having the channel statistics at the users and employing conjugate beamforming at the APs, the closed-form expression for the achievable DL rate of the kth user (weaker user) at the jth user (stronger user) in the lth cluster is given by* $R_{lj}^{lk,CB} = \log_2(1 + SINR_{lj}^{lk,CB})$*, where the* $SINR_{lj}^{lk,CB}$ *is given by (17).*

Proof: Proof follows steps similar to the steps in Appendix A and is omitted due to the space limit. ∎

1) User Ordering: By the assumption that users are ordered based on their channel quality, NOMA uses the power domain to transmit multiple signals over the same resource, and performs SIC at the receivers to decode the corresponding signals [9]. Considering the expression in (16), we take the term $\mathbf{h}_{lk}^{\mathrm{vir}} = \left[\frac{\gamma_{1l}}{\sum_{i=1}^{K}\beta_{1li}}\beta_{1lk}, \frac{\gamma_{2l}}{\sum_{i=1}^{K}\beta_{2li}}\beta_{2lk} \cdots \frac{\gamma_{Ml}}{\sum_{i=1}^{K}\beta_{Mli}}\beta_{Mlk} \right]^T, \forall l, k$, as the virtual channel of the kth user at the lth cluster. Next we sort the users based on the quality of this virtual channel, i.e., $||\mathbf{h}_{l1}^{\mathrm{vir}}||_2 \geq ||\mathbf{h}_{l2}^{\mathrm{vir}}||_2 \geq \cdots \geq ||\mathbf{h}_{lK}^{\mathrm{vir}}||_2, \forall l$. Note that determining the optimal user ordering is a NP-hard combinatorial problem, which is beyond the scope of this work.

C. Normalized Conjugate Beamforming

In this section, we consider the achievable rate with normalized conjugate beamforming given in (8). Exploiting the NOMA scheme of Section III-A, the signal received for the kth user in the lth cluster using the normalized conjugate beamforming at the APs is given by

$$\begin{aligned}
r_{lk}^{lk,\mathrm{NCB}} &= r_{lk}^{\mathrm{NCB}} - \sqrt{\rho_d} \sum_{k''=k+1}^{K} \mathbb{E}\left\{ \sum_{m=1}^{M} \sqrt{\eta_{mlk''}} \mathbf{g}_{mlk}^T \frac{\hat{\mathbf{f}}_{ml}^*}{||\hat{\mathbf{f}}_{ml}||} \right\} s_{lk''} \\
&= \sqrt{\rho_d} \sum_{m=1}^{M} \sum_{l'=1}^{L} \sum_{k'=1}^{K} \sqrt{\mu_{ml'k'}} \mathbf{g}_{mlk}^T \frac{\hat{\mathbf{f}}_{ml'}^*}{||\hat{\mathbf{f}}_{ml'}||} s_{l'k'} + n_{lk} \\
&- \sqrt{\rho_d} \sum_{k''=k+1}^{K} \mathbb{E}\left\{ \sum_{m=1}^{M} \sqrt{\eta_{mlk''}} \mathbf{g}_{mlk}^T \frac{\hat{\mathbf{f}}_{ml}^*}{||\hat{\mathbf{f}}_{ml}||} \right\} s_{lk''} \\
&= \underbrace{\sqrt{\rho_d}\mathbb{E}\left\{ \sum_{m=1}^{M} \sqrt{\mu_{mlk}} \mathbf{g}_{mlk}^T \frac{\hat{\mathbf{f}}_{ml}^*}{||\hat{\mathbf{f}}_{ml}||} \right\}}_{\mathrm{DS}_{lk}} s_{lk} \\
&+ \underbrace{\sqrt{\rho_d} \left(\sum_{m=1}^{M} \sqrt{\mu_{mlk}} \mathbf{g}_{mlk}^T \frac{\hat{\mathbf{f}}_{ml}^*}{||\hat{\mathbf{f}}_{ml}||} - \mathbb{E}\left\{ \sum_{m=1}^{M} \sqrt{\mu_{mlk}} \mathbf{g}_{mlk}^T \frac{\hat{\mathbf{f}}_{ml}^*}{||\hat{\mathbf{f}}_{ml}||} \right\} \right)}_{\mathrm{BU}_{lk}} s_{lk} \\
&+ \sum_{k'\neq k}^{k-1} \underbrace{\sqrt{\rho_d} \sum_{m=1}^{M} \sqrt{\mu_{mlk'}} \mathbf{g}_{mlk}^T \frac{\hat{\mathbf{f}}_{ml}^*}{||\hat{\mathbf{f}}_{ml}^*||}}_{\mathrm{IUI}_{lk'}} s_{lk'} \\
&+ \sum_{k''=k+1}^{K} \underbrace{\sqrt{\rho_d} \left(\sum_{m=1}^{M} \sqrt{\mu_{mlk''}} \mathbf{g}_{mlk}^T \frac{\hat{\mathbf{f}}_{ml}^*}{||\hat{\mathbf{f}}_{ml}||} - \mathbb{E}\left\{ \sum_{m=1}^{M} \sqrt{\mu_{mlk''}} \mathbf{g}_{mlk}^T \frac{\hat{\mathbf{f}}_{ml}^*}{||\hat{\mathbf{f}}_{ml}^*||} \right\} \right)}_{\mathrm{ISIC}_{lk''}} s_{lk''} \\
&+ \sum_{l'\neq l}^{L} \sum_{k'=1}^{K} \underbrace{\sqrt{\rho_d} \sum_{m=1}^{M} \sqrt{\mu_{ml'k'}} \mathbf{g}_{mlk}^T \frac{\hat{\mathbf{f}}_{ml'}^*}{||\hat{\mathbf{f}}_{ml'}||}}_{\mathrm{ICI}_{l'k'}} s_{l'k'} + n_{lk},
\end{aligned} \tag{18}$$

where the superscript NCB refers to normalized conjugate beamforming.

Proposition 2. *The terms* DS_{lk}*,* BU_{lk}*,* $IUI_{lk'}$*,* $ISIC_{lk''}$*, and* $ICI_{l'k'}$ *are mutually uncorrelated.*

$$\sigma_{ml} = \frac{\sqrt{\pi}\left(N^{\frac{3}{2}}\left(\beta_{ml1}-\beta_{ml2}\right)+\frac{\beta_{ml2}\left(1-\beta_{ml1k}\tau_p\rho_p\right)}{\left(\frac{1}{N\beta_{ml2}\tau_p\rho_p}\right)^{\frac{3}{2}}}+\beta_{ml1}\left(N\beta_{ml1}\tau_p\rho_p\right)^{\frac{3}{2}}\left(2\beta_{ml2}\tau_p\rho_p-1\right)\right)}{2N\left(\beta_{ml1}-\beta_{ml2}\right)\left(\beta_{ml1}\tau_p\rho_p-1\right)\left(\beta_{ml2}\tau_p\rho_p-1\right)}, \text{ for } K=2. \tag{21}$$

The closed-form expression for the achievable DL rate of the kth user in the lth cluster is given in the following theorem:

Theorem 3. *Exploiting the channel statistics at the users and employing normalized conjugate beamforming at the APs, the closed-form expression for the achievable DL rate of the kth user in the lth cluster is given by* $R_{lk}^{NCB} = \log_2(1+SINR_{lk}^{NCB})$*, where the* $SINR_{lk}^{NCB}$ *is given by:*

$$SINR_{lk}^{lk,NCB} = \frac{\alpha A_1}{\alpha\left(A_2-A_3-A_4-A_6\right)+NA_5+\frac{1}{\rho_d}}, \tag{19}$$

where $\alpha = N^2\tau_p\rho_p$ *and*

$$A_1 = \left(\sum_{m=1}^{M}\sqrt{\mu_{mlk}}\frac{\beta_{mlk}}{\sigma_{ml}}\right)^2,\ A_2 = \sum_{k'\neq k}^{K}\left(\sum_{m=1}^{M}\frac{\sqrt{\mu_{mlk'}}\beta_{mlk}}{\sigma_{ml}}\right)^2, \tag{20a}$$

$$A_3 = \sum_{k'\neq k}^{K}\sum_{m=1}^{M}\frac{\mu_{mlk'}\beta_{mlk}^2}{\sigma_{ml}^2}, A_4 = \sum_{k''=k+1}^{K}\left(\sum_{m=1}^{M}\frac{\sqrt{\mu_{mlk''}}\beta_{mlk}}{\sigma_{ml}}\right)^2, \tag{20b}$$

$$A_5 = \sum_{l'=1}^{L}\sum_{k'=1}^{K}\sum_{m=1}^{M}\mu_{ml'k'}\beta_{mlk},\ A_6 = \sum_{m=1}^{M}\mu_{mlk}\frac{\beta_{mlk}^2}{\sigma_{ml}^2}, \tag{20c}$$

where the term σ_{ml} *is given by (21).*

Proof: Please refer to Appendix B. ∎

Theorem 4. *Having the channel statistics at the users and employing conjugate beamforming at the APs, the closed-form expression for the achievable DL rate of the kth user (weaker user) at the jth user (stronger user) in the lth cluster is given by* $R_{lj}^{lk,NCB} = \log_2(1+SINR_{lj}^{lk,NCB})$*, where the* $SINR_{lj}^{lk,NCB}$ *is*

$$SINR_{lj}^{lk,NCB} = \frac{\alpha B_1}{\alpha\left(B_2-B_3-B_4-B_5\right)+NB_6+\frac{1}{\rho_d}}, \tag{22}$$

where

$$B_1 = \left(\sum_{m=1}^{M}\sqrt{\mu_{mlk}}\frac{\beta_{mlj}}{\sigma_{ml}}\right)^2,\ B_2 = \sum_{k'\neq k}^{K}\left(\sum_{m=1}^{M}\frac{\sqrt{\mu_{mlk'}}\beta_{mlj}}{\sigma_{ml}}\right)^2, \tag{23a}$$

$$B_3 = \sum_{k'\neq k}^{K}\sum_{m=1}^{M}\frac{\mu_{mlk'}\beta_{mlj}^2}{\sigma_{ml}^2}, B_4 = \sum_{k''=k+1}^{K}\left(\sum_{m=1}^{M}\frac{\sqrt{\mu_{mlk''}}\beta_{mlj}}{\sigma_{ml}}\right)^2, \tag{23b}$$

$$B_5 = \sum_{l'=1}^{L}\sum_{k'=1}^{K}\sum_{m=1}^{M}\mu_{ml'k'}\beta_{mlj},\ B_6 = \sum_{m=1}^{M}\mu_{mlk}\frac{\beta_{mlj}^2}{\sigma_{ml}^2}. \tag{23c}$$

Proof: Proof follows the steps similar to the steps in Appendix B and is omitted due to the space limit. ∎

1) User Ordering: Using the SINR formula given in (22) and (23), we take the term $\mathbf{h}_{lk}^{\text{vir}} = \left[\frac{\beta_{1lk}}{\sigma_{1l}}\frac{\beta_{2lk}}{\sigma_{2l}}\cdots\frac{\beta_{Mlk}}{\sigma_{Ml}}\right]^T, \forall l,k$ as the virtual channel of the kth user at the lth cluster. Following this, we sort the users based on the quality of this virtual channel, i.e., $||\mathbf{h}_{l1}^{\text{vir}}||_2 \geq ||\mathbf{h}_{l2}^{\text{vir}}||_2 \geq \cdots \geq ||\mathbf{h}_{lK}^{\text{vir}}||_2, \forall l$.

D. Bandwidth Efficiency

Based on [11], the BE is measured in terms of bits/s/Hz sum-rate. In this paper, we consider the smallest BE of the users as the performance metric to characterise our cell-free massive MIMO system. The BE (in bit/s/Hz) of the kth user in the lth cluster can be defined by

$$\text{S}_{lk}^{lk,\text{final,B}} = \left(1-\frac{\tau_p}{\tau_c}\right)\log_2\left(1+\text{SINR}_{lk}^{lk,\text{final,B}}\right), \tag{24}$$

where τ_c denotes the number of samples for each coherence interval, and the superscript B refers either to CB or NCB for the cases of conjugate and normalized conjugate beamforming techniques, respectively. Finally, note that

$$\text{SINR}_{lk}^{lk,\text{final}} = \min\left(\text{SINR}_{lj}^{lk}, \text{SINR}_{lk}^{lk}\right), \forall l,k. \tag{25}$$

IV. Max-Min Bandwidth Efficiency

In this section, we consider the max-min user BE problem in cell-free massive MIMO, where the minimum user DL BE is maximized, while satisfying per-AP power constraints. In the following subsections, we show that the max-min BE problem of both conjugate and normalized conjugate beamforming may be found by a bisection search method. Furthermore, it is shown that SOCP can be exploited for solving the power minimization problem in each iteration of the bisection search method for conjugate beamforming, while SDP for the normalized conjugate beamforming designs.

A. Max-Min BE with Conjugate Beamforming

The max-min BE of conjugate beamforming is presented in this subsection. This problem under per-AP power constraints can be formulated as follows:

$$P_1: \max_{\eta_{mlk}} \min_{k=1\cdots K, l=1\cdots L} \text{S}_{lk}^{lk,\text{final,CB}} \tag{26a}$$

$$\text{s.t.} \sum_{l=1}^{L}\sum_{k=1}^{K}\eta_{mlk}\gamma_{ml} \leq \frac{1}{N}, \forall m,\ \eta_{mlk} \geq 0, \quad \forall m,\forall l,\forall k. \tag{26b}$$

Since log(.) is a monotonically increasing function, Problem P_1 can be re-written as Problem P_2 as (27).

By defining new slack variables ν_m, $\tilde{\nu}_m$, $\varrho_{lk'j}$, and $\tilde{\varrho}_{lk'k}$, Problem P_2 can be re-formulated as (28), where $\varsigma_{mlk} = \sqrt{\eta_{mlk}}$.

$$P_2: \max_{\eta_{mlk}} \min_{\{k=1\cdots K, l=1\cdots L\}} \text{SINR}_{lk}^{lk,\text{final,CB}} = \min\left(\text{SINR}_{lk}^{lk,\text{CB}}, \text{SINR}_{lj}^{lk,\text{CB}}\right), \forall j < k$$

$$= \min\left(\frac{N^2\left(\sum_{m=1}^{M}\sqrt{\eta_{mlk}}\frac{\gamma_{ml}}{\sum_{i=1}^{K}\beta_{mli}}\beta_{mlj}\right)^2}{N^2\sum_{k'=1}^{k-1}\left(\sum_{m=1}^{M}\sqrt{\eta_{mlk'}}\frac{\gamma_{ml}}{\sum_{i=1}^{K}\beta_{mli}}\beta_{mlj}\right)^2 + N\sum_{l'=1}^{L}\sum_{k'=1}^{K}\sum_{m=1}^{M}\eta_{ml'k'}\beta_{mlj}\gamma_{ml'} + \frac{1}{\rho_d}}, \right.$$

$$\left. \frac{N^2\left(\sum_{m=1}^{M}\sqrt{\eta_{mlk}}\frac{\gamma_{ml}}{\sum_{i=1}^{K}\beta_{mli}}\beta_{mlk}\right)^2}{N^2\sum_{k'=1}^{k-1}\left(\sum_{m=1}^{M}\sqrt{\eta_{mlk'}}\frac{\gamma_{ml}}{\sum_{i=1}^{K}\beta_{mli}}\beta_{mlk}\right)^2 + N\sum_{l'=1}^{L}\sum_{k'=1}^{K}\sum_{m=1}^{M}\eta_{ml'k'}\beta_{mlk}\gamma_{ml'} + \frac{1}{\rho_d}} \right), \forall j < k \tag{27a}$$

$$\text{s.t.} \quad \sum_{l=1}^{L}\sum_{k=1}^{K}\eta_{mlk}\gamma_{ml} \le \frac{1}{N}, \ \forall m, \quad \eta_{mlk} \ge 0, \quad \forall m, \forall l, \forall k. \tag{27b}$$

$$P_3: \max_{\{\varsigma_{mlk}, \varrho_{lk'j}, \nu_m, \tilde{\varrho}_{lk'j}, \tilde{\nu}_m\}} \quad t \tag{28a}$$

$$\text{s.t.} \quad \frac{N^2\left(\sum_{m=1}^{M}\varsigma_{mlk}\frac{\gamma_{ml}\beta_{mlj}}{\sum_{i=1}^{K}\beta_{mli}}\right)^2}{N^2\sum_{k'=1}^{k-1}\varrho_{lk'j}^2 + N\sum_{m=1}^{M}\beta_{mlj}\nu_m^2 + \frac{1}{\rho_d}} \ge t, \forall j < k, \tag{28b}$$

$$\left(\sum_{l'\neq l}^{L}\sum_{k'=1}^{K}\gamma_{ml'}\varsigma_{ml'k'}^2 + \sum_{k'=1}^{k}\gamma_{ml}\varsigma_{mlk'}^2\right) \le \nu_m^2, \ \forall m, \forall j < k, \quad 0 \le \nu_m \le \frac{1}{\sqrt{N}}, \ \forall m, \tag{28c}$$

$$\sum_{m=1}^{M}\varsigma_{mlk'}\frac{\gamma_{ml}\beta_{mlj}}{\sum_{i=1}^{K}\beta_{mli}} \le \varrho_{lk'j}, 1 \le k' \le k-1, \forall j < k, \tag{28d}$$

$$\frac{N^2\left(\sum_{m=1}^{M}\varsigma_{mlk}\frac{\gamma_{ml}\beta_{mlk}}{\sum_{i=1}^{K}\beta_{mli}}\right)^2}{N^2\sum_{k'=1}^{k-1}\tilde{\varrho}_{lk'k}^2 + N\sum_{m=1}^{M}\beta_{mlk}\tilde{\nu}_m^2 + \frac{1}{\rho_d}} \ge t, \tag{28e}$$

$$\left(\sum_{l'\neq l}^{L}\sum_{k'=1}^{K}\gamma_{ml'}\varsigma_{ml'k'}^2 + \sum_{k'=1}^{k}\gamma_{ml}\varsigma_{mlk'}^2\right) \le \tilde{\nu}_m^2, \ \forall m, \ \ 0 \le \tilde{\nu}_m \le \frac{1}{\sqrt{N}}, \ \forall m, \tag{28f}$$

$$\sum_{m=1}^{M}\varsigma_{mlk'}\frac{\gamma_{ml}\beta_{mlk}}{\sum_{i=1}^{K}\beta_{mli}} \le \tilde{\varrho}_{lk'k}, 1 \le k' \le k-1, \tag{28g}$$

$$\varsigma_{mlk} \ge 0, \quad \forall m, \forall l, \forall k, \sum_{l=1}^{L}\sum_{k=1}^{K}\varsigma_{mlk}^2\gamma_{ml} \le \frac{1}{N}, \ \forall m, \tag{28h}$$

$$P_4: \min_{\{\varsigma_{mlk},\varrho_{lk'j},\nu_m\}} \quad \sum_{m=1}^{M}\sum_{l'=1}^{L}\sum_{k'=1}^{K}\gamma_{ml'}\varsigma_{ml'k'}^2 \tag{29a}$$

$$\text{s.t.} \quad \frac{N^2\left(\sum_{m=1}^{M}\varsigma_{mlk}\frac{\gamma_{ml}\beta_{mlj}}{\sum_{i=1}^{K}\beta_{mli}}\right)^2}{N^2\sum_{k'=1}^{k-1}\varrho_{lk'j}^2+N\sum_{m=1}^{M}\beta_{mlj}\nu_m^2+\frac{1}{\rho_d}} \geq t, \forall j<k, \ \frac{N^2\left(\sum_{m=1}^{M}\varsigma_{mlk}\frac{\gamma_{ml}\beta_{mlk}}{\sum_{i=1}^{K}\beta_{mli}}\right)^2}{N^2\sum_{k'=1}^{k-1}\tilde{\varrho}_{lk'k}^2+N\sum_{m=1}^{M}\beta_{mlk}\tilde{\nu}_m^2+\frac{1}{\rho_d}} \geq t, \tag{29b}$$

$$(28c)-(28d),(28f)-(28h).$$

$$P_5: \min_{\{\varsigma_{mlk},\varrho_{lk'j},\nu_m\}} \quad \sum_{m=1}^{M}\sum_{l'=1}^{L}\sum_{k'=1}^{K}\gamma_{ml'}\varsigma_{ml'k'}^2 \tag{30a}$$

$$\text{s.t.} \quad ||\mathbf{z}_{lj}|| \leq \frac{N\sum_{m=1}^{M}\varsigma_{mlk}\frac{\gamma_{ml}\beta_{mlj}}{\sum_{i=1}^{K}\beta_{mli}}}{\sqrt{t}}, \forall j<k, \ ||\tilde{\mathbf{z}}_{lk}|| \leq \frac{N\sum_{m=1}^{M}\varsigma_{mlk}\frac{\gamma_{ml}\beta_{mlk}}{\sum_{i=1}^{K}\beta_{mli}}}{\sqrt{t}}, \tag{30b}$$

$$(28c)-(28d),(28f)-(28h),$$

Note that this problem cannot be directly solved in its present form, rather a series of power minimization problems has to be solved with the same target rate for all users, where the corresponding target rate is updated in the next iteration according to the feasibility condition of the power minimization problem [41]. The feasibility of the following power minimization problem is examined for a given target SINR t at all users in each iteration of the bisection search (29).

Problem P_4 can be formulated as a standard SOCP. More precisely, for a given t, Problem P_4 can be reformulated as Problem P_5 in 30), where $\mathbf{z}_{lj} \triangleq \left[N\mathbf{v}_{lj,1}^T \ \sqrt{N}\mathbf{v}_{lj,2}^T \ \frac{1}{\sqrt{\rho_d}}\right]^T$, and $\mathbf{v}_{lj,1} = \left[\varrho_{l1j}\cdots\varrho_{lk-1j}\right]^T$, $\mathbf{v}_{lj,2} = [\sqrt{\beta_{1lj}}\nu_1\cdots\sqrt{\beta_{Mlj}}\nu_M]^T$, $\tilde{\mathbf{z}}_{lk} \triangleq \left[N\tilde{\mathbf{v}}_{lk,1}^T \ \sqrt{N}\tilde{\mathbf{v}}_{lk,2}^T \ \frac{1}{\sqrt{\rho_d}}\right]^T$, and $\tilde{\mathbf{v}}_{lk,1} = [\tilde{\varrho}_{l1k}\cdots\tilde{\varrho}_{lk-1k}]^T$, $\tilde{\mathbf{v}}_{lk,2} = \left[\sqrt{\beta_{1lk}}\tilde{\nu}_1\cdots\sqrt{\beta_{Mlk}}\tilde{\nu}_M\right]^T$. It can be seen that (30) represents a second order cone (SOC) [41]. Hence, Problem P_5 is a standard SOCP, which is a convex problem. The bisection search method imposed for maximizing the donwlink max-min SINR is exploited to find the optimal solution [41]. In this bisection search approach, first the upper and lower bounds of the achievable SINR are set to $t_{\max}$ and $t_{\min}$, respectively and the initial target SINR t is chosen as $(t_{\max}+t_{\min})/2$. If Problem P_5 is feasible for a given target SINR t, then the lower bound $t_{\min}$ will be set to t and a new target SINR is chosen as $(t_{\max}+t_{\min})/2$ for the next iteration. This procedure is continued until the difference between the upper and the lower bounds becomes smaller than a predefined threshold ϵ. This bisection based search method based aided max-min DL BE scheme is summarized in Algorithm 1, which provides the optimal solution [41].

Algorithm 1 Bisection search method to solve Problem P_1

1. Initialize $t_{\min}$, $t_{\max}$ and ϵ
2. repeat
3. Solve Problem P_5, with $t = \frac{t_{\max}+t_{\min}}{2}$
4. **if** Problem P_5 is feasible, **then** $t_{\min} = t$
5. **else**, $t_{\max} = t$
6. until $(t_{\max}-t_{\min}) \leq \epsilon$

B. Max-Min BE of Normalized Conjugate Beamforming

This section investigates the max-min DL BE of normalized conjugate beamforming. The max-min DL BE problem can be formulated under per-AP power constraints as follows:

$$P_6: \max_{\mu_{mlk}} \quad \min_{\{k=1\cdots K, l=1\cdots L\}} \quad \mathrm{S}_{lk}^{lk,\text{final,NCB}} \tag{31a}$$

$$\text{s.t.} \quad \sum_{l=1}^{L}\sum_{k=1}^{K}\mu_{mlk} \leq 1, \forall m, \quad \mu_{mlk} \geq 0, \quad \forall m,\forall l,\forall k. \tag{31b}$$

Problem P_6 may also be expressed in an equivalent form as Problem P_7 in (32). By introducing new slack variables

$$P_7: \max_{\mu_{mlk}} \quad t \tag{32a}$$

$$\text{s.t.} \frac{\alpha\left(\sum_{m=1}^{M}\sqrt{\mu_{mlk}}\frac{\beta_{mlj}}{\sigma_{ml}}\right)^2}{\alpha\left[\sum\limits_{k'\neq k}^{K}\left(\sum\limits_{m=1}^{M}\frac{\sqrt{\mu_{mlk'}}\beta_{mlj}}{\beta_{mlk'}}\right)^2-\sum\limits_{k'\neq k}^{K}\sum\limits_{m=1}^{M}\frac{\mu_{mlk'}\beta_{mlj}^2}{\beta_{mlk'}^2}\right]+N\sum\limits_{l'\neq l}^{L}\sum\limits_{k'=1}^{K}\sum\limits_{m=1}^{M}\mu_{ml'k'}\beta_{mlj}-\alpha\sum\limits_{m=1}^{M}\mu_{mlk}\frac{\beta_{mlj}}{\sigma_{ml}}+\frac{1}{\rho_d}}\geq t,\forall j<k, \tag{32b}$$

$$\frac{\alpha\left(\sum_{m=1}^{M}\sqrt{\mu_{mlk}}\frac{\beta_{mlk}}{\sigma_{ml}}\right)^2}{\alpha\left[\sum\limits_{k'\neq k}^{K}\left(\sum\limits_{m=1}^{M}\frac{\sqrt{\mu_{mlk'}}\beta_{mlk}}{\beta_{mlk'}}\right)^2-\sum\limits_{k'\neq k}^{K}\sum\limits_{m=1}^{M}\frac{\mu_{mlk'}\beta_{mlk}^2}{\beta_{mlk'}^2}\right]+N\sum\limits_{l'\neq l}^{L}\sum\limits_{k'=1}^{K}\sum\limits_{m=1}^{M}\mu_{ml'k'}\beta_{mlk}-\alpha\sum\limits_{m=1}^{M}\mu_{mlk}\frac{\beta_{mlk}}{\sigma_{ml}}+\frac{1}{\rho_d}}\geq t,\forall k, \tag{32c}$$

$$\sum_{l=1}^{L}\sum_{k=1}^{K}\mu_{mlk}\leq 1,\ \forall m,\ \mu_{mlk}\geq 0,\quad \forall m,\forall l,\forall k. \tag{32d}$$

$$P_8: \max_{w_{mlk}} \quad t \tag{33a}$$

$$\text{s.t.} \frac{\alpha\mathbf{w}_{lk}^T\left(\mathbf{\Delta}_{lj}\right)\mathbf{w}_{lk}}{\alpha\sum\limits_{k'\neq k}^{K}\mathbf{w}_{lk}^T\mathbf{\Lambda}_{ljk'}\mathbf{w}_{lk'}-\alpha\sum\limits_{k'\neq k}^{K}\mathbf{w}_{lk'}^T\mathbf{\Upsilon}_{ljk'}\mathbf{w}_{lk'}+N\sum\limits_{l'\neq 1}^{L}\sum\limits_{k'=1}^{K}\mathbf{w}_{l'k}\mathbf{B}_{lj}\mathbf{w}_{l'k'}-\alpha\mathbf{w}_{lk}^T\mathbf{D}_{lj}\mathbf{w}_{lk}+\frac{1}{\rho_d}}\geq t,\ \forall j<k, \tag{33b}$$

$$\frac{\alpha\mathbf{w}_{lk}^T\left(\mathbf{\Delta}_{lk}\right)\mathbf{w}_{lk}}{\alpha\sum\limits_{k'\neq k}^{K}\mathbf{w}_{lk}^T\mathbf{\Lambda}_{lkk'}\mathbf{w}_{lk'}-\alpha\sum\limits_{k'\neq k}^{K}\mathbf{w}_{lk'}^T\mathbf{\Upsilon}_{lkk'}\mathbf{w}_{lk'}+N\sum\limits_{l'\neq 1}^{L}\sum\limits_{k'=1}^{K}\mathbf{w}_{l'k}\mathbf{B}_{lk}\mathbf{w}_{l'k'}-\alpha\mathbf{w}_{lk}^T\mathbf{D}_{lk}\mathbf{w}_{lk}+\frac{1}{\rho_d}}\geq t, \tag{33c}$$

$$\sum_{l=1}^{L}\sum_{k=1}^{K}w_{mlk}^2\leq 1,\ \forall m,\quad w_{mlk}\geq 0,\quad \forall m,\forall l,\forall k, \tag{33d}$$

$w_{mlk}=\sqrt{\mu_{mlk}}$, we have Problem P_8, defined in (33), where

$$\mathbf{\Delta}_{lj}=\mathbf{\Gamma}_j\mathbf{\Gamma}_j^T,\text{where } \mathbf{\Gamma}_j=\left[\frac{\beta_{1lj}}{\sigma_{1l}}\frac{\beta_{2lj}}{\sigma_{2l}}\cdots\frac{\beta_{Mlj}}{\sigma_{Ml}}\right]^T, \tag{34a}$$

$$\mathbf{\Lambda}_{ljk'}=\boldsymbol{\theta}_{ljk'}\boldsymbol{\theta}_{ljk'}^T,\ \text{where } \boldsymbol{\theta}_{ljk'}=\left[\frac{\beta_{1lj}}{\sigma_{1l}}\frac{\beta_{2lj}}{\sigma_{2l}}\cdots\frac{\beta_{Mlj}}{\sigma_{Ml}}\right]^T, \tag{34b}$$

$$\mathbf{\Upsilon}_{ljk'}=\text{diag}\left[\frac{\beta_{1lj}^2}{\sigma_{1l}^2}\frac{\beta_{2lj}^2}{\sigma_{2l}^2}\cdots\frac{\beta_{Mlj}^2}{\sigma_{Ml}^2}\right], \tag{34c}$$

$$\mathbf{B}_{lj}=\text{diag}\left[\beta_{1lk}\beta_{2lj}\cdots\beta_{Mlj}\right], \tag{34d}$$

$$\mathbf{D}_{lk}=\text{diag}\left[\frac{\beta_{1lj}^2}{\sigma_{1l}^2}\frac{\beta_{2lj}^2}{\sigma_{2l}^2}\cdots\frac{\beta_{Mlj}^2}{\sigma_{Ml}^2}\right], \tag{34e}$$

$$\mathbf{\Delta}_{lk}=\mathbf{\Gamma}_k\mathbf{\Gamma}_k^T,\text{where } \mathbf{\Gamma}_k=\left[\frac{\beta_{1lk}}{\sigma_{1l}}\frac{\beta_{2lk}}{\sigma_{2l}}\cdots\frac{\beta_{Mlk}}{\sigma_{Ml}}\right]^T, \tag{34f}$$

$$\mathbf{\Lambda}_{lkk'}=\boldsymbol{\theta}_{lkk'}\boldsymbol{\theta}_{lkk'}^T,\text{where } \boldsymbol{\theta}_{lkk'}=\left[\frac{\beta_{1lk}}{\sigma_{1l}}\frac{\beta_{2lk}}{\sigma_{2l}}\cdots\frac{\beta_{Mlk}}{\sigma_{Ml}}\right]^T, \tag{34g}$$

$$\mathbf{\Upsilon}_{lkk'}=\text{diag}\left[\frac{\beta_{1lk}^2}{\sigma_{1l}^2}\frac{\beta_{2lk}^2}{\sigma_{2l}^2}\cdots\frac{\beta_{Mlk}^2}{\sigma_{Ml}^2}\right], \tag{34h}$$

$$\mathbf{B}_{lk}=\text{diag}\left[\beta_{1lk}\beta_{2lk}\cdots\beta_{Mlk}\right], \tag{34i}$$

$$\mathbf{D}_{lk}=\text{diag}\left[\frac{\beta_{1lk}^2}{\sigma_{1l}^2}\frac{\beta_{2lk}^2}{\sigma_{2l}^2}\cdots\frac{\beta_{Mlk}^2}{\sigma_{Ml}^2}\right]. \tag{34j}$$

Then the classic bisection search method can be used for finding the optimal solution of the original Problem P_8 by iteratively solving the power minimization problem P_9 for a given target SINR t at all users. Problem P_9 is defined as P_9 in (35). Note that due to the negative terms in the denominator in (35a), Problem P_9 cannot be formulated as a standard SOCP, as it was in Problem P_5. However, in the following, we impose semidefinite relaxation (SDR) for solving this non-convex problem [42]. First, we introduce the new variable $\mathbf{W}_{lk}=\mathbf{w}_{lk}\mathbf{w}_{lk}^T$, which enables us to reformulate Problem P_9 into a standard SDP using SDR. By utilising the identity $\mathbf{w}^T\mathbf{R}\mathbf{w}=\text{Tr}\left[\mathbf{R}\mathbf{w}\mathbf{w}^T\right]=\text{Tr}\left[\mathbf{R}\mathbf{W}\right]$, Problem P_9 can be rewritten as P_{10} in (36).

Note that again $\mathbf{W}_{lk}\succeq 0$ means that $\mathbf{W}_{lk}$ is a positive semidefinite matrix. By relaxing all rank-one constraints in Problem P_{10}, we arrive at a standard SDP, which can be optimally solved by convex optimization software. In particular, if the solutions of P_{10} are rank-one matrices (i.e., $\text{rank}[\mathbf{W}_{lk}]=1,\forall l,k$), then it is the solution to P_9. Otherwise, the randomization techniques of [42] can be utilised for determining a set of rank-one solutions. Using [42, Theorem 3.1], if P_{10} is feasible, then it has at least one solution with $\text{rank}[\mathbf{W}_{lk}]=1,\forall l,k$. The power allocation $\mathbf{w}_{lk}$ can be determined from a rank-one $\mathbf{W}_{lk}$ solution using $\mathbf{w}_{lk}=\sqrt{\lambda_{lk}}\mathbf{u}_{lk}$ where $\sqrt{\lambda_{lk}}$ and $\mathbf{u}_{lk}$ denote the maximum eigenvalue and the corresponding eigenvector of $\mathbf{W}_{lk}$, respectively. This bisection

$$P_9: \min_{w_{mlk}} \quad \sum_{l'=1}^{L}\sum_{k'=1}^{K} \mathbf{w}_{lk}^T \mathbf{w}_{lk} \tag{35a}$$

$$\text{s.t.} \quad \frac{\alpha \mathbf{w}_{lk}^T \mathbf{\Delta}_{lj} \mathbf{w}_{lk}}{\alpha \sum_{k'\neq k}^{K} \mathbf{w}_{lk}^T \mathbf{\Delta}_{ljk'} \mathbf{w}_{lk'} - \alpha \sum_{k'\neq k}^{K} \mathbf{w}_{lk}^T \mathbf{\Upsilon}_{ljk'} \mathbf{w}_{lk'} - \alpha \sum_{k''\neq k}^{K} \mathbf{w}_{lk''}^T \mathbf{\Delta}_{ljk''} \mathbf{w}_{lk''} + N \sum_{l'=1}^{L}\sum_{k'=1}^{K} \mathbf{w}_{l'k} \mathbf{B}_{lj} \mathbf{w}_{l'k'} - \alpha \mathbf{w}_{lk}^T \mathbf{D}_{lj} \mathbf{w}_{lk} + \frac{1}{\rho_d}} \geq t, \forall l, \forall k, \forall j < k, \tag{35b}$$

$$\frac{\alpha \mathbf{w}_{lk}^T \mathbf{\Delta}_{lk} \mathbf{w}_{lk}}{\alpha \sum_{k'\neq k}^{K} \mathbf{w}_{lk}^T \mathbf{\Delta}_{lkk'} \mathbf{w}_{lk'} - \alpha \sum_{k'\neq k}^{K} \mathbf{w}_{lk}^T \mathbf{\Upsilon}_{lkk'} \mathbf{w}_{lk'} - \alpha \sum_{k''\neq k}^{K} \mathbf{w}_{lk''}^T \mathbf{\Delta}_{lkk''} \mathbf{w}_{lk''} + N \sum_{l'=1}^{L}\sum_{k'=1}^{K} \mathbf{w}_{l'k'} \mathbf{B}_{lk} \mathbf{w}_{l'k'} - \alpha \mathbf{w}_{lk}^T \mathbf{D}_{lk} \mathbf{w}_{lk} + \frac{1}{\rho_d}} \geq t, \forall l, k, \tag{35c}$$

$$\sum_{l=1}^{L}\sum_{k=1}^{K} w_{mlk}^2 \leq 1, \ \forall m, \ w_{mlk} \geq 0, \quad \forall m, \forall l, \forall k. \tag{35d}$$

$$P_{10}: \min_{\mathbf{W}_{lk}} \quad \sum_{l'=1}^{L}\sum_{k'=1}^{K} \mathrm{Tr}\left[\mathbf{W}_{lk}\right] \tag{36a}$$

$$\text{s.t.} \quad \alpha \mathrm{Tr}\left[\mathbf{\Delta}_{lj} \mathbf{W}_{lk}\right] - t\alpha \sum_{k'\neq k}^{K} \mathrm{Tr}\left[\mathbf{\Delta}_{ljk'} \mathbf{W}_{lk'} + \alpha \sum_{k'\neq k}^{K} \mathrm{Tr}\left[\mathbf{\Upsilon}_{ljk'} \mathbf{W}_{lk'}\right]\right] + \alpha \sum_{k''\neq k}^{K} \mathrm{Tr}\left[\mathbf{\Delta}_{ljk''} \mathbf{W}_{lk''}\right]$$
$$- N \sum_{l'=1}^{L}\sum_{k'=1}^{K} \mathrm{Tr}\left[\mathbf{B}_{lj} \mathbf{W}_{l'k'}\right] + \alpha \mathrm{Tr}\left[\mathbf{D}_{lj} \mathbf{W}_{lk}\right] k \geq \frac{t}{\rho_d}, \forall l, \forall k, \forall j < k, \tag{36b}$$

$$\alpha \mathrm{Tr}\left[\mathbf{\Delta}_{lk} \mathbf{W}_{lk}\right] - t\alpha \sum_{k'\neq k}^{K} \mathrm{Tr}\left[\mathbf{\Delta}_{lkk'} \mathbf{W}_{lk'} + \alpha \sum_{k'\neq k}^{K} \mathrm{Tr}\left[\mathbf{\Upsilon}_{lkk'} \mathbf{W}_{lk'}\right]\right] + \alpha \sum_{k''\neq k}^{K} \mathrm{Tr}\left[\mathbf{\Delta}_{lkk''} \mathbf{W}_{lk''}\right]$$
$$- N \sum_{l'=1}^{L}\sum_{k'=1}^{K} \mathrm{Tr}\left[\mathbf{B}_{lk} \mathbf{W}_{l'k'}\right] + \alpha \mathrm{Tr}\left[\mathbf{D}_{lk} \mathbf{W}_{lk}\right] \geq \frac{t}{\rho_d}, \forall l, \forall k, \tag{36c}$$

$$\sum_{l=1}^{L}\sum_{k=1}^{K} w_{mlk} \leq 1, \ \forall m, \quad \mathbf{W}_{lk} = \mathbf{W}_{lk}^T, \ \forall l, \forall k, \mathbf{W}_{lk} \succeq 0, \ \forall l, \forall k, \quad \mathrm{rank}\left[\mathbf{W}_{lk}\right] = 1, \quad \forall l, \forall k. \tag{36d}$$

Algorithm 2 Bisection search method to solve Problem P_6

1. Initialize $t_{\min}$, $t_{\max}$ and ϵ
2. repeat
3. Solve Problem P_{10}, with $t = \frac{t_{\max}+t_{\min}}{2}$
4. **if** Problem P_{10} is feasible, **then** $t_{\min} = t$
5. **else**, $t_{\max} = t$
6. until $(t_{\max} - t_{\min}) \leq \epsilon$

method based on SDP is imposed for finding the optimal solution of Problem P_6. Based on the analysis in [41, Chapter 4], we set the threshold ϵ used for terminating Algorithms 1 and 2 to a small value. The iterative procedure of these algorithms is terminated, once the difference between the upper and lower bounds of the SINR becomes lower than the predefined threshold ϵ. In the simulations, we set $\epsilon = 0.0001$, hence the difference between the optimal solution and the solution achieved by the proposed schemes is less than 0.0001.

V. Complexity analysis

In this section, we provide the computational complexity analysis for the proposed schemes with conjugate beamforming and normalized conjugate beamforming.

A. Computational Complexity of Solving Problem P_1

The iterative bisection search method solves a SOCP at each iteration with $O\left(n_v^2 n_c\right)$ arithmetic operations, where n_v is the number of optimization variables and n_c refers to the total number of SOC constraints [43]. Moreover, note that the total number of iterations required is given by $\log_2(\frac{t_{\max}-t_{\min}}{\epsilon})$ [41]. In Problem P_5, the total number of variables is $n_v = LMK$, and there are $n_c = LK$ SOC constraints. Hence, the number of arithmetic operations required for for solving Problem P_1 is $\log_2(\frac{t_{\max}-t_{\min}}{\epsilon}) \times O\left(M^2 L^3 K^3\right)$.

B. Computational Complexity of of Solving Problem P_6

The complexity of solving a SDP under n_c semidefinite constraints and a $n_d \times n_d$ dimensional semidefinite cone is $O(n_c n_d^3 + n_c^2 n_d^2 + n_c^3)$ [44]. In Problem P_{10}, we have $n_c = LK$

semidefinite constraints and the dimension of the semidefinite cone is $M \times M$ (i.e., $n_d = M$). The bisection search scheme solves a SDP at each iteration. Hence, based on [41], the computational complexity of solving Problem P_6 is $\log_2(\frac{t_{\max}-t_{\min}}{\epsilon}) \times \left(LKM^3 + L^2K^2M^2 + (LK)^3\right)$.

VI. Numerical Results and Discussion

In this section, we provide numerical results to validate the performance of the proposed max-min rate scheme relying on different parameters. A cell-free massive MIMO system relying on M APs and K_{tot} single-antenna users is considered in a $D \times D$ coverage area, where both APs and users are uniformly distributed at random. In the following subsections, we define the simulation parameters and then present the corresponding simulation results.

A. Simulation Parameters

The channel coefficients between users and APs are modelled in Section II, where the coefficient β_{mlk} is given by $\beta_{mlk} = \text{PL}_{mlk} 10^{\frac{\sigma_{sh} z_{mkl}}{10}}$, and PL_{mlk} is the path loss from the kth user in the lth cluster to the mth AP, while the second term $10^{\frac{\sigma_{sh} z_{mlk}}{10}}$ denotes the shadow fading with standard deviation $\sigma_{sh} = 8$ dB, and $z_{mlk} \sim \mathcal{N}(0,1)$ [15]. In the simulations, an uncorrelated shadowing model and a three-slope path loss model similar to [15] are considered. The noise power is given by $p_n = \text{BW} \times k_B \times T_0 \times W$, where BW = 20 MHz denotes the bandwidth, $k_B = 1.381 \times 10^{-23}$ represents the Boltzmann constant, and $T_0 = 290$ (Kelvin) is the noise temperature. Moreover, the $W = 9$dB, and denotes the noise figure. It is assumed that $\bar{p}_p$ and $\bar{\rho}_d$ denote the power of the pilot sequence and of the DL data, respectively, where $\rho_p = \dfrac{\bar{\rho}_p}{p_n}$ and $\rho_d = \dfrac{\bar{\rho}_d}{p_n}$. In our simulations, we set $\bar{\rho}_p = 100$ mW and $\bar{\rho_d} = 200$ mW. Similar to [15], we assume that the simulation area is wrapped around at the edges, which can simulate an area without boundaries. Hence, the square simulation area has eight neighbours. We evaluate the max-min BE of the system over 200 random realizations of the locations of APs, users and shadow fading. Finally, note that after solving the max-min optimization problem, all users will achieve the same per-user throughput.

B. Clustering Schemes

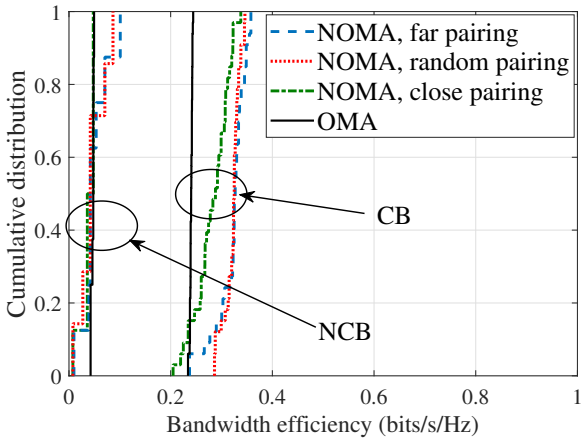
Since there are a large number of users in massive MIMO systems it is not practically feasible to simultaneously accommodate all users in a single cluster, while performing SIC at the users' ends [6], [36], [45]. Note that in [37], [38], [46], [47], the authors proposed to group the users into small clusters, where NOMA was employed within each cluster with the aid of SIC. To reduce the computational complexity introduced by SIC at the users' ends, the users can be divided into multiple clusters, and the NOMA principle can be employed within each clusters [4]. Based on [48], user pairing is a key technique in NOMA systems, which facilitates the practical implementation of NOMA for many users by reducing the complexity of SIC. Hence, following a similar approach to those in [4], [48], [49], we propose to pair the users into clusters. However, in contrast to [37], [38], the CSI is not available at the CPU and the receiver. Hence, it is not possible to exploit the user pairing schemes proposed in [37], [38]. The investigation of clustering algorithms with more than two users per cluster will be considered in our future work. Three different user pairing schemes are compared. Based on [37], the computational complexity of optimal clustering in the DL NOMA system is extremely high, and therefore not suitable for practical implementation in real-time systems. In [6], [32], the authors propose to pair the users based on their position. The clustering schemes proposed in this paper are different from those of [6], [32]. We propose three different clustering schemes as follows: 1) In the first scheme, the users who have the smallest distance from each other are paired. We continue to pair the closest users until all the users are grouped into clusters; 2) In the second scheme, the users who have the largest distance from each other, are paired. We continue to pair the farthest users until all the users are grouped into clusters; 3) Finally, we compare these schemes to the choice of randomly pairing the users.

C. Mode Selection

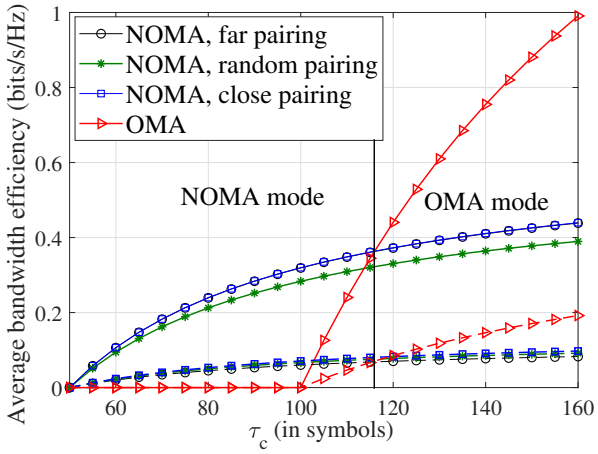
For OMA, there is no pilot contamination thanks to the orthogonal pilots assigned to the users. However, the length of pilot sequences has to be equal to the total number of users ($\tau_c \geq K_{\text{tot}}$). But long pilot sequences leave less time for data transmission, hence reducing the overall throughput. On the other hand, in NOMA, the length of pilots only has to be equal to the number of clusters. Having two users per-cluster, we get $\tau_c \geq K_{\text{tot}}/2$. Hence, compared to OMA, there is more time left for payload data transmission. Moreover, within each cluster, the user having a higher received signal power performs SIC, while the other users still suffers from some residual inter-user interference. As a result, the system performance can be improved by switching between OMA and NOMA modes, depending both on the number of users and on the length of the channel's coherence time. The mode set is defined as Mode = {OMA, NOMA}. Then, the aim of the proposed design is to select the optimal mode maximizing the throughput of the system.

D. Simulation Results

In this subsection, we evaluate the performance of the proposed DL max-min BE schemes. First, to assess the performance, a NOMA-based cell-free massive MIMO system relying on 20 APs ($M = 20$) and supporting 100 users ($K_{\text{tot}} = 100$) is considered who are randomly distributed over the coverage area of size 1×1 km^2 and $\tau_c = 110$. Fig. 2a presents the cumulative distribution of the achievable DL BE of both conjugate and of the normalized conjugate beamforming techniques, where the schemes proposed in Subsection IV-A and IV-B are used for determining the max-min bandwidth efficiency, respectively. As seen in Fig. 2a, in the NOMA scheme, random clustering has the better performance, while the scenario in which the far users are clustered has a performance as good as random clustering.

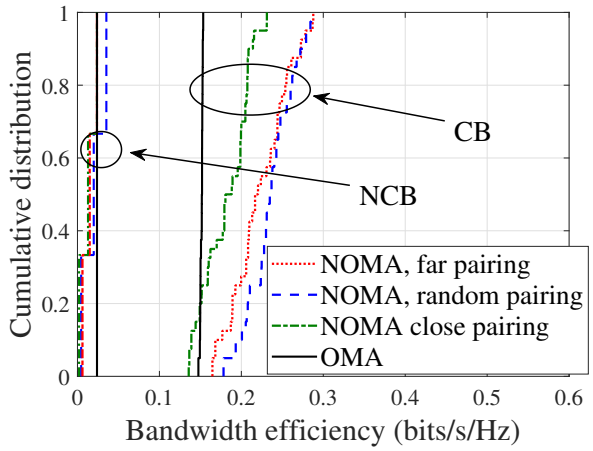


(a) The cumulative distribution of the per-user DL BE of cell-free massive MIMO with per-AP power constraints where CB and NCB refers to the conjugate beamforming and normalized conjugate beamforming techniques.

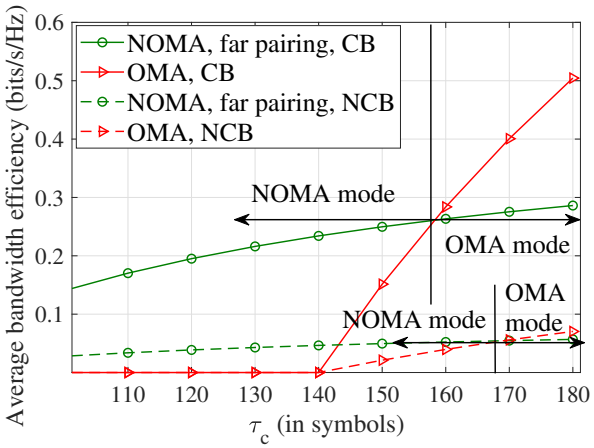


(b) The average DL BE of cell-free massive MIMO versus coherence time (in symbols). We solve the max-min BE problems with per-AP power constraints. The simulation parameters are the same as Fig. 2a.

Figure 2. We set $M = 20$, $N = 15$, $K = 2$, $K_{\text{tot}} = 100$, $\tau_c = 110$, $D = 1$ km, $\bar{\rho}_p = 100$ mW and $\bar{\rho}_d = 200$ mW.



(a) The cumulative distribution of the per-user DL BE of cell-free massive MIMO with per-AP power constraints where CB and NCB refers to the conjugate beamforming and normalized conjugate beamforming techniques.



(b) The average DL BE of cell-free massive MIMO versus coherence time (in symbols). We solve the max-min BE problems with per-AP power constraints. The simulation parameters are the same as Fig. 3a.

Figure 3. We set $M = 20$, $N = 15$, $K = 2$, $K_{\text{tot}} = 140$, $\tau_c = 150$, $D = 1$ km, $\bar{\rho}_p = 100$ mW and $\bar{\rho}_d = 200$ mW.

By contrast, the scenario supporting users that are clustered in each others proximity has poor performance. As the figure shows, the performance of conjugate beamforming is superior to that normalized conjugate beamforming. Moreover, NOMA outperforms OMA using conjugate beamforming, whereas the performance gap between NOMA and OMA with normalized conjugate beamforming is small. Given the above-mentioned parameters, as Fig. 2a demonstrates, NOMA-based cell-free massive MIMO provides a better performance than that of the system using OMA in terms of its per-user BE. Next, to investigate the effect of the channel's coherence time, the average BE of cell-free massive MIMO is plotted versus τ_c in Fig. 2b, while the other system parameters are the same as those used in Fig. 2b. Based on the results of Fig. 2b, one could find an optimal mode switching point depending on the length of the channel's coherence time to maximize the system performance. As Fig. 2b shows by increasing the length of the channel's coherence time, OMA outperforms NOMA in terms of its user BE. Let us now assume that there are $K_{\text{tot}} = 140$ users in the area and $M = 20$ APs are uniformly distributed, where each AP has $N = 15$ antennas with $\tau_c = 150$. The cumulative distribution of the achievable DL BE with conjugate beamforming and normalized conjugate beamforming is plotted in Fig. 3a. As the figure shows, the average DL BE of the cell-free massive MIMO system relying on conjugate beamforming is better than that of the system having normalized conjugate beamforming. Next, Fig. 3b quantifies the average BE (having solved the max-min bandwidth efficiency problems) versus the channel's coherence time τ_c. As shown in Fig. 3b, $\tau_c = 158$ is the mode switching point between OMA and NOMA for improving the throughput of conjugate beamforming, while for normalized conjugate beamforming $\tau_c = 140$ is the mode switching point. As it is stated in Subsection VI-C, when the number of users is

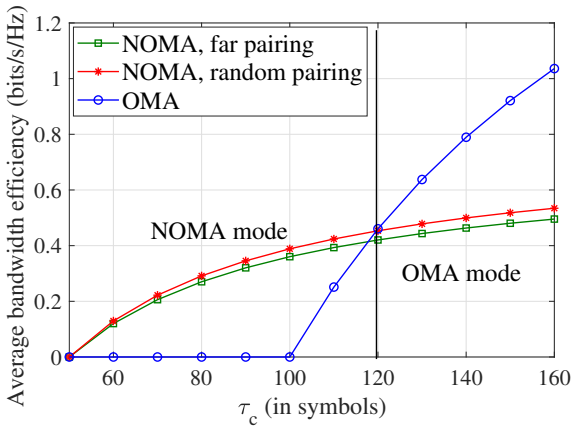


Figure 4. The average DL BE of cell-free massive MIMO versus coherence time (in symbols) with conjugate beamforming. We solve the max-min BE problems with per-AP power constraints. We set $M = 60$, $N = 5$, $K = 2$, $K_{\text{tot}} = 100$, $D = 1$ km, $\bar{\rho}_p = 100$ mW and $\bar{\rho}_d = 200$ mW.

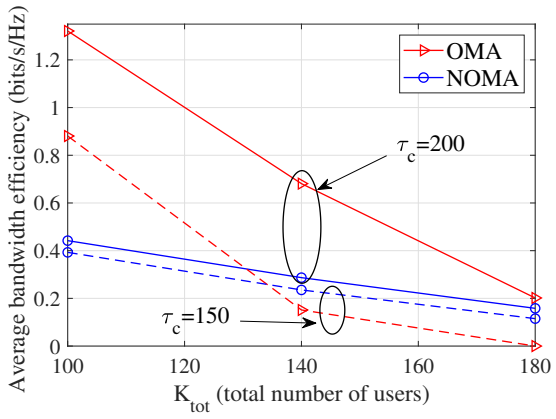


Figure 5. The average DL BE of cell-free massive MIMO versus total number of users in the area with conjugate beamforming. We solve the max-min BE problems with per-AP power constraints for two values of coherence time, i.e., $\tau_c = 150$ and $\tau_c = 200$. Note that for the NOMA scheme, we exploit random pairing approach. Moreover, we set $M = 20$, $N = 15$, $K = 2$, $D = 1$ km, $\bar{\rho}_p = 100$ mW and $\bar{\rho}_d = 200$ mW.

high, the time left for payload data transmission is short, hence resulting in low BE.

Next, we consider conjugate beamforming in conjunction with $M = 60$ APs, each equipped with $N = 5$ antennas, where there are $K_{\text{tot}} = 100$ users in the area. Fig. 4 investigates the average BE to solve the max-min BE problem of conjugate beamforming. The mode switching point is $\tau_c = 120$.

We investigate the impact of the total number of users in the area on the system's performance. In Fig. 5, we consider conjugate beamforming having $M = 20$, $N = 15$, and two different coherence times, namely $\tau_c = 150, 200$. As indicated in the figure, the solid and dashed lines denote $\tau_c = 200$ and $\tau_c = 150$, respectively. As Fig. 5 shows, by increasing the total number of users, the NOMA scheme outperforms the OMA scheme in terms of the average DL BE. Finally, we investigate the signal and interference power of both conjugate beamforming and of normalized conjugate beamforming. Based on the analysis in [35], and to draw a fair comparison,

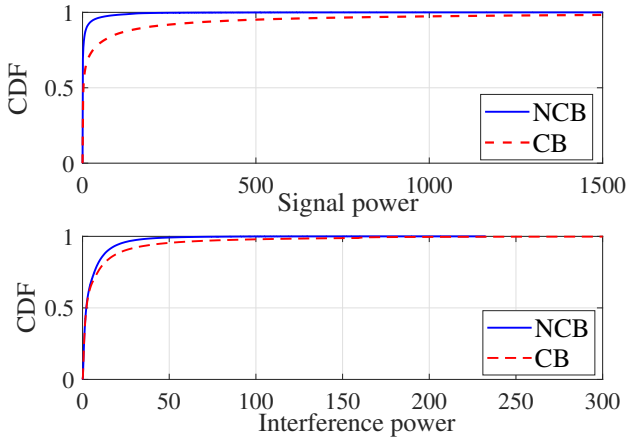


Figure 6. CDF of DL signal power and DL interference power of cell-free massive MIMO for the OMA scheme with equal power allocation, i.e., $\mu_{mk} = \frac{\gamma_{mk}}{N\sum_{k'=1}^{K}\gamma_{mk'}}$ and $\eta_{mk} = \frac{1}{N\sum_{k'=1}^{K}\gamma_{mk'}}$, for NCB and CB, respectively. We set $M = 20$ APs, each equipped with $N = 15$ antennas, where there are $K_{\text{tot}} = 100$ users in the area. Moreover, note that $D = 1$ km, $\bar{\rho}_p = 100$ mW and $\bar{\rho}_d = 200$ mW.

Algorithm 3 *Proposed Scheme for the Cell-free massive MIMO Operation*

1. *Initialize K_{tot}, τ_c, N, M, D, ρ_p and ρ_d.*
2. *Solve the optimization Problems P_1 and P_6 using Algorithms 1 and 2.*
3. *Calculate the average DL BE of the system with optimal power elements obtained in Step 2 using (24).*
4. *Choose the optimal operating mode from the set Mode = { OMA, NOMA }, and define the operating region. An example of the operating region is provided in Fig. 7.*
5. *Set the optimal operational mode based on using the operating region.*

we set $\mu_{mk} = \frac{\gamma_{mk}}{N\sum_{k'=1}^{K}\gamma_{mk'}}$ and $\eta_{mk} = \frac{1}{N\sum_{k'=1}^{K}\gamma_{mk'}}$ for normalized conjugate beamforming and conjugate beamforming, respectively. Fig. 6 shows the cumulative distribution function (CDF) of the signal and interference power in the DL of cell-free massive MIMO for the OMA scheme associated with $M = 20$, $N = 15$, $K_{\text{tot}} = 100$, $D = 1$ km, $\bar{\rho}_p = 100$ mW and $\bar{\rho}_d = 200$ mW. As Fig. 6 demonstrates, the normalized conjugate beamforming introduces a similar interference power to that of the conjugate beamforming, however at a lower signal power. Furthermore, based on the results in Fig. 6, one could conclude that the signal power of cell-free massive MIMO is much higher than the interference power for both conjugate and normalized conjugate beamforming. This explains the performance gap between the systems using conjugate and normalized conjugate beamforming.

E. Proposed Off-line Method for Cell-free Massive MIMO to Achieve the Best System Performance

Since there are many system parameters, we do not provide any analysis to determine the optimal mode switching point.

K_{tot} \ τ_c	110	114	120	130	140	150	158	160	170	180	190	196	200
180	N	N	N	N	N	N	N	N	N	N	N	N=O	O
140	N	N	N	N	N	N	N=O	O	O	O	O	O	O
100	N	N=O	O	O	O	O	O	O	O	O	O	O	O

Figure 7. Operating region for different modes with different total number of users K_{tot} and coherence time τ_c (in symbols). Here, we set M = 20, N = 15, D = 1 km, $\bar{\rho}_p$ = 100 mW and $\bar{\rho}_d$ = 200 mW. Note that the letters N and O refer to Mode = NOMA and Mode = OMA, respectively. Moreover N=O refers to the case where both systems with the NOMA and OMA schemes show the same performance in terms of the system DL BE.

These parameters include: the number of users K_{tot} and APs M, the number of antennas per-AP N, the channel coherence time τ_c, and the size of area D. It is set aside for future research to find the optimal switching point in terms of the above-mentioned parameters. The numerical results show that we require a high-flexibility system to switch between the NOMA and OMA modes, depending on the system parameters. Similar conclusions have been reached in [6], [36], where the authors indicated that a massive MIMO system having collocated antennas and switching between the OMA and NOMA operating modes has the best performance. Finding the optimal switching point between the OMA and NOMA modes was done based on off-line simulations in [6]. In bold contrast to [6], we propose an off-line algorithm to achieve the best performance in the context of cell-free massive MIMO systems. In the proposed scheme, we determine the average DL BE of the system for the given parameters and find the optimum operating region, which maximises the average DL BE from the set Mode = { OMA, NOMA }. Next, the operational mode is fixed for a given set of parameters. To provide an example, we present the operating region of the pair of different modes for different values of K_{tot} as well as τ_c and for fixed values of M, N, and D in Fig. 7, where the symbols N and O refer to the NOMA and OMA mode, respectively. Moreover, N=O represents the scenario where the performance of the NOMA mode matches that of the OMA mode. Hence, given the operating region seen in Fig. 7, the wireless operator could fix the operational mode based on the values of K_{tot} and τ_c. The details of proposed scheme are summarized in Algorithm 3.

Note that we have to run Algorithm 3 for every system parameters to define the operating region. A practical technique of implementing this is considering a user-birth/death process. In this case, a Markov model allows the system's evaluation from supporting K_{tot} users to either supporting $(K_{\text{tot}}+1)$ or $(K_{\text{tot}}-1)$ users on a near-instantaneous basis [50], which will be considered in our future work.

VII. Conclusions

We have considered a NOMA-based cell-free massive MIMO system relying on both conjugate and normalized conjugate beamforming techniques, in which the users are grouped into clusters. In the NOMA technique, the non-orthogonality is due to assigning the same pilots to users within the same cluster. Moreover, it is assumed that orthogonal pilots are assigned to different clusters avoiding inter-cluster interference. We have also assumed there is no DL training, since the SIC must be performed only relying on the statistics of the channel, which imposes errors on the received signals. A closed-form expression of the BE has been derived. We have then studied the problem of max-min BE under per-AP power constraints, where the minimum BE of all users is maximized. We have developed a SOCP and used SDP to efficiently solve the non-convex optimization problems of conjugate beamforming and of normalized conjugate beamforming, respectively. Additionally, the complexity of the proposed schemes has been investigated. Finally, we have investigated the effect of the channel's coherence time and of the length of pilots on the system's performance and proposed an OMA/NOMA mode switching scheme for maximizing the average per-user BE of the system assuming the max-min optimization. Based on the numerical results, the switching point depends both on the channel's coherence time and on the total number of users.

Appendix A: Proof of Theorem 1

The desired signal of the kth user in the lth cluster is given by

$$\begin{aligned}
&\text{DS}_{lk} \\
&= \sqrt{\rho_d}\mathbb{E}\left\{\sum_{m=1}^{M}\eta_{mlk}\mathbf{g}_{mlk}^{T}c_{ml}\left(\sqrt{\tau_p\rho_p}\sum_{k'=1}^{K}\mathbf{g}_{mlk'}+\mathbf{W}_{p,m}\boldsymbol{\phi}_{lk}\right)^{*}\right\} \\
&= N\sqrt{\rho_d}\sum_{m=1}^{M}\sqrt{\eta_{mlk}}\gamma_{ml}\frac{\beta_{mlk}}{\sum_{i=1}^{K}\beta_{mli}}. \qquad (37)
\end{aligned}$$

Hence, $|\text{DS}_{lk}|^2 = N^2\left(\sum_{m=1}^{M}\sqrt{\rho_d\eta_{mlk}}\frac{\gamma_{ml}}{\sum_{i=1}^{K}\beta_{mli}}\beta_{mlj}\right)^2$. Moreover, the term $\mathbb{E}\{|\text{BU}_k|^2\}$ can be obtained as

$$\begin{aligned}
&\mathbb{E}\left\{|\text{BU}_{lk}|^2\right\} \\
&= \rho_d\mathbb{E}\left\{\left|\sum_{m=1}^{M}\sqrt{\eta_{mlk}}\mathbf{g}_{mlk}^{T}\hat{\mathbf{f}}_{ml}^{*}-\mathbb{E}\left\{\sum_{m=1}^{M}\sqrt{\eta_{mlk}}\mathbf{g}_{mlk}^{T}\hat{\mathbf{f}}_{ml}^{*}\right\}\right|^2\right\} \\
&= \rho_d\sum_{m=1}^{M}\eta_{mlk}\left(\mathbb{E}\left\{\left|\mathbf{g}_{mlk}^{T}\hat{\mathbf{f}}_{ml}^{*}-\mathbb{E}\left\{\mathbf{g}_{mlk}^{T}\hat{\mathbf{f}}_{ml}^{*}\right\}\right|^2\right\}\right) \\
&= \rho_d\sum_{m=1}^{m}\eta_{mlk}\left(\underbrace{\mathbb{E}\left\{\left|\mathbf{g}_{mlk}^{T}\hat{\mathbf{f}}_{ml}^{*}\right|^2\right\}}_{I_1}-\left|\underbrace{\mathbb{E}\left\{\mathbf{g}_{mlk}^{T}\hat{\mathbf{f}}_{ml}^{*}\right\}}_{I_2}\right|^2\right), \qquad (38)
\end{aligned}$$

where

$$I_1 = \mathbb{E}\left\{\left|\mathbf{g}_{mlk}^T c_{ml}\left(\sqrt{\tau_p\rho_p}\sum_{i=1}^{K}\mathbf{g}_{mli}+\mathbf{W}_{p,m}\boldsymbol{\phi}_{lk}\right)^*\right|^2\right\}$$
$$= \underbrace{\tau_p\rho_p c_{ml}^2\mathbb{E}\left\{|\mathbf{g}_{mlk}|^4\right\}}_{\tau_1}+\underbrace{\tau_p\rho_p c_{ml}^2\mathbb{E}\left\{\left|\sum_{i\neq k}^{K}\mathbf{g}_{mli}^*\right|^2\right\}}_{\tau_2}+\tau_3, \quad (39)$$

where $\tau_1 = \tau_p\rho_p c_{ml}^2 N(N+1)\beta_{mlk}^2$, $\tau_2 = N\beta_{mlk}\gamma_{ml} - Nc_{ml}^2\beta_{mlk} - N\tau_p\rho_p c_{ml}^2\beta_{mlk}^2$ and $\tau_3 = Nc_{ml}^2\beta_{mlk}$. Hence, $\mathbb{E}\left\{|\mathrm{BU}_{lk}|^2\right\} = \rho_d N\sum_{m=1}^{M}\eta_{mlk}\beta_{mlk}\gamma_{ml}$. Next, $\mathbb{E}\left\{|\mathrm{IUI}_{lk'}|^2\right\}$ is calculated as follows:

$$\mathbb{E}\left\{|\mathrm{IUI}_{lk'}|^2\right\} = \rho_d\mathbb{E}\left\{\left|\sum_{m=1}^{M}\sqrt{\eta_{mlk'}}\mathbf{g}_{mlk}^T\hat{\mathbf{f}}_{ml}^*\right|^2\right\}$$
$$= \rho_d\mathbb{E}\left\{\left|\sum_{m=1}^{M}\sqrt{\eta_{mlk'}}\mathbf{g}_{mlk}^T c_{ml}\left(\sqrt{\tau_p\rho_p}\sum_{i=1}^{K}\mathbf{g}_{mli}+\underbrace{\mathbf{W}_{p,m}\boldsymbol{\phi}_{lk}}_{\tilde{\mathbf{w}}_{mlk}}\right)^*\right|^2\right\}$$
$$= \underbrace{\rho_d\,\mathbb{E}\left\{\left|\sum_{m=1}^{M}\sqrt{\eta_{mlk'}}c_{ml}\mathbf{g}_{mlk}^T\tilde{\mathbf{w}}_{mlk}^*\right|^2\right\}}_{A}$$
$$+\underbrace{\rho_d\,\tau_p\rho_p\mathbb{E}\left\{\left|\sum_{m=1}^{M}\sqrt{\eta_{mlk'}}c_{ml}\mathbf{g}_{mlk}^T\left(\sum_{i=1}^{K}\mathbf{g}_{mli}^*\right)\right|^2\right\}}_{B}, \quad (40)$$

where the third equality in (40) is due to the fact that for two independent random variables X and Y and $\mathbb{E}\{X\} = 0$, we have $\mathbb{E}\left\{|X+Y|^2\right\} = \mathbb{E}\left\{|X|^2\right\} + \mathbb{E}\left\{|Y|^2\right\}$ [15]. Since $\tilde{\mathbf{w}}_{mlk} = \boldsymbol{\phi}_{lk}^H\mathbf{W}_{\mathbf{p},\mathbf{m}}$ is independent from the term $\mathbf{g}_{mlk}$, the term A in (40) immediately is given by $A = N\sum_{m=1}^{M}\eta_{mlk'}c_{ml}^2\beta_{mlk}$. The term B in (40) can be obtained as

$$B = \underbrace{\tau_p\rho_p\mathbb{E}\left\{\left|\sum_{m=1}^{M}\sqrt{\eta_{mlk'}}c_{ml}||\mathbf{g}_{mlk}||^2\right|^2\right\}}_{C}$$
$$+\underbrace{\tau_p\rho_p\mathbb{E}\left\{\left|\sum_{m=1}^{M}\sqrt{\eta_{mlk'}}c_{ml}\mathbf{g}_{mlk}^T\left(\sum_{i\neq k}^{K}\mathbf{g}_{mli}^*\right)^H\right|^2\right\}}_{D}. \quad (41)$$

The first term in (41) is given by

$$C = \tau_p\rho_p\mathbb{E}\left\{\left|\sum_{m=1}^{M}\sqrt{\eta_{mlk'}}c_{ml}||\mathbf{g}_{mlk}||^2\right|^2\right\}$$
$$= \mathbb{E}\left\{\sum_{m=1}^{M}\sum_{n=1}^{M}\sqrt{\eta_{mlk'}}\sqrt{\eta_{nlk'}}c_{ml}c_{nl}||\mathbf{g}_{mlk}||^2||\mathbf{g}_{nlk}||^2\right\}$$
$$= \tau_p\rho_p N(N+1)\sum_{m=1}^{M}\eta_{mlk'}c_{ml}^2\beta_{mlk}^2$$
$$+\tau_p\rho_p N^2\sum_{m=1}^{M}\sum_{n\neq m}^{M}\sqrt{\eta_{mlk'}}\sqrt{\eta_{nlk'}}c_{ml}c_{nl}\beta_{mlk}\beta_{nlk}$$
$$-N\tau_p\rho_p\sum_{m=1}^{M}\eta_{mlk'}c_{ml}^2\beta_{mlk}^2+N^2\left(\sum_{m=1}^{M}\eta_{mlk'}\gamma_{ml}\frac{\beta_{mlk}}{\sum_{i=1}^{K}\beta_{mli}}\right)^2. \quad (42)$$

The second term in (41) can be obtained as

$$D = \tau_p\rho_p\mathbb{E}\left\{\left|\sum\nolimits_{m=1}^{M}\sqrt{\eta_{mlk'}}c_{ml}\mathbf{g}_{mlk}^T\left(\sum\nolimits_{i\neq k}^{K}\mathbf{g}_{mli}^*\right)^H\right|^2\right\}$$
$$= N\tau_p\rho_p\sum\nolimits_{m=1}^{M}\sum\nolimits_{i\neq k}^{K}\eta_{mlk'}c_{ml}^2\beta_{mlk}\beta_{mli}$$
$$= N\sqrt{\tau_p\rho_p}\sum_{m=1}^{M}\eta_{mlk'}c_{ml}\beta_{mlk}\beta_{mlk'}$$
$$-N\sum_{m=1}^{M}\eta_{mlk'}c_{ml}^2\beta_{mlk}-N\tau_p\sum_{m=1}^{M}\eta_{mlk'}c_{ml}^2\beta_{mlk}. \quad (43)$$

Finally, we obtain

$$\mathbb{E}\{|\mathrm{IUI}_{lk'}|^2\} = N\rho_d\sum_{m=1}^{M}\eta_{mlk'}\beta_{mlk}\gamma_{ml}$$
$$+N^2\rho_d\left(\sum_{m=1}^{M}\sqrt{\eta_{mlk'}}\gamma_{ml}\frac{\beta_{mlk}}{\sum_{i=1}^{K}\beta_{mli}}\right)^2. \quad (44)$$

In the next step, we calculate the term $\mathbb{E}\left\{|\mathrm{ICI}_{l'k'}|^2\right\}$ as follows:

$$\mathbb{E}\left\{|\mathrm{ICI}_{l'k'}|^2\right\} = \mathbb{E}\left\{\left|\sum_{m=1}^{M}\eta_{ml'k'}\mathbf{g}_{mlk}^T\hat{\mathbf{f}}_{ml'}^*\right|^2\right\}$$
$$= \rho_d\mathbb{E}\left\{\left|\sum_{m=1}^{M}\sqrt{\eta_{ml'k'}}c_{ml'}\mathbf{g}_{mlk}^T\left(\sqrt{\tau_p\rho_p}\sum_{i=1}^{K}\mathbf{g}_{ml'i}+\tilde{\mathbf{w}}_{ml'k'}\right)^*\right|^2\right\}$$
$$= \underbrace{\rho_d\,\mathbb{E}\left\{\left|\sum_{m=1}^{M}\sqrt{\eta_{ml'k}}c_{ml'}\mathbf{g}_{mlk'}^T\tilde{\mathbf{w}}_{ml'k'}^*\right|^2\right\}}_{E}$$
$$+\underbrace{\rho_d\,\tau_p\rho_p\mathbb{E}\left\{\left|\sum_{m=1}^{M}\sqrt{\eta_{ml'k'}}c_{ml'}\mathbf{g}_{mlk}^T\left(\sum_{i=1}^{K}\mathbf{g}_{ml'i}^*\right)\right|^2\right\}}_{F}, \quad (45)$$

where the term E is obtained by $E = N\sum_{m=1}^{M}\eta_{ml'k'}c_{ml'}^2\beta_{mlk}$, and the term F can be calculated as

$$\begin{aligned}F &= \sqrt{\tau_p}N\sum_{m=1}^{M}\eta_{ml'k'}c_{ml'}\beta_{mlk}\beta_{ml'k'} - N\sum_{m=1}^{M}\eta_{ml'k'}c_{ml'}^2\beta_{mlk}\\ &= N\sum_{m=1}^{M}\eta_{ml'k'}\beta_{mlk}\gamma_{ml'} - N\sum_{m=1}^{M}\eta_{ml'k'}c_{ml'}^2\beta_{mlk}. \end{aligned} \quad (46)$$

Finally, we have

$$\mathbb{E}\left\{|\mathrm{ICI}_{l'k'}|^2\right\} = N\rho_d\sum_{m=1}^{M}\eta_{ml'k'}\beta_{mlk}\gamma_{ml'}. \quad (47)$$

Next, the term $\mathbb{E}\{|\mathrm{ISIC}_{lk''}|^2\}$ is obtained as follows:

$$\begin{aligned}E\left\{|\mathrm{ISIC}_{lk''}|^2\right\} &= \rho_d\underbrace{\mathbb{E}\left\{\left|\sum_{m=1}^{M}\sqrt{\eta_{mlk''}}\mathbf{g}_{mlk'}^T\hat{\mathbf{f}}_{ml}^*\right|^2\right\}}_{I_1}\\ &\quad - \rho_d\underbrace{\left|\mathbb{E}\left\{\sum_{m=1}^{M}\sqrt{\eta_{mlk''}}\mathbf{g}_{mlk}^T\mathbf{f}_{ml}^*\right\}\right|^2}_{I_2}, \end{aligned} \quad (48)$$

where the term I_1 is obtained by

$$I_1 = N^2\left(\sum_{m=1}^{M}\sqrt{\eta_{mlk''}}\gamma_{ml}\frac{\beta_{mlk}}{\sum_{i=1}^{K}\beta_{mli}}\right)^2 + N\sum_{m=1}^{M}\eta_{mlk''}\beta_{mlk}\gamma_{ml}, \quad (49)$$

and

$$\begin{aligned}I_2 &= \sum_{m=1}^{M}\sqrt{\eta_{mlk''}}\mathbb{E}\left\{c_{ml}\mathbf{g}_{mlk}^T\left(\sqrt{\tau_p\rho_p}\sum_{i=1}^{K}\mathbf{g}_{mli} + \mathbf{w}_{lk}\right)\right\}\\ &\quad - N\sum_{m=1}^{M}\sqrt{\eta_{mlk''}}\gamma_{ml}\frac{\beta_{mlk}}{\sum_{i=1}^{K}\beta_{mli}}. \end{aligned} \quad (50)$$

Using (48), (49) and (50), we have

$$\mathbb{E}\left\{|\mathrm{ISIC}_{lk''}|^2\right\} = N\rho_d\sum_{m=1}^{M}\eta_{mlk''}\beta_{mlk}\gamma_{ml}. \quad (51)$$

This completes the proof of Theorem 1. ∎

Appendix B: Proof of Theorem 3

The desired signal for the kth user in the lth cluster can be derived as follows:

$$\begin{aligned}\mathrm{DS}_{lk} &= \sqrt{\rho_d}\mathbb{E}\left\{\sum_{m=1}^{M}\mu_{mlk}\mathbf{g}_{ml}^T\frac{\hat{\mathbf{f}}_{ml}^*}{||\hat{\mathbf{f}}_{ml}||}\right\}\\ &= \sqrt{\rho_d}\sum_{m=1}^{M}\mu_{mlk}\mathbb{E}\left\{\mathbf{g}_{ml}^T\frac{\left(c_{ml}\left(\sqrt{\tau_p\rho_p}\sum_{k'=1}^{K}\mathbf{g}_{mlk'} + \mathbf{W}_{p,m}\boldsymbol{\phi}_{lk}\right)\right)^*}{||c_{ml}\left(\sqrt{\tau_p\rho_p}\sum_{k'=1}^{K}\mathbf{g}_{mlk'} + \mathbf{W}_{p,m}\boldsymbol{\phi}_{lk}\right)||}\right\}\\ &\overset{(\tilde{a})}{\approx} \sqrt{\rho_d}\sum_{m=1}^{M}\mu_{mlk}\frac{\mathbb{E}\left\{\mathbf{g}_{ml}^T\left(\sqrt{\tau_p\rho_p}\sum_{k'=1}^{K}\mathbf{g}_{mlk'} + \mathbf{W}_{p,m}\boldsymbol{\phi}_{lk}\right)^*\right\}}{\mathbb{E}\left\{||\sqrt{\tau_p\rho_p}\sum_{k'=1}^{K}\mathbf{g}_{mlk'} + \mathbf{W}_{p,m}\boldsymbol{\phi}_{lk}||\right\}}\\ &= N\sqrt{\tau_p\rho_p}\sum_{m=1}^{M}\sqrt{\mu_{mlk}}\frac{\beta_{mlk}}{\sigma_{ml}}, \end{aligned} \quad (52)$$

where in step $(\tilde{a})$, we use the analysis in [35]. Moreover, note that using [51, Corollary 1] we have:

$$\sigma_{ml} = \int_0^{\infty}\sqrt{x}\sum_{i=1}^{K+1}\left\{\prod_{j=1,j\neq i}^{K+1}\left(1 - \frac{\varpi_j}{\varpi_i}\right)^{-1}\right\}\frac{e^{-\frac{x}{\varpi_i}}}{\varpi_i}dx, \forall l, \quad (53)$$

where $\varpi_i = N\tau_p\rho_p\beta_{mli}, 1 \leq i \leq K$ and $\varpi_{K+1} = N$. For the case of $K = 2$, the term σ_{ml} is given in (21). Hence, we have $|\mathrm{DS}_{lk}|^2 = \rho_d\tau_p\rho_p N^2\left(\sum_{m=1}^{M}\sqrt{\mu_{mlk}}\frac{\beta_{mlk}}{\sigma_{ml}}\right)^2$. Next, the term $\mathbb{E}\{|\mathrm{BU}_k|^2\}$ is calculated as follows:

$$\begin{aligned}&\mathbb{E}\left\{|\mathrm{BU}_{lk}|^2\right\}\\ &= \rho_d\mathbb{E}\left\{\left|\sum_{m=1}^{M}\sqrt{\mu_{mlk}}\mathbf{g}_{mlk}^T\frac{\hat{\mathbf{f}}_{ml}^*}{||\hat{\mathbf{f}}_{ml}||} - \mathbb{E}\left\{\sum_{m=1}^{M}\sqrt{\mu_{mlk}}\mathbf{g}_{mlk}^T\frac{\hat{\mathbf{f}}_{ml}^*}{||\hat{\mathbf{f}}_{ml}||}\right\}\right|^2\right\}\\ &= \rho_d\sum_{m=1}^{M}\mu_{mlk}\left(\underbrace{\mathbb{E}\left\{\left|\mathbf{g}_{mlk}^T\frac{\hat{\mathbf{f}}_{ml}^*}{||\hat{\mathbf{f}}_{ml}||}\right|^2\right\}}_{I_3} - \left|\underbrace{\mathbb{E}\left\{\mathbf{g}_{mlk}^T\frac{\hat{\mathbf{f}}_{ml}^*}{||\hat{\mathbf{f}}_{ml}||}\right\}}_{I_4}\right|^2\right), \end{aligned} \quad (54)$$

where

$$I_3 = \mathbb{E}\left\{\left|\mathbf{g}_{mlk}^T\frac{\hat{\mathbf{f}}_{ml}^*}{||\hat{\mathbf{f}}_{ml}||}\right|^2\right\} = \mathbb{E}\left\{||\mathbf{g}_{mlk}||^2\right\} = N\beta_{mlk}, \quad (55)$$

$$I_4 = \sqrt{\tau_p\rho_p}N\left(\sum_{m=1}^{M}\sqrt{\mu_{mlk}}\frac{\beta_{mlk}}{\sigma_{ml}}\right). \quad (56)$$

As a result, we have

$$\mathbb{E}\left\{|\mathrm{BU}_{lk}|^2\right\} = N\sum_{m=1}^{M}\mu_{mlk}\beta_{mlk} - N^2\tau_p\rho_p\sum_{m=1}^{M}\mu_{mlk}\gamma_{ml}. \quad (57)$$

Next, $\mathbb{E}\left\{|\mathrm{IUI}_{lk'}|^2\right\}$ is calculated as follows:

$$\begin{aligned}\mathbb{E}\{|\mathrm{IUI}_{lk'}|^2\} &= \rho_d\mathbb{E}\left\{\left|\sum_{m=1}^{M}\sqrt{\mu_{mlk'}}\mathbf{g}_{mlk}^T\frac{\hat{\mathbf{f}}_{ml}^*}{||\hat{\mathbf{f}}_{ml}||}\right|^2\right\}\\ &= \rho_d\mathbb{E}\left\{\sum_{m=1}^{M}\sum_{n=1}^{M}\sqrt{\mu_{mlk'}\mu_{nlk'}}\mathbf{g}_{mlk}^T\mathbf{g}_{nlk}\frac{\hat{\mathbf{f}}_{ml}^*}{||\hat{\mathbf{f}}_{ml}||}\frac{\hat{\mathbf{f}}_{nl}}{||\hat{\mathbf{f}}_{nl}||}\right\}\\ &= N^2\tau_p\rho_d\sum_{m=1}^{M}\sum_{n\neq m}^{M}\frac{\sqrt{\mu_{mlk'}\mu_{nlk'}}\beta_{mlk}\beta_{nlk}}{\sigma_{ml}\sigma_{nl}} + N\sum_{m=1}^{M}\mu_{mlk'}\beta_{mlk}. \end{aligned} \quad (58)$$

Next, we calculate the term $\mathbb{E}\left\{|\mathrm{ICI}_{l'k'}|^2\right\}$ as follows:

$$\begin{aligned}\mathbb{E}\left\{|\mathrm{ICI}_{l'k'}|^2\right\} &= \mathbb{E}\left\{\left|\sum_{m=1}^{M}\mu_{ml'k'}\mathbf{g}_{mlk}^T\frac{\hat{\mathbf{f}}_{ml'}^*}{||\hat{\mathbf{f}}_{ml'}^*||}\right|^2\right\}\\ &= \mathbb{E}\left\{\sum_{m=1}^{M}\sum_{n=1}^{M}\sqrt{\mu_{ml'k'}\mu_{nl'k'}}\mathbf{g}_{mlk}^T\mathbf{g}_{nlk}^*\frac{\hat{\mathbf{f}}_{ml'}^*}{||\hat{\mathbf{f}}_{ml'}^*||}\frac{\hat{\mathbf{f}}_{nl'}^*}{||\hat{\mathbf{f}}_{nl'}^*||}\right\}\\ &= \mathbb{E}\left\{\sum_{m=1}^{M}\sqrt{\mu_{ml'k'}}\mathbf{g}_{mlk}^T\mathbf{g}_{mlk}^*\right\}\\ &\quad + \sum_{m=1}^{M}\sum_{n\neq m}^{M}\sqrt{\mu_{ml'k'}\mu_{nl'k'}}\mathbb{E}\left\{\mathbf{g}_{mlk}^T\frac{\hat{\mathbf{f}}_{ml'}^*}{||\hat{\mathbf{f}}_{ml'}^*||}\mathbf{g}_{nlk}^T\frac{\hat{\mathbf{f}}_{nl'}^*}{||\hat{\mathbf{f}}_{nl'}^*||}\right\}\\ &\overset{(b)}{=} N\sum_{m=1}^{M}\mu_{ml'k'}\beta_{mlk}, \end{aligned} \quad (59)$$

where step (b) comes from the following fact:

$$\mathbb{E}\left\{\mathbf{g}_{mlk}^T \frac{\hat{\mathbf{f}}_{ml'}^*}{||\hat{\mathbf{f}}_{ml'}^*||}\mathbf{g}_{nlk}^T \frac{\hat{\mathbf{f}}_{nl'}^*}{||\hat{\mathbf{f}}_{nl'}^*||}\right\} \mathbb{E}\left\{\mathbf{g}_{mlk}^T \frac{\hat{\mathbf{f}}_{ml'}^*}{||\hat{\mathbf{f}}_{ml'}^*||}\right\}\mathbb{E}\left\{\mathbf{g}_{nlk}^T \frac{\hat{\mathbf{f}}_{nl'}^*}{||\hat{\mathbf{f}}_{nl'}^*||}\right\}, \quad (60)$$

and $\mathbb{E}\left\{\mathbf{g}_{mlk}^T \hat{\mathbf{f}}_{ml'}^*\right\} = c_{ml'k'}\mathbb{E}\left\{\mathbf{g}_{mlk}^T\left(\sqrt{\tau_p\rho_p}\sum_{i=1}^K \mathbf{g}_{ml'i} + \mathbf{W}_{p,m}\boldsymbol{\phi}_{l'k'}\right)\right\} = 0$. In the following, we calculate $\mathbb{E}\{|\mathrm{ISIC}_{lk''}|^2\}$;

$$\mathbb{E}\left\{|\mathrm{ISIC}_{lk''}|^2\right\} = \rho_d \underbrace{\mathbb{E}\left\{\left|\sum_{m=1}^M \sqrt{\mu_{mlk''}}\mathbf{g}_{mlk}^T \frac{\hat{\mathbf{f}}_{ml}^*}{||\hat{\mathbf{f}}_{ml}||}\right|^2\right\}}_{I_5} - \rho_d \underbrace{\left|\mathbb{E}\left\{\sum_{m=1}^M \sqrt{\mu_{mlk''}}\mathbf{g}_{mlk}^T \frac{\mathbf{f}_{ml}^*}{||\hat{\mathbf{f}}_{ml}||}\right\}\right|^2}_{I_6}, \quad (61)$$

where $I_5 = N^2\tau_p\rho_d\sum_{m=1}^M\sum_{n\neq m}^M \frac{\sqrt{\mu_{mlk''}\mu_{nlk''}}\beta_{mlk}\beta_{nlk}}{\sigma_{ml}\sigma_{nl}} + N\sum_{m=1}^M \mu_{mlk''}\beta_{mlk}$ & $I_6 = \sum_{m=1}^M \sqrt{\mu_{mlk''}}\frac{\beta_{mlk}}{\sigma_{ml}}$. Finally, we have

$$\mathbb{E}\left\{|\mathrm{ISIC}_{lk''}|^2\right\} = N^2\tau_p\rho_d\sum_{m=1}^M\sum_{n\neq m}^M \frac{\sqrt{\mu_{mlk'}\mu_{nlk''}}\beta_{mlk'}\beta_{nlk''}}{\sigma_{ml}\sigma_{nl}} + N\sum_{m=1}^M \mu_{mlk''}\beta_{mlk} - N^2\tau_p\rho_d\left(\sum_{m=1}^M \frac{\sqrt{\mu_{mlk''}}\beta_{mlk}}{\sigma_{ml}}\right)^2. \quad (62)$$

This completes the proof of Theorem 3. ∎

References

[1] M. Bashar, K. Cumanan, A. G. Burr, , L. H. H. Q. Ngo, and P. Xiao, "NOMA/OMA mode selection-based cell-free massive MIMO," in *Proc. IEEE ICC*, May 2019, pp. 1–7.

[2] Y. Liu, Z. Qin, M. Elkashlan, Z. Ding, A. Nallanathan, and L. Hanzo, "Nonorthogonal multiple access for 5G and beyond," *IEEE Trans. Wireless Commun.*, vol. 105, no. 12, pp. 2347–2381, Dec. 2017.

[3] Y. Liu, H. Xing, C. Pan, A. Nallanathan, M. Elkashlan, and L. Hanzo, "Multiple-antenna-assisted non-orthogonal multiple access," *IEEE Trans. Wireless Commun.*, vol. 25, no. 2, pp. 17–23, Apr. 2018.

[4] Z. Ding, P. Fan, and H. V. Poor, "Impact of user pairing on 5G non-orthogonal multiple-access downlink transmissions," *IEEE Trans. Veh. Technol.*, vol. 85, no. 5, pp. 6010–6023, Aug. 2016.

[5] P. Xu and K. Cumanan, "Optimal power allocation scheme for non-orthogonal multiple access with α -fairness," *IEEE J. Sel. Areas Commun.*, vol. 35, no. 10, pp. 2357–2369, Oct. 2017.

[6] 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.

[7] G. Wunder, P. Jung, M. Kasparick, T. Wild, F. Schaich, Y. Chen, S. T. Brink, I. Gaspar, N. Michailow, A. Festag, L. Mendes, N. Cassiau, D. Ktenas, M. Dryjanski, S. Pietrzyk, B. Eged, P. Vago, and F. Wiedmann, "5GNOW: non-orthogonal, asynchronous waveforms for future mobile applications," *IEEE Commun. Mag.*, vol. 52, no. 2, pp. 97–105, Feb. 2014.

[8] Z. Ding, Z. Yang, P. Fan, and H. V. Poor, "On the performance of non-orthogonal multiple access in 5G systems with randomly deployed users," *IEEE Signal Process. Lett.*, vol. 21, no. 12, pp. 917–920, Dec. 2014.

[9] R. Zhang and L. Hanzo, "A unified treatment of superposition coding aided communications: Theory and practice," *IEEE Commun. Surveys Tutorials*, vol. 13, no. 3, pp. 503–520, Mar. 2011.

[10] F. Boccardi, R. W. Heath, A. Lozano, T. L. Marzetta, and P. Popovski, "Five disruptive technology directions for 5G," *IEEE Commun. Mag.*, vol. 52, no. 2, pp. 74–80, Feb. 2014.

[11] H. Q. Ngo, E. G. Larsson, and T. L. Marzetta, "Energy and spectral efficiency of very large multiuser MIMO systems," *IEEE Trans. Commun.*, vol. 61, no. 4, pp. 1436–1449, Apr. 2013.

[12] T. L. Marzetta, E. G. Larsson, H. Yang, and H. Q. Ngo, *Fundamentals of massive MIMO*. Cambridge University Press, 2016.

[13] Y. Xu, G. Yue, and S. Mao, "User grouping for massive MIMO in FDD systems: New design methods and analysis," *IEEE Trans. Wireless Commun.*, vol. 14, no. 12, pp. 6827–6842, Jul. 2015.

[14] M. Bashar, H. Q. Ngo, K. Cumanan, A. G. Burr, D. Maryopi, and E. G. Larsson, "On the performance of backhaul constrained cell-free massive MIMO with linear receivers," in *Proc. IEEE Asilomar*, 2018, pp. 1–6.

[15] 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.

[16] M. Bashar, K. Cumanan, A. G. Burr, M. Debbah, and H. Q. Ngo, "On the uplink max-min SINR of cell-free massive MIMO systems," *IEEE Trans. Wireless Commun.*, pp. 2021–2036, Jan. 2019.

[17] ——, "Enhanced max-min SINR for uplink cell-free massive MIMO systems," in *Proc. IEEE ICC*, May 2018, pp. 1–6.

[18] M. Bashar, K. Cumanan, A. Burr, H. Q. Ngo, E. Larsson, and P. Xiao, "On the energy efficiency of limited-backhaul cell-free massive MIMO," in *Proc. IEEE ICC*, May 2019, pp. 1–7.

[19] M. Bashar, K. Cumanan, A. G. Burr, H. Q. Ngo, E. G. Larsson, and P. Xiao, "Energy efficiency of the cell-free massive MIMO uplink with optimal uniform quantization," *IEEE Trans. Green Commun. and Net.*, pp. 1–18, To appear.

[20] M. Karakayali, G. Foschini, and R. Valenzuela, "Network coordination for spectrally efficient communications in cellular systems," *IEEE Trans. Wireless Commun. Mag.*, vol. 13, no. 4, pp. 56–61, Aug. 2006.

[21] E. Björnson, R. Zakhour, D. Gesbert, and B. Ottersten, "Cooperative multicell precoding: rate region characterization and distributed strategies with instantaneous and statistical CSI," *IEEE Trans. Signal Process.*, vol. 58, no. 8, pp. 4298–4310, Aug. 2010.

[22] A. G. Burr, M. Bashar, and D. Maryopi, "Cooperative access networks: Optimum fronthaul quantization in distributed massive MIMO and cloud RAN," in *Proc. IEEE VTC*, Jun. 2018, pp. 1–7.

[23] A. Burr, M. Bashar, and D. Maryopi, "Ultra-dense radio access networks for smart cities: Cloud-RAN, Fog-RAN and cell-free Massive MIMO," in *Proc. IEEE PIMRC*, Sep. 2018, pp. 1–5.

[24] D. Maryopi, M. Bashar, and A. Burr, "On the uplink throughput of zero forcing in cell-free massive MIMO with coarse quantization," *IEEE Trans. Veh. Technol.*, vol. 68, no. 7, pp. 7220–7224, Jul. 2019.

[25] M. Bashar, K. Cumanan, A. G. Burr, H. Q. Ngo, and M. Debbah, "Cell-free massive MIMO with limited backhaul," in *Proc. IEEE ICC*, May 2018, pp. 1–7.

[26] M. Bashar, H. Q. Ngo, A. Burr, D. Maryopi, K. Cumanan, and E. G. Larsson, "On the performance of backhaul constrained cell-free massive MIMO with linear receivers," in *Proc. IEEE Asilomar*, Nov. 2018, pp. 1–7.

[27] M. Bashar, K. Cumanan, A. G. Burr, H. Q. Ngo, M. Debbah, and P. Xiao, "Max-min rate of cell-free massive MIMO uplink with optimal uniform quantization," *IEEE Trans. Commun.*, pp. 1–18, To appear.

[28] M. Bashar, *Cell-free massive MIMO and Millimeter Wave Channel Modelling for 5G and Beyond*. Ph.D. dissertation, University of York, United Kingdom, 2019.

[29] G. Interdonato, E. Bjornson, H. Q. Ngo, P. Frenger, and E. G. Larsson, "Ubiquitous cell-free massive MIMO communications," *EURASIP J. Wireless Commun. Netw*, pp. 1–19, To appear.

[30] T. L. Marzetta, "Noncooperative cellular wireless with unlimited numbers of base station antennas," *IEEE Trans. Commun.*, vol. 9, no. 11, pp. 3590–3600, Nov. 2010.

[31] A. Ashikhmin, T. L. Marzetta, and L. Li, "Interference reduction in multi-cell massive MIMO systems I: Large-scale fading precoding and decoding," *IEEE Trans. Inf. Theory*, vol. 64, no. 9, pp. 6340–6361, Sep. 2018.

[32] Y. Li and G. A. A. Baduge, "NOMA-Aided cell-free massive MIMO systems," *IEEE Wireless Commun. Lett.*, pp. 1–4, 2018.

[33] H. Q. Ngo, L. Tran, T. Q. Duong, M. Matthaiou, and E. G. Larsson, "On the total energy efficiency of cell-free massive MIMO," *IEEE Trans. Green Commun. and Net.*, vol. 2, no. 1, pp. 25–39, Mar. 2017.

[34] 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.

[35] 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 CAMAD*, Oct. 2016, pp. 1–6.

[36] K. Senel, H. V. Cheng, E. Björnson, and E. G. Larsson, "What role can NOMA play in massive MIMO?" *IEEE J. Sel. Topics Signal Process.*, pp. 1–16, Feb. 2019.

[37] S. Ali, E. Hossain, and D. I. Kim, "Non-orthogonal multiple access (NOMA) for downlink multiuser MIMO systems: User clustering, beamforming, and power allocation," *IEEE Access*, vol. 5, pp. 565–577, Mar. 2017.

[38] B. Kimy, S. Lim, H. Kim, S. Suh, J. Kwun, S. Choi, C. Lee, S. Lee, and D. Hong, "Non-orthogonal multiple access in a downlink multiuser beamforming system," in *Proc. IEEE MILCOM*, Nov. 2013, pp. 1–6.

[39] Z. Wei, L. Yang, D. W. K. Ng, J. Yuan, and L. Hanzo, "On the performance gain of NOMA over OMA in uplink communication systems," *[online]. Available: https://arxiv.org/abs/1903.01683*, pp. 1–51, Submitted.

[40] M. F. Hanif, Z. Ding, T. Ratnarajah, and G. K. Karagiannidis, "A minorization-maximization method for optimizing sum rate in the downlink of non-orthogonal multiple access systems," *IEEE Trans. Signal Process.*, vol. 64, no. 1, pp. 76–88, Jan. 2016.

[41] S. Boyd and L. Vandenberghe, *Convex Optimization*. Cambridge, UK: Cambridge University Press, 2004.

[42] M. Bengtsson and B. Ottersten, "Optimal downlink beamforming using semidefinite optimization," in *Proc. IEEE Allerton*, Sep. 1999, pp. 987–996.

[43] M. S. Lobo, L. Vandenberghe, S. Boyd, and H. Lebret, "Applications of second-order cone programming," *Linear Algebra Applicat*, pp. 193–228, Nov. 1998.

[44] I. Polik and T. Terlaky, *Interior Point Methods for Nonlinear Optimization*. New York, NY, USA: Springer, 2010.

[45] Z. Ding, L. Dai, R. Schober, and H. V. Poor, "NOMA meets finite resolution analog beamforming in massive MIMO and millimeter-wave networks," *IEEE Commun. Lett.*, vol. 21, no. 8, pp. 1879–1882, Aug. 2017.

[46] Y. Liu, M. Elkashlan, Z. Ding, and G. K. Karagiannidis, "Fairness of user clustering in MIMO non-orthogonal multiple access systems," *IEEE Commun. Lett.*, vol. 20, no. 7, pp. 1465–1468, Jul. 2016.

[47] M. S. Ali, H. Tabassum, and E. Hossain, "Dynamic user clustering and power allocation for uplink and downlink non-orthogonal multiple access (NOMA) systems," *IEEE Access*, vol. 4, pp. 6325–6343, Aug. 2016.

[48] Z. Ding, M. Peng, and H. V. Poor, "Cooperative non-orthogonal multiple access in 5G systems," *IEEE Commun. Lett.*, vol. 18, no. 8, pp. 1462–1465, Dec. 2015.

[49] M. Salehi, H. Tabassum, and E. Hossain, "Meta distribution of SIR in large-scale uplink and downlink NOMA networks," *IEEE Trans. Commun.*, pp. 1–17, Dec. 2018.

[50] L. Yang and L. Hanzo, "Adaptive rate DS-CDMA systems using variable spreading factors," *IEEE Trans. Veh. Technol.*, vol. 53, no. 1, pp. 72–81, Jan. 2004.

[51] A. Bletsas, H. Shin, and M. Z. Win, "Cooperative communications with outage-optimal opportunistic relaying," *IEEE Trans. Wireless Commun.*, vol. 6, no. 9, pp. 3450–3460, Sep. 2007.

Manijeh Bashar (S'16) received the B.Sc. degree in electrical engineering from the University of Guilan, Iran, in 2009, and the M.Sc. degree in communication systems engineering (with honors) from the Shiraz University of Technology, Iran, in 2013. She received the Ph.D. degree in communications engineering from the University of York, U.K. in 2019. In 2017, she was an Academic Visitor with the Department of Electronics and Nanoengineering, Aalto University, Espoo, Finland, with a Short Term Scientific Mission (STSM) Scholarship Award from European COST-IC1004 "Cooperative Radio Communications for Green Smart Environments".

She is currently a research fellow at the Institute for Communication Systems, home of 5G Innovation Centre (5GIC) at the University of Surrey. Her current research interests include cooperative communications for 5G networks including distributed Massive MIMO, Cloud-RAN, Fog-RAN, NOMA, deep learning, resource allocation, and also millimetre-wave channel modelling.

She received the K. M. Stott Prize for excellent Ph.D. research in electronics engineering from the University of York, U.K. in 2019. She has been awarded First place (based on jury) in the IEEE WCNC'18 three-minute Ph.D. thesis competition for her research in cell-free Massive MIMO. She has been a member of Technical Program Committees for the IEEE ICC 2020.

Dr. Cumanan (M'10) received the BSc degree with first class honors in electrical and electronic engineering from the University of Peradeniya, Sri Lanka in 2006 and the PhD degree in signal processing for wireless communications from Loughborough University, Loughborough, UK, in 2009.

He is currently a lecturer at the Department of Electronic Engineering, The University of York, UK. From March 2012 to November 2014, he was working as a research associate at School of Electrical and Electronic Engineering, Newcastle University, UK. Prior to this, he was with the School of Electronic, Electrical and System Engineering, Loughborough University, UK. In 2011, he was an academic visitor at Department of Electrical and Computer Engineering, National University of Singapore, Singapore. From January 2006 to August 2006, he was a teaching assistant with Department of Electrical and Electronic Engineering, University of Peradeniya, Sri Lanka. His research interests include non-orthogonal multiple access (NOMA), cell-free massive MIMO, physical layer security, cognitive radio networks, convex optimization techniques and resource allocation techniques. He has published more than 80 journal articles and conference papers which attracted more than 1200 Google scholar citations. He has been also recently appointed as an associate editor for IEEE Access journal.

Dr. Cumanan was the recipient of an overseas research student award scheme (ORSAS) from Cardiff University, Wales, UK, where he was a research student between September 2006 and July 2007.

Alister Burr was born in London, U.K, in 1957. He received the BSc degree in Electronic Engineering from the University of Southampton, U.K in 1979 and the PhD from the University of Bristol in 1984. Between 1975 and 1985 he worked at Thorn-EMI Central Research Laboratories in London. In 1985 he joined the Department of Electronics (now Electronic Engineering) at the University of York, U.K, where he has been Professor of Communications since 2000. His research interests are in wireless communication systems, especially MIMO, cooperative systems, physical layer network coding, and iterative detection and decoding techniques. He has published around 250 papers in refereed international conferences and journals, and is the author of "Modulation and Coding for Wireless Communications" (published by Prentice-Hall/PHEI), and co-author of "Wireless Physical-Layer Network Coding (Cambridge University Press, 2018). In 1999 he was awarded a Senior Research Fellowship by the U.K. Royal Society, and in 2002 he received the J. Langham Thompson Premium from the Institution of Electrical Engineers. He has also given more than 15 invited presentations, including three keynote presentations. He was chair, working group 2, of a series of European COST programmes including IC1004 "Cooperative Radio Communications for Green Smart Environments" (which have been influential in 3GPP standardisation), and has also served as Associate Editor for IEEE Communications Letters, Workshops Chair for IEEE ICC 2016, and TPC co-chair for PIMRC 2018.

Hien Quoc Ngo received the B.S. degree in electrical engineering from the Ho Chi Minh City University of Technology, Vietnam, in 2007, the M.S. degree in electronics and radio engineering from Kyung Hee University, South Korea, in 2010, and the Ph.D. degree in communication systems from Linköping University (LiU), Sweden, in 2015. In 2014, he visited the Nokia Bell Labs, Murray Hill, New Jersey, USA. From January 2016 to April 2017, Hien Quoc Ngo was a VR researcher at the Department of Electrical Engineering (ISY), LiU. He was also a Visiting Research Fellow at the School of Electronics, Electrical Engineering and Computer Science, Queen's University Belfast, UK, funded by the Swedish Research Council.

Hien Quoc Ngo is currently a Lecturer at Queen's University Belfast, UK. His main research interests include massive (large-scale) MIMO systems, cell-free massive MIMO, physical layer security, and cooperative communications. He has co-authored many research papers in wireless communications and co-authored the Cambridge University Press textbook *Fundamentals of Massive MIMO* (2016).

Dr. Hien Quoc Ngo received the IEEE ComSoc Stephen O. Rice Prize in Communications Theory in 2015, the IEEE ComSoc Leonard G. Abraham Prize in 2017, and the Best PhD Award from EURASIP in 2018. He also received the IEEE Sweden VT-COM-IT Joint Chapter Best Student Journal Paper Award in 2015. He was an *IEEE Communications Letters* exemplary reviewer for 2014, an *IEEE Transactions on Communications* exemplary reviewer for 2015, and an *IEEE Wireless Communications Letters* exemplary reviewer for 2016. He was awarded the UKRI Future Leaders Fellowship in 2019. Dr. Hien Quoc Ngo currently serves as an Editor for the IEEE Wireless Communications Letters, Digital Signal Processing, and IEICE Transactions on Fundamentals of Electronics, Communications and Computer Sciences. He was a Guest Editor of IET Communications, special issue on "Recent Advances on 5G Communications" and a Guest Editor of IEEE Access, special issue on "Modelling, Analysis, and Design of 5G Ultra-Dense Networks", in 2017. He has been a member of Technical Program Committees for several IEEE conferences such as ICC, GLOBECOM, WCNC, and VTC.

Lajos Hanzo (http://www-mobile.ecs.soton.ac.uk) FREng, FIEEE, FIET, Fellow of EURASIP, DSc received his degree in electronics in 1976 and his doctorate in 1983. In 2009 he was awarded an honorary doctorate by the Technical University of Budapest and in 2015 by the University of Edinburgh. In 2016 he was admitted to the Hungarian Academy of Science. During his 40-year career in telecommunications he has held various research and academic posts in Hungary, Germany and the UK. Since 1986 he has been with the School of Electronics and Computer Science, University of Southampton, UK, where he holds the chair in telecommunications. He has successfully supervised 120+ PhD students, co-authored 18 John Wiley/IEEE Press books on mobile radio communications totalling in excess of 10 000 pages, published 1850+ research contributions at IEEE Xplore, acted both as TPC and General Chair of IEEE conferences, presented keynote lectures and has been awarded a number of distinctions. Currently he is directing an academic research team, working on a range of research projects in the field of wireless multimedia communications sponsored by industry, the Engineering and Physical Sciences Research Council (EPSRC) UK, the European Research Council's Advanced Fellow Grant and the Royal Society's Wolfson Research Merit Award. He is an enthusiastic supporter of industrial and academic liaison and he offers a range of industrial courses. He is also a Governor of the IEEE ComSoc and VTS. During 2008 - 2012 he was the Editor-in-Chief of the IEEE Press and a Chaired Professor also at Tsinghua University, Beijing. For further information on research in progress and associated publications please refer to http://www-mobile.ecs.soton.ac.uk

Pei Xiao Pei Xiao is a professor of Wireless Communications at the Institute for Communication Systems, home of 5G Innovation Centre (5GIC) at the University of Surrey. He is the technical manager of 5GIC, leading the research team in the new physical layer work area, and coordinating/supervising research activities across all the work areas within 5GIC (www.surrey.ac.uk/5gic/research). Prior to this, he worked at Newcastle University and Queen's University Belfast. He also held positions at Nokia Networks in Finland. He has published extensively in the fields of communication theory, RF and antenna design, signal processing for wireless communications, and is an inventor on over 10 recent 5GIC patents addressing bottleneck problems in 5G systems.