Collaborative Beamforming Aided Fog Radio Access Networks

W. Zhu[1,2], H. D. Tuan[2], E. Dutkiewicz[2], and L. Hanzo[3]

***Abstract*—The success of fog radio access networks (F-RANs) is critically dependent on the potential quality of service (QoS) that they can offer to users in the face of capacity-constrained fronthaul links and limited caches at their remote radio heads (RRHs). In this context, the collaborative beamforming design is very challenging, since it constitutes a large-dimensional nonlinearly constrained optimization problem. The paper develops a new technique for tackling these critical challenges in fog computing. We show that all the associated constraints can be efficiently dealt with maximizing the geometric mean (GM) of the user throughputs (GM-throughput) subject to the affordable total transmit power constraints. To elaborate, the GM-throughput maximization judiciously exploits the fronthaul links and the RRHs' caches by relying on our novel algorithm, which evaluates low-complexity closed-form expressions in each of its iterations. The problem of F-RAN energy-efficiency is also addressed while maintaining the target throughput. Numerical examples are provided for quantifying the efficiency of the proposed algorithms.**

***Index Terms*—Frog radio access network (F-RAN), multi-input single output (MISO), collaborative beamforming, geometric mean (GM) maximization**

I. INTRODUCTION

Fog radio access networks (F-RANs) [1]–[3] have been conceived for bringing distributed computation, communication, control, and storage closer to end users for supporting a range of emerging Internet-of-Things (IoT) applications, such as augmented reality/virtual reality (AR/VR), device-to-device (D2D) communications, smart living and smart cities, etc. (see e.g. [4]–[7] and references therein). For providing low-latency services for all users [8], F-RANs rely on caching memory at their remote radio heads (RRH) [9]–[11], which are uniformly distributed over the service area and thus are close to all users. Usually, each user's request is served by a number of the nearest RRHs and due to the limited capacity of the RRHs' caches, the baseband unit (BBU) still has to fetch some segments of the users' requests that are not in the RRHs' caches through the costly and limited-capacity fronthaul links [12], [13]. Thus, collaborative beamforming (CBF) at RRHs plays a pivotal role in maintaining the quality-of-service (QoS) provided by F-RANs. Compared to the traditional CBF of sensor or generic wireless networks [14]–[20], under which multiple single-antenna nodes form a virtual beamformer for fetching and assembling the contents, the RRHs in a F-RAN equipped with multiple-antenna arrays typically employ multi-user beamformers for transmitting the contents [21]–[27]. Then the concerted F-RAN is only deemed successful, if the throughput of the users is above the minimum threshold subject to the constraints on both the fronthaul capacity and the transmit power. Since the throughput is a complex nonlinear function of all the beamformer weights and there are numerous beamformers, determining the CBF weights constitutes a large-dimensional nonlinearly constrained optimization problem [21], [22], [25]–[28]. The family of advanced optimization algorithms such as those proposed in [26], [28], which iteratively solve convex problems were only capable of handling scenarios of up to five four-antenna based RRHs serving ten users.

At the time of writing, most contributions on CBF are based on the conventional proper Gaussian signaling (PGS), which is based on linearly beamforming proper Gaussian source signals. However, it has been shown e.g. in [29]–[35] that PGS is outperformed by improper Gaussian signaling (IGS), which is generated by widely linearly beamforming proper Gaussian source signals [36]. However, in contrast to PGS, which is based on single beamformers, IGS is based on pairs of correlated beamformers. Hence, the design of the CBF weights under IGS involves twice the number of decision variables compared to its PGS counterpart. As such, convex-solver based algorithms employed for CBF weight optimization under IGS are only capable of handling limited-dimensional scenarios of up to five users [35].

Against the above background, this paper offers the following contributions:

- We show that the problem of providing similar throughputs for all users subject to capacity-constrained fronthaul links and limited RRH caches can be formulated as the problem of maximizing the geometric mean of their throughputs (GM-throughput) subject to the total affordable transmit power constraints. We will demonstrate that the GM-throughput maximization results in similar throughputs for the users and at the same time it judiciously exploits the constrained capacity of the fronthaul links and finite RRH caches;
- We develop a sophisticated but low-complexity optimization algorithm for this GM-throughput maximization problem, which has the explicit benefit of relying on solving closed-form expressions in each iteration. Hence it is eminently suitable for large-scale F-RANs;

[1]School of Communication and Information Engineering, Shanghai University, Shanghai 200444, China (email: wenbozhu@shu.edu.cn); [2]School of Electrical and Data Engineering, University of Technology Sydney, Broadway, NSW 2007, Australia (email: wenbo.zhu@student.uts.edu.au, tuan.hoang@uts.edu.au, eryk.dutkiewicz@uts.edu.au); [3]School of Electronics and Computer Science, University of Southampton, Southampton, SO17 1BJ, U.K (email: lh@ecs.soton.ac.uk)

H. D. Tuan would like to acknowledge the financial support of the Australian Research Council's Discovery Projects under Grant DP190102501.

L. Hanzo would like to acknowledge the financial support of the Engineering and Physical Sciences Research Council projects EP/P034284/1 and EP/P003990/1 (COALESCE) as well as of the European Research Council's Advanced Fellow Grant QuantCom (Grant No. 789028)

TABLE I
BOLDLY AND EXPLICITLY CONTRASTING OUR CONTRIBUTIONS TO THE LITERATURE

	this paper	[21], [22] [24], [25]	[26]	[28]	[37]
FRAN	✓	✓	✓	✓	
Content-service	✓	✓	✓	✓	
IGS	✓			✓	✓
QoS	✓		✓	✓	✓
Energy-efficiency	✓			✓	
Scalable Complexity	✓				✓
Convex-solver based			✓	✓	
Semi-definite relaxation		✓			

- We also employ IGS for improving the GM-throughput of F-RANs. The problem of maximizing the GM-throughput under widely linear beamforming is much more computationally challenging than that under linear beamforming, not only because the number of decision variables is doubled but the throughput becomes a log-determinant function. Again, another algorithm is developed for its computation, which still only relies on closed-form expressions in each iteration.

Our novel contributions are boldly and explicitly contrasted to the state-of-the-art in Table I at a glance.

The paper is organized as follows. Section II is devoted to the linear beamforming aided design of F-RANs capable of supporting a certain target GM-throughput. Then a similar problem is considered in Section III for designing widely linear beamformers. Our simulations characterizing FRANs having with 20 multiple-antenna RRHs and supporting 30 users are discussed in Section IV. Section V concludes the paper.

Notation. Our notations are standard, where only optimization variables are boldfaced to emphasize their appearance in nonlinear functions; $|\mathcal{A}|$ is the cardinality of the set $\mathcal{A}$; The dot product $\langle A, B\rangle$ of two matrices A and B is the trace$(A^H B)$; Accordingly, the Frobenius norm of the matrix A is defined as $||A|| = \sqrt{\langle A, A\rangle}$; We use $[A]^2$ to refer to AA^H, so the quadratic form $x^H Ax$ associated with a Hermitian symmetric matrix A is often expressed by $\langle A, [x]^2\rangle$; $\mathcal{CN}(0,1)$ is the set of circular Gaussian random variables with zero mean and unit covariance, while $\mathcal{C}(0,1)$ is the set of non-circular Gaussian random variables with zero means and unit covariance; Note that $\mathbb{E}(x^2) = 0$ for $x \in \mathcal{CN}(0,1)$ but $\mathbb{E}(x^2) \neq 0$ for $x \in \mathcal{C}(0,1)$.

The following inequality [38] that holds for all matrices $\mathbf{V}$ and $\bar{V}$ of size $n\times$ and positive definite matrices $\mathbf{Y}$ and $\bar{Y}$ of size $n \times n$ is frequently used:

$$\begin{aligned}\ln|I_n + [\mathbf{V}]^2\mathbf{Y}^{-1}| \geq \ln|I_n + [\bar{V}]^2\bar{Y}^{-1}| \\ - \langle\bar{Y}^{-1}, [\bar{V}]^2\rangle + 2\Re\{\langle\bar{Y}^{-1}\bar{V}, \mathbf{V}\rangle\} \\ - \langle\bar{Y}^{-1} - \left([\bar{V}]^2 + \bar{Y}\right)^{-1}, [\mathbf{V}]^2 + \mathbf{Y}\rangle.\end{aligned} \tag{1}$$

II. PROPER GAUSSIAN SIGNALING FOR FRAN

Fig. 1 illustrates a F-RAN of a baseband unit (BBU) linked with N_R RRHs (indexed by $i \in \mathcal{N}_R \triangleq \{1, \ldots, N_R\}$) through the fronthaul links to provide contents for K users (UEs) (indexed by $k \in \mathcal{K} \triangleq \{1, \ldots, K\}$). The RRHs are equipped with N_t-element antenna arrays, while the UEs are equipped with single antennas. Each UE k is assumed to request the content f_k from a library of F files having the total capacity of F_c. Under the uncoded strategy, each file f is split into L subfiles (f, ℓ), $\ell \in \mathcal{L} \triangleq \{1, \ldots, L\}$. The fractional caching capacity at each RRH is fixed at $\mu = N_c/F_c$, so each RRH can store a fraction μ of each file during the pre-fetching phase [39]. For binary-valued $c^i_{f,\ell}$ satisfying $c^i_{f,\ell} = 1$ if and only if (f, ℓ) is cached by RRH i, we have $\sum_{f=1}^{F}\sum_{\ell=1}^{L} c^i_{f,\ell}/L \leq \mu F$, $i \in \mathcal{N}_R$.

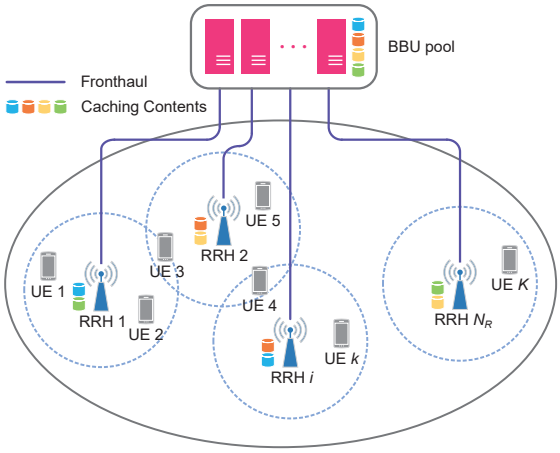

Fig. 1. F-RAN Architecture

The set of requested files is defined as

$$\mathcal{F}_{req} \triangleq \{f_k \in \mathcal{F} : \ k = 1, \ldots, K\}.$$

In what follows, for convenience of presentation, we use the notation $c^i_{k,\ell}$ to refer to $c^i_{f_k,\ell}$.

Given (f_k, ℓ), we define

$$d^i_{k,\ell} = \begin{cases} 1 & \text{if subfile } (f_k, \ell) \text{ is transferred to RRH } i, \\ 0 & \text{otherwise.} \end{cases} \tag{2}$$

To satisfy the rate requirement, the subfile (f_k, ℓ) requested by UE k is fetched through the fronthaul links to the number N_F of RRHs having the highest channel gains from them to this UE among all the RRHs that do not have (f_k, ℓ) in their cache. For $i = 1, \ldots, N_R$, the set of subfiles that are either in RRH i's cache or are received by RRH i from the BBU is defined as

$$\mathcal{N}_i \triangleq \{(k, \ell) : k \in \mathcal{K}, \text{ and } c^i_{k,\ell} = 1 \text{ or } d^i_{k,\ell} = 1\}. \tag{3}$$

We set $\mathbf{v}^i_{k,\ell} = \emptyset$ if $(k, \ell) \notin \mathcal{N}_i$ to define

$$\mathbf{v}_{k,\ell} \triangleq \begin{bmatrix} \mathbf{v}^1_{k,\ell} \\ \mathbf{v}^2_{k,\ell} \\ \cdots \\ \mathbf{v}^{N_R}_{k,\ell} \end{bmatrix}, (k, \ell) \in \mathcal{K} \times \mathcal{L}, \tag{4}$$

which concatenates all beamforming (BF) vectors for the information source $s_{k,\ell} \in \mathcal{C}(0,1)$ encoding (f_k, ℓ), and then we define $\mathbf{v} \triangleq \{\mathbf{v}_{k,\ell} : (k, \ell) \in \mathcal{K} \times \mathcal{L}\}$, which represents

the set of all beamformers. With $n_{k,\ell}$ N_t-dimensional vectors $\mathbf{v}_{k,\ell}^i$ in (4), the dimension of $\mathbf{v}_{k,\ell}$ in (4) is given by

$$N_{k,\ell} = N_t \cdot n_{k,\ell}. \tag{5}$$

We define the matrix $\Lambda_{k,\ell}^i \in \mathbb{R}^{N_t \times N_{k,\ell}}$ to represent $\mathbf{v}_{k,\ell}^i$ associated with $(k,\ell) \in \mathcal{N}_i$ as a component of $\mathbf{v}_{k,\ell}$ as[1]

$$\Lambda_{k,\ell}^i \mathbf{v}_{k,\ell} = \mathbf{v}_{k,\ell}^i. \tag{6}$$

Let us also define $\Xi^i \triangleq \mathcal{N}_i \setminus \{(k,\ell) \ : \ c_{k,\ell}^i = 1\}$. The BBU must fetch the following beamformed subfiles to RRH i:

$$\xi_i = \sum_{(k,\ell)\in\Xi^i} \mathbf{v}_{k,\ell}^i s_{k,\ell}, \tag{7}$$

where $\mathbf{v}_{k,\ell}^i \in \mathbb{C}^{N_t}$ is the baseband beamformer for $s_{k,\ell} \in \mathcal{CN}(0,1)$.

Let $h_{k,i} \in \mathbb{C}^{1\times N_t}$ represent the channel spanning from RRH i to UE k, which is assumed to be known [21]–[27], and $z_k \in \mathbb{C}$ be the noise having the covariance of σ, which incorporates both the background noise and other uncertainties, such as the channel estimation error [40].The robust design relying on imperfect channel state information, which explicitly incorporates channel estimation error into the optimization formulation is beyond the scope of this paper. The signal received at UE k is formulated as

$$\begin{aligned}
y_k =& \sum_{i=1}^{N_R} h_{k,i} \left(\sum_{(k',\ell)\in\mathcal{N}_i} \mathbf{v}_{k',\ell}^i s_{k',\ell} \right) + z_k \\
=& \sum_{\ell=1}^{L} \left[\sum_{i:(k,\ell)\in\mathcal{N}_i} h_{k,i}\mathbf{v}_{k,\ell}^i \right] s_{k,\ell} \\
&+ \sum_{k'\in\mathcal{K}\setminus\{k\}} \sum_{\ell=1}^{L} \sum_{i:(k',\ell)\in\mathcal{N}_i} h_{k,i}\mathbf{v}_{k',\ell}^i s_{k',\ell} + z_k \qquad (8) \\
=& \sum_{\ell=1}^{L} \left[\sum_{i:(k,\ell)\in\mathcal{N}_i} h_{k,i}\Lambda_{k,\ell}^i \mathbf{v}_{k,\ell} \right] s_{k,\ell} \\
&+ \sum_{k'\in\mathcal{K}\setminus\{k\}} \sum_{\ell=1}^{L} \sum_{i:(k',\ell)\in\mathcal{N}_i} h_{k,i}\Lambda_{k',\ell}^i \mathbf{v}_{k',\ell} s_{k',\ell} + z_k \\
=& \sum_{\ell=1}^{L} \mathcal{A}_{k,k,\ell}\mathbf{v}_{k,\ell}s_{k,\ell} + \sum_{k'\in\mathcal{K}\setminus\{k\}} \sum_{\ell=1}^{L} \mathcal{A}_{k,k',\ell}\mathbf{v}_{k',\ell}s_{k',\ell} + z_k, \qquad (9)
\end{aligned}$$

for

$$\mathcal{A}_{k,k',\ell} \triangleq \sum_{i:(k',\ell)\in\mathcal{N}_i} h_{k,i}\Lambda_{k',\ell}^i \in \mathbb{C}^{1\times N_{k,\ell}}. \tag{10}$$

Based on (9), UE k successively detects $s_{k,1}, s_{k,2}, \ldots, s_{k,L}$ as follows:

- It detects $s_{k,1}$ by treating $\mathcal{A}_{k,k,1}\mathbf{v}_{k,1}s_{k,1}$ in (9) as the signal of interest (to be detected) and the remainder

$$\sum_{\ell=2}^{L} \mathcal{A}_{k,k,\ell}\mathbf{v}_{k,\ell}s_{k,\ell} + \sum_{k'\in\mathcal{K}\setminus\{k\}} \sum_{\ell=1}^{L} \mathcal{A}_{k,k',\ell}\mathbf{v}_{k',\ell}s_{k',\ell} + z_k$$

as the interference plus noise. As such the throughput of $s_{k,1}$ is given by

$$r_{k,1}(\mathbf{v}) = \ln\left[1 + |\mathcal{A}_{k,k,1}\mathbf{v}_{k,1}|^2/\psi_{k,1}(\mathbf{v})\right], \tag{11}$$

where

$$\begin{aligned}
\psi_{k,1}(\mathbf{v}) \triangleq& \sum_{\ell'=2}^{L} |\mathcal{A}_{k,k,\ell'}\mathbf{v}_{k,\ell'}|^2 \\
&+ \sum_{k'\in\mathcal{K}\setminus\{k\}} \sum_{\ell=1}^{L} |\mathcal{A}_{k,k',\ell}\mathbf{v}_{k',\ell}|^2 + \sigma. \qquad (12)
\end{aligned}$$

- The UE then detects $s_{k,\ell}$ for $1 < \ell \le K$ by successive cancelation (SC): it subtracts the detected signal

$$\sum_{\ell'=1}^{\ell-1} \mathcal{A}_{k,k,\ell'}\mathbf{v}_{k,\ell'}s_{k,\ell'} \tag{13}$$

from the RHS of (9), yielding

$$\sum_{\ell'=\ell}^{L} \mathcal{A}_{k,k,\ell'}\mathbf{v}_{k,\ell'}s_{k,\ell'} + \sum_{k'\in\mathcal{K}\setminus\{k\}} \sum_{\ell=1}^{L} \mathcal{A}_{k,k',\ell}\mathbf{v}_{k'\ell}s_{k',\ell} + z_k, \tag{14}$$

for detecting $s_{k,\ell}$ by treating $\mathcal{A}_{k,k,\ell}\mathbf{v}_{k,\ell}s_{k,\ell}$ in (14) as the signal of interest, and the remainder

$$\begin{aligned}
&\sum_{\ell'=\ell+1}^{L} \mathcal{A}_{k,\ell',k}\mathbf{v}_{k,\ell'}s_{k,\ell'} + \sum_{k'\in\mathcal{K}\setminus\{k\}} \sum_{\ell=1}^{L} \mathcal{A}_{k,k',\ell}\mathbf{v}_{k',\ell}s_{k',\ell} \\
&\qquad + z_k, \qquad (15)
\end{aligned}$$

as the interference plus noise. As such, the throughput of $s_{k,\ell}$ is[2]

$$r_{k,\ell}(\mathbf{v},\boldsymbol{x}) = \ln\left[1 + |\mathcal{A}_{k,k,\ell}\mathbf{v}_{k,\ell}|^2/\psi_{k,\ell}(\mathbf{v})\right], \tag{16}$$

where we have

$$\begin{aligned}
\psi_{k,\ell}(\mathbf{v}) \triangleq& \sum_{\ell'=\ell+1}^{L} |\mathcal{A}_{k,k,\ell'}\mathbf{v}_{k,\ell'}|^2 \\
&+ \sum_{k'\in\mathcal{K}\setminus\{k\}} \sum_{\ell=1}^{L} |\mathcal{A}_{k,k',\ell}\mathbf{v}_{k',\ell}|^2 + \sigma. \qquad (17)
\end{aligned}$$

A. GM-throughput maximization

We now consider the following problem of maximizing the GM of the subfile throughput (GM-throughput):

$$\max_{\mathbf{v}} \varphi(\mathbf{v}) \triangleq \left(\prod_{k=1}^{K} \prod_{\ell=1}^{L} r_{k,\ell}(\mathbf{v}) \right)^{1/(KL)} \tag{18a}$$

$$||\mathbf{v}||^2 \le P, \tag{18b}$$

[1]Assume that $\Lambda_{k,\ell}^i = \mathsf{Row}[\Lambda_{k,\ell}^{i_j}]_{j=1,\ldots,n_{k,\ell}}$ with $\Lambda_{k,\ell}^{i_j} \in \mathbb{R}^{N_t\times N_t}$. Assume furthermore that $\mathbf{v}_{k,\ell}^i$ is from the $[(r_i-1)N_t+1]$-st entry to the (r_iN_t-1)-st entry in $\mathbf{v}_{k,\ell}$. Then we have $\Lambda_{k,\ell}^{i_{r_i}} = I_{N_t}$ and $\Lambda_{k,\ell}^{i_j} = 0$ for $j \ne r_i$.

[2]$\mathbf{v}_{k,\ell} \in \mathbb{C}^{N_{k,\ell}\times 1}$ by (5) and $\mathcal{A}_{k,k,\ell} \in 1\times N_{k,\ell}$ by (10) so $\mathcal{A}_{k,k,\ell}\mathbf{v}_{k,\ell} \in \mathbb{C}$

where (18b) is the sum transmit power constraint given the total power budget P.

Let us mention that this problem formulation ignores the so-called fronthaul link constraint on transferring ξ_i defined by (7) concerning the BBU to RRH link i:

$$\sum_{(k,\ell)\in\Xi^i} r_{k,\ell}(\mathbf{v}) \leq \log_2(e)C, i = 1,\ldots,N_R, \tag{19}$$

where C is the fronthaul capacity. Instead, after computing (18) we will calculate

$$C_F = \frac{1}{\log_2(e)} \max_{i=1,\ldots,N_R} [\sum_{(k,\ell)\in\Xi^i} r_{k,\ell}(\mathbf{v})] \tag{20}$$

to show that it is far below the practical values of C in (19), which range from 5 bps/Hz to 10 bps/Hz.

Furthermore, the problem (18) does not impose throughput constraints, such as

$$r_{k,\ell}(\mathbf{v}) \geq \bar{r}, (k,\ell) \in \mathcal{K}\times\mathcal{L}, \tag{21}$$

for the successful delivery of all subfiles (k,ℓ), since optimizing the GM-throughput in (18) will be shown to lead to fair user-throughput distributions, automatically satisfying (21).

Finally, the problem (18) does not include the transmit power constraints at the RRHs, as routinely done in cell-free multiple input multiple output networks [41], [42], because again, optimizing the GM-throughput in (18) will be shown to lead to fairly similar transmit powers at the RHHs.

In other words, the problem (18) provides a new approach to F-RAN CBF designs which achieve the targets (19) and (21) along with similar transmit powers at the RRHs.

Following [37], to solve problem (18), we use its following equivalent max-min formulation:

$$\max_{\mathbf{v}} \min_{\gamma_{k,\ell}>0, \prod_{k=1}^K\prod_{\ell=1}^L \gamma_{k,\ell}=1} \frac{\sum_{k=1}^K\sum_{\ell=1}^L \gamma_{k,\ell} r_{k,\ell}(\mathbf{v})}{KL}$$
$$\text{s.t.} \quad (18b). \tag{22}$$

We will now use alternating optimization between $\mathbf{v}$ and $\gamma \triangleq \{\gamma_{k,\ell}, (k,\ell)\in\mathcal{K}\times\mathcal{L}\}$.

Let $v^{(\kappa)}$ be a feasible point for (18) that is found from the $(\kappa-1)$-st alternating iteration in $\mathbf{v}$, and

$$r_{\max}^{(\kappa)} \triangleq \max_{(k,\ell)\in\mathcal{K}\times\mathcal{L}} r_{k,\ell}(v^{(\kappa)}). \tag{23}$$

Then, the optimal solution of the alternating optimization in γ, which is formulated as

$$\min_{\gamma_{k,\ell}>0, \prod_{k=1}^K\prod_{\ell=1}^L \gamma_{k,\ell}=1} \frac{\sum_{k=1}^K\sum_{\ell=1}^L \gamma_{k,\ell} r_{k,\ell}(v^{(\kappa)})}{KL} \tag{24}$$

is given by $\gamma_{k,\ell}^{opt,(\kappa)} = \frac{r_{\max}^{(\kappa)}}{\varphi(v^{(\kappa)}) r_{k,l}(v^{(\kappa)})}, (k,\ell)\in\mathcal{K}\times\mathcal{L}$. Thus the alternating optimization in $\mathbf{v}$ at the κ-th iteration round is expressed as

$$\max_{\mathbf{v}} \frac{r_{\max}^{(\kappa)}}{KL} \sum_{k=1}^K\sum_{\ell=1}^L \frac{r_{k,\ell}(\mathbf{v})}{r_{k,\ell}(v^{(\kappa)})} \quad \text{s.t.} \quad (18b), \tag{25}$$

which is equivalent to

$$\max_{\mathbf{v}} \ \varphi^{(\kappa)}(\mathbf{v}) \triangleq \sum_{k=1}^K\sum_{\ell=1}^L \gamma_{k,\ell}^{(\kappa)} r_{k,\ell}(\mathbf{v}) \quad \text{s.t.} \quad (18b), \tag{26}$$

for

$$\gamma_{k,\ell}^{(\kappa)} \triangleq \frac{r_{\max}^{(\kappa)}}{r_{k,\ell}(v^{(\kappa)})}, (k,\ell)\in\mathcal{K}\times\mathcal{L}. \tag{27}$$

Furthermore, applying the inequality (1) yields the following concave lower-bounding approximation for $r_{k,\ell}(\mathbf{v})$:

$$\begin{aligned} r_{k,\ell}(\mathbf{v}) \geq & r_{k,\ell}^{(\kappa)}(\mathbf{v}) \\ \triangleq & r_{k,\ell}(v^{(\kappa)}) - \frac{|\mathcal{A}_{k,k,\ell} v_{k,\ell}^{(\kappa)}|^2}{\psi_{k,\ell}(v^{(\kappa)})} \\ & + 2\frac{\Re\{\langle \mathcal{A}_{k,k,\ell} v_{k,\ell}^{(\kappa)}, \mathcal{A}_{k,k,\ell}\mathbf{v}_{k,\ell}\rangle\}}{\psi_{k,\ell}(v^{(\kappa)})} \\ & - \frac{|\mathcal{A}_{k,k,\ell} v_{k,\ell}^{(\kappa)}|^2 \left(|\mathcal{A}_{k,k,\ell}\mathbf{v}_{k,\ell}|^2 + \psi_{k,\ell}(\mathbf{v})\right)}{(|\mathcal{A}_{k,k,\ell} v^{(\kappa)}|^2 + \psi_{k,\ell}(v^{(\kappa)}))\psi_{k,\ell}(v^{(\kappa)})} \end{aligned} \tag{28}$$

$$= \chi_{k,\ell}^{(\kappa)} + 2\Re\{\langle g_{k,\ell}^{(\kappa)}, \mathbf{v}_{k,\ell}\rangle\} - \sum_{\ell'=\ell}^L \langle \mathcal{C}_{k,\ell,k,\ell'}^{(\kappa)}, [\mathbf{v}_{k,\ell'}]^2\rangle$$
$$- \sum_{k'\in\mathcal{K}\setminus\{k\}} \sum_{\ell'=1}^L \langle \mathcal{C}_{k,\ell,k',\ell'}^{(\kappa)}, [\mathbf{v}_{k',\ell'}]^2\rangle, \tag{29}$$

for $\nu_{k,\ell}^{i,(\kappa)} = \frac{1}{\psi_{k,\ell}(v^{(\kappa)})}$, $\zeta_{k,\ell}^{(\kappa)} \triangleq \frac{|\mathcal{A}_{k,k,\ell} v_{k,\ell}^{(\kappa)}|^2}{(|\mathcal{A}_{k,k,\ell} v^{(\kappa)}|^2 + \psi_{k,\ell}(v^{(\kappa)}))} \nu_{k,\ell}^{i,(\kappa)}$, $\chi_{k,\ell}^{(\kappa)} \triangleq r_{k,\ell}(v^{(\kappa)}) - |\mathcal{A}_{k,k,\ell} v^{(\kappa)}|^2 \nu_{k,\ell}^{i,(\kappa)} - \zeta_{k,\ell}^{(\kappa)}\sigma$, and $g_{k,\ell}^{(\kappa)} \triangleq \nu_{k,\ell}^{i,(\kappa)} [\mathcal{A}_{k,k,\ell}^H]^2 v_{k,\ell}^{(\kappa)}$, $\mathcal{C}_{k,\ell,k',\ell'}^{(\kappa)} \triangleq \zeta_{k,\ell}^{(\kappa)} [\mathcal{A}_{k,k',\ell'}^H]^2$.

Then,

$$\begin{aligned} \varphi^{(\kappa)}(\mathbf{v}) \geq & \tilde{\varphi}^{(\kappa)}(\mathbf{v}) \\ \triangleq & \sum_{k=1}^K\sum_{\ell=1}^L \gamma_{k,\ell}^{(\kappa)} r_{k,\ell}^{(\kappa)}(\mathbf{v}) \\ = & \chi^{(\kappa)} + 2\sum_{k=1}^K\sum_{\ell=1}^L \gamma_{k,\ell}^{(\kappa)} \Re\{\langle g_{k\ell}^{(\kappa)}, \mathbf{v}_{k,\ell}\rangle\} \\ & - \sum_{k=1}^K\sum_{\ell=1}^L \gamma_{k,\ell}^{(\kappa)} \left(\sum_{\ell'=\ell}^L \langle \mathcal{C}_{k,\ell,k,\ell'}^{(\kappa)}, [\mathbf{v}_{k,\ell'}]^2\rangle \right. \\ & \left. + \sum_{k'\in\mathcal{K}\setminus\{k\}} \sum_{\ell'=1}^L \langle \mathcal{C}_{k,\ell,k',\ell'}^{(\kappa)}, [\mathbf{v}_{k',\ell'}]^2\rangle \right) \\ = & \chi^{(\kappa)} + 2\sum_{k=1}^K\sum_{\ell=1}^L \gamma_{k,\ell}^{(\kappa)} \Re\{\langle g_{k\ell}^{(\kappa)}, \mathbf{v}_{k,\ell}\rangle\} \\ & - \sum_{k=1}^K\sum_{\ell=1}^L \langle \mathcal{C}_{k,\ell}^{(\kappa)}, [\mathbf{v}_{k,\ell}]^2\rangle, \end{aligned} \tag{30}$$

$$\tag{31}$$

for $\chi^{(\kappa)} \triangleq \sum_{k=1}^K\sum_{\ell=1}^L \gamma_{k,\ell}^{(\kappa)} \chi_{k,\ell}^{(\kappa)}$, and $\mathcal{C}_{k,\ell}^{(\kappa)} \triangleq \sum_{\ell'=1}^{\ell} \gamma_{k,\ell'}^{(\kappa)} \mathcal{C}_{k,\ell',k,\ell}^{(\kappa)} + \sum_{k'\in\mathcal{K}\setminus\{k\}} \sum_{\ell'=1}^L \gamma_{k',\ell'}^{(\kappa)} \mathcal{C}_{k',\ell',k,\ell}^{(\kappa)}$,

because

$$\sum_{k=1}^{K}\sum_{\ell=1}^{L}\gamma_{k,\ell}^{(\kappa)}\sum_{\ell'=\ell}^{L}\langle\mathcal{C}_{k,\ell,k,\ell'}^{(\kappa)},[\mathbf{v}_{k,\ell'}]^2\rangle = \sum_{k=1}^{K}\Bigg[\gamma_{k,1}^{(\kappa)}\sum_{\ell'=1}^{L}\langle\mathcal{C}_{k,1,k,\ell'}^{(\kappa)},[\mathbf{v}_{k,\ell'}]^2\rangle + \gamma_{k,2}^{(\kappa)}\sum_{\ell'=2}^{L}\langle\mathcal{C}_{k,2,k,\ell'}^{(\kappa)},[\mathbf{v}_{k,\ell'}]^2\rangle + \cdots + \gamma_{k,L}^{(\kappa)}\langle\mathcal{C}_{k,L,k,L}^{(\kappa)},[\mathbf{v}_{k,L}]^2\rangle\Bigg] = \sum_{k=1}^{K}\sum_{\ell=1}^{L}\langle\sum_{\ell'=1}^{\ell}\gamma_{k,\ell'}^{(\kappa)}\mathcal{C}_{k,\ell',k,\ell}^{(\kappa)},[\mathbf{v}_{k,\ell}]^2\rangle,$$

while

$$\sum_{k=1}^{K}\sum_{\ell=1}^{L}\gamma_{k,\ell}^{(\kappa)}\sum_{k'\in\mathcal{K}\setminus\{k\}}\sum_{\ell'=1}^{L}\langle\mathcal{C}_{k,\ell,k',\ell'}^{(\kappa)},[\mathbf{v}_{k',\ell'}]^2\rangle = \sum_{k'=1}^{K}\sum_{\ell'=1}^{L}\gamma_{k',\ell'}^{(\kappa)}\sum_{k\in\mathcal{K}\setminus\{k'\}}\sum_{\ell=1}^{L}\langle\mathcal{C}_{k',\ell',k,\ell}^{(\kappa)},[\mathbf{v}_{k,\ell}]^2\rangle = \sum_{k=1}^{K}\sum_{\ell=1}^{L}\langle\sum_{k'\in\mathcal{K}\setminus\{k\}}\sum_{\ell'=1}^{L}\gamma_{k',\ell'}^{(\kappa)}\mathcal{C}_{k',\ell',k,\ell}^{(\kappa)},[\mathbf{v}_{k,\ell}]^2\rangle.$$

We solve the following convex problem at the κ-th iteration to generate the next feasible point $v^{(\kappa+1)}$ for (18):

$$\max_{\mathbf{v},\boldsymbol{x}}\tilde{\varphi}^{(\kappa)}(\mathbf{v}) \quad \text{s.t.} \quad (18b), \tag{32}$$

which admits the closed-form solution[3]

$$v_{k,\ell}^{(\kappa+1)} = \begin{cases} \gamma_{k,\ell}^{(\kappa)}(\mathcal{C}_{k,\ell}^{(\kappa)})^{-1}g_{k\ell}^{(\kappa)} \\ \text{if } \sum_{k=1}^{K}\sum_{\ell=1}^{L}||\gamma_{k,\ell}^{(\kappa)}(\mathcal{C}_{k,\ell}^{(\kappa)})^{-1}g_{k\ell}^{(\kappa)}||^2 \le P, \\ \\ \gamma_{k,\ell}^{(\kappa)}(\mathcal{C}_{k,\ell}^{(\kappa)}+\lambda I_{N_{k,\ell}})^{-1}g_{k\ell}^{(\kappa)} \\ \text{otherwise}, \end{cases} \tag{33}$$

where λ is found from bisection based search, so that we have

$$\sum_{k=1}^{K}\sum_{\ell=1}^{L}||\gamma_{k,\ell}^{(\kappa)}(\mathcal{C}_{k,\ell}^{(\kappa)}+\lambda I_{N_{k,\ell}})^{-1}g_{k\ell}^{(\kappa)}||^2 = P. \tag{34}$$

Note that $\varphi^{(\kappa)}(v^{(\kappa+1)}) \ge \tilde{\varphi}^{(\kappa)}(v^{(\kappa+1)})$ by (30), while $\tilde{\varphi}^{(\kappa)}(v^{(\kappa+1)}) > \tilde{\varphi}^{(\kappa)}(v^{(\kappa)}) = \varphi^{(\kappa)}(v^{(\kappa)})$ as far as $\tilde{\varphi}^{(\kappa)}(v^{(\kappa+1)}) \ne \tilde{\varphi}^{(\kappa)}(v^{(\kappa)})$ because $v^{(\kappa+1)}$ and $v^{(\kappa)}$ are the optimal solution and a feasible point for the problem (32). We thus arrive at

$$\varphi^{(\kappa)}(v^{(\kappa+1)}) > \varphi^{(\kappa)}(v^{(\kappa)}), \tag{35}$$

i.e. $v^{(\kappa+1)}$ is a better feasible point for the problem (26), and as thus Algorithm 1 provides steep descent iterations for solving problem (18), namely on GM-throughput maximization.

The computational complexity of (33) is on the order of $\mathcal{O}(N_{k,\ell}\log N_{k,\ell})$ [43], so the computational complexity of each iteration of Algorithm 1 is formulated as:

$$\sum_{(k,\ell)\in\mathcal{K}\times\mathcal{L}}\mathcal{O}(N_{k,\ell}\log N_{k,\ell}), \tag{36}$$

[3]The inverse here is actually a pseudo-inverse, if the matrices considered are near to singularity

Algorithm 1 PGS CBF Algorithm

1: Set $\kappa = 0$. Take a feasible point $v^{(0)}$ to satisfy the power constraint (18b).
2: **Repeat until convergence of the objective function in (18).** Iterate $v^{(\kappa+1)}$ by (33). Reset $v^{(\kappa)} \leftarrow v^{(\kappa+1)}$ and update $\gamma_{k,\ell}^{(\kappa)}$ by (27).
3: **Output** $v^{(\kappa)}$ as the optimal solution of (18).

Algorithm 2 Energy-efficient PGS Algorithm

1: Set $\kappa = 0$. Take a feasible point $v^{(0)}$ to satisfy the power constraint (18b).
2: **Repeat until convergence of the objective function in (39).** Iterate $v^{(\kappa+1)}$ by (45). Reset $v^{(\kappa)} \leftarrow v^{(\kappa+1)}$ and update $\gamma_{k,\ell}^{(\kappa)}$ by (27).
3: **Output** $v^{(\kappa)}$ as the optimal solution of (39).

which represents a scalable complexity. This is in contrast to the computational complexity order $\mathcal{O}((\sum_{(k,\ell)\in\mathcal{K}\times\mathcal{L}}N_{k,\ell})^3)$ of the convex-solver based algorithms [28].

B. Energy-efficiency maximization

Let us now define the total power consumption as

$$\pi_c(\mathbf{v}) \triangleq \alpha||\mathbf{v}||^2 + P_{sc} + N_R\left[P_0 + P_{bt}KL\varphi(\mathbf{v})\right], \tag{37}$$

where α is the reciprocal of the power amplifier's drain efficiency at APs, P_{sc} is the power dissipation of the circuit components, P_0 is the power consumption of each backhaul, and P_{bt} is the traffic-dependent power (in Watt per bits/s). Their values are given in Table II in Section V. The first and second terms in (37) represent the transmission power consumption, while the third term reflects the power consumption of the backhauls. Note that in contrast to [44], we insert $KL(\prod_{k=1}^{K}\prod_{\ell=1}^{L}r_{k,\ell}(\mathbf{v}))^{1/KL}$ into (37) instead of $\sum_{k=1}^{K}\sum_{\ell=1}^{L}r_{k,\ell}(\mathbf{v})$. We consider the following problem of energy-efficiency (EE) maximization:

$$\max_{\mathbf{v}}\frac{KL\varphi(\mathbf{v})}{\pi_c(\mathbf{v})} \quad \text{s.t.} \quad (18b). \tag{38}$$

The rationale behind considering (38) is that by maximizing the objective in (38) one achieves both the subfiles' throughput target, because the GM-throughput in its numerator must be maximized and also high EE, because the power consumption in its denominator must be minimized.

Note that we have

$$\frac{KL\varphi(\mathbf{v})}{\pi_c(\mathbf{v})} = \left(\frac{\alpha||\mathbf{v}||^2 + P_{sc} + N_RP_0}{KL\varphi(\mathbf{v})} + N_RP_{bt}\right)^{-1},$$

so the problem (38) is equivalent to the following problem:

$$\max_{\mathbf{v}} f(\mathbf{v}) \triangleq \frac{\varphi(\mathbf{v})}{\pi_e(\mathbf{v})} \quad \text{s.t.} \quad (18b), \tag{39}$$

where we have

$$\pi_e(\mathbf{v}) \triangleq \alpha||\mathbf{v}||^2 + P_{sc} + N_RP_0. \tag{40}$$

In fact,

$$\max\ (38) = \left(\frac{1}{KL\max\ (39)} + N_R P_{bt}\right)^{-1}. \quad (41)$$

Let $v^{(\kappa)}$ be a feasible point for (39) that is found from the $(\kappa-1)$-st iteration, and accordingly $r_{\max}^{(\kappa)}$ and $\gamma_{k,\ell}^{(\kappa)}$ are defined by (23) and (27). Similarly to (26), we develop a steep descent for (39) by examining the following problem:

$$\max_{\mathbf{v}}\ f^{(\kappa)}(\mathbf{v}) \triangleq \sum_{k=1}^{K}\sum_{\ell=1}^{L}\gamma_{k,\ell}^{(\kappa)} r_{k,\ell}(\mathbf{v}) - \theta^{(\kappa)}\pi_e(\mathbf{v}) \quad \text{s.t.} \quad (18b), \quad (42)$$

for

$$\theta^{(\kappa)} \triangleq KL\frac{r_{\max}^{(\kappa)}}{\pi_e(v^{(\kappa)})}. \quad (43)$$

Recalling the function $\tilde{\varphi}^{(\kappa)}(\mathbf{v})$ from (31), we solve the following convex problem to generate $v^{(\kappa+1)}$:

$$\max_{\mathbf{v}}\ \tilde{\varphi}^{(\kappa)}(\mathbf{v}) - \theta^{(\kappa)}\pi_e(\mathbf{v}) \quad \text{s.t.} \quad (18b), \quad (44)$$

which admits the following closed-form solution:

$$v_{k,\ell}^{(\kappa+1)} = \begin{cases} \gamma_{k,\ell}^{(\kappa)}\left(\mathcal{C}_{k,\ell}^{(\kappa)} + \theta^{(\kappa)}\alpha I_{N_{k,\ell}}\right)^{-1} g_{k\ell}^{(\kappa)} \\ \text{if } \sum_{k=1}^{K}\sum_{\ell=1}^{L}||\gamma_{k,\ell}^{(\kappa)}\left(\mathcal{C}_{k,\ell}^{(\kappa)} + \theta^{(\kappa)}\alpha I_{N_{k,\ell}}\right)^{-1} g_{k\ell}^{(\kappa)}||^2 \\ \qquad\qquad \leq P, \\ \gamma_{k,\ell}^{(\kappa)}\left(\mathcal{C}_{k,\ell}^{(\kappa)} + \theta^{(\kappa)}\alpha I_{N_{k,\ell}} + \lambda I_{N_{k,\ell}}\right)^{-1} g_{k\ell}^{(\kappa)} \\ \text{otherwise}, \end{cases} \quad (45)$$

where $\lambda > 0$ is found from bisection search, so that we have

$$\sum_{k=1}^{K}\sum_{\ell=1}^{L}||\gamma_{k,\ell}^{(\kappa)}\left(\mathcal{C}_{k,\ell}^{(\kappa)} + \theta^{(\kappa)}\alpha I_{N_{k,\ell}} + \lambda I_{N_{k,\ell}}\right)^{-1} g_{k\ell}^{(\kappa)}||^2 = P. \quad (46)$$

Similarly to (35), we can show that

$$\tilde{f}^{(\kappa)}(v^{(\kappa+1)}) > \tilde{f}^{(\kappa)}(v^{(\kappa)}), \quad (47)$$

whenever we have $\tilde{f}^{(\kappa)}(v^{(\kappa+1)}) \neq \tilde{f}^{(\kappa)}(v^{(\kappa)})$, i.e. $v^{(\kappa+1)}$ is a better feasible point than $v^{(\kappa)}$ for (42). Thus Algorithm 2 provides steep descent iterations for the energy-efficient maximization in (39).

The computational complexity of each iteration of Algorithm 2 is the same as that of its counterpart of Algorithm 1, which is defined by (36).

III. Widely linear processing

In (4), each $s_{k,\ell}$ for $(k,\ell) \in \mathcal{N}_i$ is linearly beamformed by the weight vector $\mathbf{v}_{k,\ell}^i \in \mathbb{C}^{N_t}$. Now, $s_{k,\ell}$ for $(k,\ell) \in \mathcal{N}_i$ is beamformed by a pair of beamformers $\mathbf{w}_{k,\ell}^{i,1} \in \mathbb{C}^{N_t}$ and $\mathbf{w}_{k,\ell}^{i,1} \in \mathbb{C}^{N_t}$ as

$$\lambda_{k,\ell}^i(s_{k,\ell}) = \mathbf{w}_{k,\ell}^{i,1}s_{k,\ell} + \mathbf{w}_{k,\ell}^{i,2}s_{k,\ell}^*, \quad (48)$$

i.e. $s_{k,\ell}$ is widely linearly beamformed [36]. In contrast to the signal $\mathbf{v}_{k,\ell}^i s_{k,\ell}$ in (4), which is proper Gaussian as $s_{k,\ell}$ itself, because we have $\mathbb{E}(\mathbf{v}_{k,\ell}^i s_{k,\ell}(\mathbf{v}_{k,\ell}^i s_{k,\ell})^T) = \mathbf{v}_{k,\ell}^i\mathbb{E}(s_{k,\ell}^2)(\mathbf{v}_{k,\ell}^i)^T = 0$, the signal defined by (48) is an improper Gaussian signal because $\mathbb{E}(\lambda_{k,\ell}^i(s_{k,\ell})(\lambda_{k,\ell}^i(s_{k,\ell}))^T) \neq 0$. Thus, from now on we will refer to the beamforming scheme using (4) as proper Gaussian signaling (PGS), while to that using (48) as improper Gaussian signaling (IGS).

Let us now introduce the definitions

$$\begin{aligned}\tilde{\lambda}_{k,\ell}^i(s_{k,\ell}) &\triangleq \begin{bmatrix}\Re\{\lambda_{k,\ell}^i(s_{k,\ell})\} \\ \Im\{\lambda_{k,\ell}^i(s_{k,\ell})\}\end{bmatrix} \\ &= \begin{bmatrix}\Re\{\mathbf{w}_{k,\ell}^{i,1}\} + \Re\{\mathbf{w}_{k,\ell}^{i,2}\} & -\Im\{\mathbf{w}_{k,\ell}^{i,1}\} + \Im\{\mathbf{w}_{k,\ell}^{i,2}\} \\ \Im\{\mathbf{w}_{k,\ell}^{i,1}\} + \Im\{\mathbf{w}_{k,\ell}^{i,2}\} & \Re\{\mathbf{w}_{k,\ell}^{i,1}\} - \Re\{\mathbf{w}_{k,\ell}^{i,2}\}\end{bmatrix} \\ &\quad \times \begin{bmatrix}\Re\{s_{k,\ell}\} \\ \Im\{s_{k,\ell}\}\end{bmatrix} \quad (49) \\ &= \mathbf{V}_{k,\ell}^i\tilde{s}_{k,\ell}, \quad (50)\end{aligned}$$

where we have

$$\tilde{s}_{k,\ell} \triangleq \begin{bmatrix}\Re\{s_{k,\ell}\} \\ \Im\{s_{k,\ell}\}\end{bmatrix}, \quad (51)$$

and

$$\begin{aligned}\mathbf{V}_{k,\ell}^i &\triangleq \begin{bmatrix}\mathbf{v}_{k,\ell}^{i,11} & \mathbf{v}_{k,\ell}^{i,12} \\ \mathbf{v}_{k,\ell}^{i,21} & \mathbf{v}_{k,\ell}^{i,22}\end{bmatrix} \quad (52) \\ &= \begin{bmatrix}\Re\{\mathbf{w}_{k,\ell}^{i,1}\} + \Re\{\mathbf{w}_{k,\ell}^{i,2}\} & -\Im\{\mathbf{w}_{k,\ell}^{i,1}\} + \Im\{\mathbf{w}_{k,\ell}^{i,2}\} \\ \Im\{\mathbf{w}_{k,\ell}^{i,1}\} + \Im\{\mathbf{w}_{k,\ell}^{i,2}\} & \Re\{\mathbf{w}_{k,\ell}^{i,1}\} - \Re\{\mathbf{w}_{k,\ell}^{i,2}\}\end{bmatrix} \\ &\in \mathbb{R}^{(2N_t)\times 2}, \quad (53)\end{aligned}$$

and $N_{k,\ell}$ is defined in (5). Note that the transform (53) is indeed legitimate, since its inverse is given by

$$\begin{bmatrix}\Re\{\mathbf{w}_{k,\ell}^{i,1}\} & \Im\{\mathbf{w}^{i,1}\} \\ \Re\{\mathbf{w}_{k,\ell}^{i,2}\} & \Im\{\mathbf{w}_{k,\ell}^{i,2}\}\end{bmatrix} = \frac{1}{2}\begin{bmatrix}\mathbf{v}_{k,\ell}^{i,11} + \mathbf{v}_{k,\ell}^{i,22} & \mathbf{v}_{k,\ell}^{i,21} - \mathbf{v}_{k,\ell}^{i,12} \\ \mathbf{v}_{k,\ell}^{i,11} - \mathbf{v}_{k,\ell}^{i,22} & \mathbf{v}_{k,\ell}^{i,21} + \mathbf{v}_{k,\ell}^{i,12}\end{bmatrix}. \quad (54)$$

Furthermore, we have:

$$||\mathbf{w}_{k,\ell}^{i,1}||^2 + ||\mathbf{w}_{k,\ell}^{i,2}||^2 = \frac{1}{2}\sum_{j_1=1}^{2}\sum_{j_2=1}^{2}||\mathbf{v}_{k,\ell}^{i,j_1j_2}||^2. \quad (55)$$

Again, recalling that $N_{k,\ell}$ is defined in (5), we set $\mathbf{V}_{k,\ell}^i = \emptyset$ if $(k,\ell) \notin \mathcal{N}_i$ to define

$$\mathbf{V}_{k,\ell} \triangleq \begin{bmatrix}\mathbf{V}_{k,\ell}^1 \\ \mathbf{V}_{k,\ell}^2 \\ \cdots \\ \mathbf{V}_{k,\ell}^{N_R}\end{bmatrix} \in \mathbb{R}^{(2N_{k,\ell})\times 2}, (k,\ell) \in \mathcal{K}\times\mathcal{L}, \quad (56)$$

which is the beamforming matrix for the subfile $s_{k,\ell}$, and then $\mathbf{V} \triangleq \{\mathbf{V}_{k,\ell} : (k,\ell) \in \mathcal{K}\times\mathcal{L}\}$, which is the set of all beamforming matrices. We also introduce the matrix $\tilde{\Lambda}_{k,\ell}^i \in \mathbb{R}^{(2N_t)\times(2N_{k,\ell})}$ to present $\mathbf{V}_{k,\ell}^i$ with $(k,\ell) \in \mathcal{N}_i$ as a component of $\mathbf{V}_{k,\ell}$,[4] which satisfies

$$\tilde{\Lambda}_{k,\ell}^i\mathbf{V}_{k,\ell} = \mathbf{V}_{k,\ell}^i. \quad (57)$$

[4] $\tilde{\Lambda}_{k,\ell}^i \triangleq \mathsf{Row}[\tilde{\Lambda}_{k,\ell}^{ij}]_{j=1,\ldots,n_{k,\ell}}$ with each $\tilde{\Lambda}_{k,\ell}^{ij} \in \mathbb{R}^{(2N_t)\times(2N_t)}$. Assume that $\mathbf{V}_{k,\ell}^i$ is the $((r_i-1)2N_t+1)$-th row to $(r_i 2N_t-1)$-th row in $\mathbf{V}_{k,\ell}$, then $\tilde{\Lambda}_{k,\ell}^{ir_i} = I_{2N_t}$ and $\tilde{\Lambda}_{k,\ell}^{ij} = 0$ for $j \neq r_i$.

The signal received at UE k is formulated as

$$y_k = \sum_{i=1}^{N_R} h_{k,i} \sum_{(k',\ell)\in\mathcal{N}_i} (\mathbf{w}_{k',\ell}^{i,1} s_{k',\ell} + \mathbf{w}_{k',\ell}^{i,2} s_{k',\ell}^*) + z_k. \quad (58)$$

Its equivalent real composite form is given by

$$\begin{aligned}
\tilde{y}_k &\triangleq \begin{bmatrix} \Re\{y_k\} \\ \Im\{y_k\} \end{bmatrix} && (59)\\
&= \sum_{i=1}^{N_R} \tilde{\mathcal{H}}_{k,i} \left(\sum_{(k',\ell)\in\mathcal{N}_i} \mathbf{V}_{k',\ell}^i \tilde{s}_{k',\ell} \right) + \tilde{z}_k && (60)\\
&= \sum_{\ell=1}^{L} \left[\sum_{i:(k,\ell)\in\mathcal{N}_i} \tilde{\mathcal{H}}_{k,i} \mathbf{V}_{k,\ell}^i \right] \tilde{s}_{k,\ell} \\
&\quad + \sum_{k'\in\mathcal{K}\setminus\{k\}} \sum_{\ell=1}^{L} \sum_{i:(k',\ell)\in\mathcal{N}_i} \tilde{\mathcal{H}}_{k,i} \mathbf{V}_{k',\ell}^i \tilde{s}_{k',\ell} + \tilde{z}_k && (61)\\
&= \sum_{\ell=1}^{L} \left[\sum_{i:(k,\ell)\in\mathcal{N}_i} \tilde{\mathcal{H}}_{k,i} \tilde{\Lambda}_{k,\ell}^i \mathbf{V}_{k,\ell} \right] \tilde{s}_{k,\ell} \\
&\quad + \sum_{k'\in\mathcal{K}\setminus\{k\}} \sum_{\ell=1}^{L} \sum_{i:(k',\ell)\in\mathcal{N}_i} \tilde{\mathcal{H}}_{k,i} \tilde{\Lambda}_{k',\ell}^i \mathbf{V}_{k',\ell} \tilde{s}_{k',\ell} + \tilde{z}_k \\
&= \sum_{\ell=1}^{L} \tilde{\mathcal{A}}_{k,k,\ell} \mathbf{V}_{k,\ell} \tilde{s}_{k,\ell} + \sum_{k'\in\mathcal{K}\setminus\{k\}} \sum_{\ell=1}^{L} \tilde{\mathcal{A}}_{k,k',\ell} \mathbf{V}_{k',\ell} \tilde{s}_{k',\ell} \\
&\quad + \tilde{z}_k, && (62)
\end{aligned}$$

for

$$\tilde{\mathcal{H}}_{k,i} \triangleq \begin{bmatrix} \Re\{h_{k,i}\} & -\Im\{h_{k,i}\} \\ \Im\{h_{k,i}\} & \Re\{h_{k,i}\} \end{bmatrix}, \tilde{z}_k \triangleq \begin{bmatrix} \Re\{z_k\} \\ \Im\{z_k\} \end{bmatrix}, \quad (63)$$

and

$$\tilde{\mathcal{A}}_{k,k',\ell} \triangleq \sum_{i:(k',\ell)\in\mathcal{N}_i} \tilde{\mathcal{H}}_{k,i} \tilde{\Lambda}_{k',\ell}^i. \quad (64)$$

Similarly to (13)-(17), in detecting $\tilde{s}_{k,1}, \ldots, \tilde{s}_{k,L}$ by successive cancelation (SC), UE k subtracts the detected signal

$$\sum_{\ell'=1}^{\ell-1} \mathcal{A}_{k,k,\ell'} \mathbf{V}_{k,\ell'} s_{k,\ell'} \quad (65)$$

from the RHS of (62) to arrive at

$$\sum_{\ell'=\ell}^{L} \tilde{\mathcal{A}}_{k,k,\ell'} \mathbf{V}_{k,\ell'} \tilde{s}_{k,\ell'} + \sum_{k'\in\mathcal{K}\setminus\{k\}} \sum_{\ell=1}^{L} \tilde{\mathcal{A}}_{k,k',\ell} \mathbf{V}_{k',\ell} \tilde{s}_{k',\ell} + \tilde{z}_k \quad (66)$$

for detecting $s_{k,\ell}$ by treating

$$\tilde{\mathcal{A}}_{k,k,\ell} \mathbf{V}_{k,\ell} \tilde{s}_{k,\ell} \quad (67)$$

in (66) as the signal of interest, and

$$\sum_{\ell'=\ell+1}^{L} \tilde{\mathcal{A}}_{k,k,\ell'} \mathbf{V}_{k,\ell'} \tilde{s}_{k,\ell'} + \sum_{k'\in\mathcal{K}\setminus\{k\}} \sum_{\ell=1}^{L} \tilde{\mathcal{A}}_{k,k',\ell} \mathbf{V}_{k',\ell} \tilde{s}_{k',\ell} + \tilde{z}_k \quad (68)$$

as the interference plus noise. Then the throughput of $s_{f_k,\ell}$ at user k is $\frac{1}{2}\tilde{r}_{k,\ell}(\mathbf{V}, \boldsymbol{x})$ [36] with

$$\tilde{r}_{k,\ell}(\mathbf{V}) \triangleq \ln \left| I_2 + [\mathcal{A}_{k,k,\ell} \mathbf{V}_{k,\ell}]^2 \Psi_{k,\ell}^{-1}(\mathbf{V}) \right|, \quad (69)$$

for

$$\begin{aligned}
\Psi_{k,\ell}(\mathbf{V}) &\triangleq \sum_{\ell'=\ell+1}^{L} [\tilde{\mathcal{A}}_{k,k,\ell'} \mathbf{V}_{k,\ell'}]^2 \\
&\quad + \sum_{k'\in\mathcal{K}\setminus\{k\}} \sum_{\ell=1}^{L} [\tilde{\mathcal{A}}_{k,k',\ell} \mathbf{V}_{k',\ell}]^2 + \sigma I_2. && (70)
\end{aligned}$$

A. GM-throughput maximization

Based on (18), we consider the following problem:[5]

$$\max_{\mathbf{V}} \Phi(\mathbf{V}) \triangleq \left(\prod_{k=1}^{K} \prod_{\ell=1}^{L} \tilde{r}_{k,\ell}(\mathbf{V}) \right)^{1/(KL)} \quad (71a)$$

$$||\mathbf{V}||^2 \le 2P. \quad (71b)$$

Note that the BBU must forward the following beamformed subfiles to RRH i instead of (7):

$$\xi_i = \sum_{(k,\ell)\in\Xi^i} \left(\mathbf{w}_{k,\ell}^{i,1} s_{k,\ell} + \mathbf{w}_{k,\ell}^{i,2} s_{k,\ell}^* \right), \quad (72)$$

or

$$\tilde{\xi}_i \triangleq \begin{bmatrix} \Re\{\xi_i\} \\ \Im\{\xi_i\} \end{bmatrix} = \sum_{(k,\ell)\in\Xi^i} \mathbf{V}_{k,\ell}^i \tilde{s}_{k,\ell}, \quad (73)$$

with $\mathbf{V}_{k,\ell}^i$ and $\tilde{s}_{k,\ell}$ defined in (53) and (51). Then the fronthaul capacity is defined as the following instead of (20):

$$\tilde{C}_F = \frac{1}{2\log_2(e)} \max_{i=1,\ldots,N_R} [\sum_{(k,\ell)\in\Xi^i} \tilde{r}_{k,\ell}(\mathbf{V})]. \quad (74)$$

Let $V^{(\kappa)}$ be a feasible point for (71) that is found from the $(\kappa-1)$-st iteration and

$$\tilde{\gamma}_{k,\ell}^{(\kappa)} \triangleq \max_{(k',\ell')\in\mathcal{K}\times\mathcal{L}} \tilde{r}_{k,\ell}(V^{(\kappa)}) / \tilde{r}_{k,\ell}(V^{(\kappa)}). \quad (75)$$

Similarly to (26) constructed for PGS, at the κ-th iteration we examine the following problem:

$$\max_{\mathbf{V},\boldsymbol{x}} \ \Phi^{(\kappa)}(\mathbf{V}, \boldsymbol{x}) \triangleq \sum_{k=1}^{K} \sum_{\ell=1}^{L} \tilde{\gamma}_{k,\ell}^{(\kappa)} \tilde{r}_{k,\ell}(\mathbf{V}) \quad \text{s.t.} \quad (71b). \quad (76)$$

Applying the inequality (1) yields the following concave lower-bounding approximation for $\tilde{r}_{k,\ell}(\mathbf{V})$:

$$\begin{aligned}
\tilde{r}_{k,\ell}(\mathbf{V}) &\ge \tilde{r}_{k,\ell}^{(\kappa)}(\mathbf{V}) \\
&\triangleq \tilde{r}_{k,\ell}(V^{(\kappa)}, x^{(\kappa)}) - \langle [\tilde{\mathcal{A}}_{k,k,\ell} V_{k,\ell}^{(\kappa)}]^2, \Psi_{k,\ell}^{-1}(V^{(\kappa)}) \rangle \\
&\quad + 2\langle \Psi_{k,\ell}^{-1}(V^{(\kappa)}) \tilde{\mathcal{A}}_{k,k,\ell} V_{k,\ell}^{(\kappa)}, \tilde{\mathcal{A}}_{k,k,\ell} \mathbf{V}_{k,\ell} \rangle \\
&\quad - \langle \Psi_{k,\ell}^{-1}(V^{(\kappa)}) - \left([\tilde{\mathcal{A}}_{k,k,\ell} V_{k,\ell}^{(\kappa)}]^2 + \Psi_{k,\ell}(V^{(\kappa)}) \right)^{-1}, \\
&\quad [\tilde{\mathcal{A}}_{k,k,\ell} \mathbf{V}_{k,\ell}]^2 + \Psi_{k,\ell}(\mathbf{V}) \rangle && (77)\\
&= \tilde{\chi}_{k,\ell}^{(\kappa)} + 2\langle \tilde{\mathcal{G}}_{k,\ell}^{(\kappa)}, \mathbf{V}_{k,\ell} \rangle - \sum_{\ell'=\ell}^{L} \langle \tilde{\mathcal{C}}_{k,\ell,k,\ell'}^{(\kappa)}, [\mathbf{V}_{k,\ell'}]^2 \rangle
\end{aligned}$$

[5] the optimal value must be divided by two in comparison to (18a)

$$- \sum_{k' \in \mathcal{K} \setminus \{k\}} \sum_{\ell=1}^{L} \langle \tilde{\mathcal{C}}_{k,\ell,k',\ell'}^{(\kappa)}, [\mathbf{V}_{k',\ell'}]^2 \rangle, \tag{78}$$

for $\tilde{\chi}_{k,\ell}^{(\kappa)} \triangleq \tilde{r}_{k,\ell}(V^{(\kappa)}) - \langle [\tilde{\mathcal{A}}_{k,k,\ell} V_{k,\ell}^{(\kappa)}]^2 \Psi_{k,\ell}^{-1}(V^{(\kappa)}) \rangle - \sigma \langle \Gamma_{k,\ell}^{(\kappa)} \rangle$, $\tilde{\mathcal{G}}_{k,\ell}^{(\kappa)} \triangleq \tilde{\mathcal{A}}_{k,k,\ell}^T \Psi_{k,\ell}^{-1}(V^{(\kappa)}, x^{(\kappa)}) \tilde{\mathcal{A}}_{k,k,\ell} V_{k,\ell}^{(\kappa)}$, $\Gamma_{k,\ell}^{(\kappa)} \triangleq \Psi_{k,\ell}^{-1}(V^{(\kappa)}) - \left([\tilde{\mathcal{A}}_{k,k,\ell} V_{k,\ell}^{(\kappa)}]^2 + \Psi_{k,\ell}(V^{(\kappa)}) \right)^{-1} \succeq 0$, $\tilde{\mathcal{C}}_{k,\ell,k',\ell'}^{(\kappa)} \triangleq \mathcal{A}_{k,k',\ell'}^T \Gamma_{k,\ell}^{(\kappa)} \mathcal{A}_{k,k',\ell'} \succeq 0$.

Thus we have

$$\begin{aligned}
\Phi^{(\kappa)}(\mathbf{V}) \geq & \tilde{\Phi}^{(\kappa)}(\mathbf{V}) \qquad (79)\\
\triangleq & \sum_{k=1}^{K} \sum_{\ell=1}^{L} \tilde{\gamma}_{k,\ell}^{(\kappa)} \tilde{r}_{k,\ell}^{(\kappa)}(\mathbf{V}) \\
= & \tilde{\chi}^{(\kappa)} + 2 \sum_{k=1}^{K} \sum_{\ell=1}^{L} \tilde{\gamma}_{k,\ell}^{(\kappa)} \langle \tilde{\mathcal{G}}_{k,\ell}^{(\kappa)}, \mathbf{V}_{k,\ell} \rangle \\
& - \sum_{k=1}^{K} \sum_{\ell=1}^{L} \tilde{\gamma}_{k,\ell}^{(\kappa)} \left(\sum_{\ell'=\ell}^{L} \langle \tilde{\mathcal{C}}_{k,\ell,k,\ell'}^{(\kappa)}, [\mathbf{V}_{k,\ell'}]^2 \rangle \right. \\
& \left. + \sum_{k' \in \mathcal{K} \setminus \{k\}} \sum_{\ell'=1}^{L} \langle \tilde{\mathcal{C}}_{k,\ell,k',\ell'}^{(\kappa)}, [\mathbf{V}_{k',\ell'}]^2 \rangle \right) \\
= & \tilde{\chi}^{(\kappa)} + 2 \sum_{k=1}^{K} \sum_{\ell=1}^{L} \tilde{\gamma}_{k,\ell}^{(\kappa)} \langle \tilde{\mathcal{G}}_{k,\ell}^{(\kappa)}, \mathbf{V}_{k,\ell} \rangle \\
& - \sum_{k=1}^{K} \sum_{\ell=1}^{L} \langle \tilde{\mathcal{C}}_{k,\ell}^{(\kappa)}, [\mathbf{V}_{k,\ell}]^2 \rangle, \qquad (80)
\end{aligned}$$

for $\tilde{\chi}^{(\kappa)} \triangleq \sum_{k=1}^{K} \sum_{\ell=1}^{L} \tilde{\gamma}_{k,\ell}^{(\kappa)} \chi_{k,\ell}^{(\kappa)}$, and $\tilde{\mathcal{C}}_{k,\ell}^{(\kappa)} \triangleq \sum_{\ell'=1}^{\ell} \tilde{\gamma}_{k,\ell'}^{(\kappa)} \mathcal{C}_{k,\ell',k,\ell}^{(\kappa)} + \sum_{k' \in \mathcal{K} \setminus \{k\}} \sum_{\ell'=1}^{L} \tilde{\gamma}_{k',\ell'}^{(\kappa)} \mathcal{C}_{k',\ell',k,\ell}^{(\kappa)}$.

We solve the following convex optimization problem at the κ-th iteration to generate the next feasible point $V^{(\kappa+1)}$ for (71):

$$\max_{\mathbf{V}, \boldsymbol{x}} \tilde{\Phi}^{(\kappa)}(\mathbf{V}) \quad \text{s.t.} \quad (71b), \tag{81}$$

which admits the following closed-form solution:

$$V_{k,\ell}^{(\kappa+1)} = \begin{cases} \tilde{\gamma}_{k,\ell}^{(\kappa)} (\tilde{\mathcal{C}}_{k,\ell}^{(\kappa)})^{-1} \tilde{\mathcal{G}}_{k,\ell}^{(\kappa)} \\ \text{if } \sum_{k=1}^{K} \sum_{\ell=1}^{L} ||\tilde{\gamma}_{k,\ell}^{(\kappa)} (\tilde{\mathcal{C}}_{k,\ell}^{(\kappa)})^{-1} \tilde{\mathcal{G}}_{k,\ell}^{(\kappa)}||^2 \leq 2P, \\ \\ \tilde{\gamma}_{k,\ell}^{(\kappa)} (\tilde{\mathcal{C}}_{k,\ell}^{(\kappa)} + \lambda I_{2N_{k,\ell}})^{-1} \tilde{\mathcal{G}}_{k,\ell}^{(\kappa)} \\ \text{otherwise}, \end{cases} \tag{82}$$

where λ is found from bisection search, so that we have

$$\sum_{k=1}^{K} \sum_{\ell=1}^{L} ||\tilde{\gamma}_{k,\ell}^{(\kappa)} (\tilde{\mathcal{C}}_{k,\ell}^{(\kappa)} + \lambda I_{2N_{k,\ell}})^{-1} \tilde{\mathcal{G}}_{k,\ell}^{(\kappa)}||^2 = 2P. \tag{83}$$

Similarly to its PGS dual pair in (35), we can show that $\Phi^{(\kappa)}(V^{(\kappa+1)}) > \Phi^{(\kappa)}(V^{(\kappa)})$ as far as $\Phi^{(\kappa)}(V^{(\kappa+1)}) \neq \Phi^{(\kappa)}(V^{(\kappa)})$, i.e. $V^{(\kappa+1)}$ is a better feasible point for the problem (76) and as such Algorithm 3 provides steep descent iterations for computing (71).

The computational complexity of each iteration of Algorithm 3 is given by (36).[6]

[6] $\mathcal{O}(2N_{k,\ell} \log(2N_{k,\ell})) = \mathcal{O}(N_{k,\ell} \log(N_{k,\ell}))$

Algorithm 3 IGS CBF Algorithm

1: Set $\kappa = 0$. Take a feasible point $V^{(0)}$ to satisfy the power constraint (71b).
2: **Repeat until convergence of the objective function in (71).** Iterate $V^{(\kappa+1)}$ by (82). Reset $V^{(\kappa)} \leftarrow V^{(\kappa+1)}$ and update $\tilde{\gamma}_{k,\ell}^{(\kappa)}$ by (75).
3: **Output** $V^{(\kappa)}$ as the optimal solution of (71).

Algorithm 4 Energy-efficient IGS Algorithm

1: Set $\kappa = 0$. Take a feasible point $V^{(0)}$ to satisfy the power constraint (71b).
2: **Repeat until convergence of the objective function in (85).** Iterate $V^{(\kappa+1)}$ by (92). Reset $V^{(\kappa)} \leftarrow V^{(\kappa+1)}$ and update $\tilde{\gamma}_{k,\ell}^{(\kappa)}$ by (75).
3: **Output** $V^{(\kappa)}$ as the optimal solution of (85).

B. Energy-efficiency maximization

Instead of its PGS dual pair in (37), the function of total power consumption is defined by

$$\tilde{\pi}_c(\mathbf{V}) \triangleq \frac{\alpha}{2} ||\mathbf{V}||^2 + P_{sc} + N_R \left(P_0 + \frac{P_{bt}}{2} KL \tilde{\varphi}(\mathbf{V}) \right), \tag{84}$$

and instead of (38), the problem of EE maximization is formulated as

$$\max_{\mathbf{v}} \frac{KL\Phi(\mathbf{v})}{2\tilde{\pi}_c(\mathbf{V})} \quad \text{s.t.} \quad (71b). \tag{85}$$

Like (39)-(41), the problem (85) is equivalent to the following problem:

$$\max_{\mathbf{v}} F(\mathbf{V}) \triangleq \frac{\Phi(\mathbf{V})}{\tilde{\pi}_e(\mathbf{V})} \quad \text{s.t.} \quad (71b) \tag{86}$$

where we have

$$\tilde{\pi}_e(\mathbf{V}) \triangleq \alpha ||\mathbf{V}||^2 + 2P_{sc} + 2N_R P_0, \tag{87}$$

because

$$\max\ (85) = \left(\frac{1}{KL \max\ (86)} + N_R P_{bt} \right)^{-1}. \tag{88}$$

Let $V^{(\kappa)}$ be the feasible point for (71) that is found from the $(\kappa-1)$-st iteration. Similar to its PGS counterpart in (42), we examine the following problem for generating $V^{(\kappa+1)}$:

$$\max_{\mathbf{V}} \ F^{(\kappa)}(\mathbf{V}) \triangleq \sum_{k=1}^{K} \sum_{\ell=1}^{L} \tilde{\gamma}_{k,\ell}^{(\kappa)} \tilde{r}_{k,\ell}(\mathbf{V}) - \tilde{\theta}^{(\kappa)} \tilde{\pi}_e(\mathbf{v}) \quad \text{s.t.} \quad (71b), \tag{89}$$

for $\tilde{\gamma}_{k,\ell}^{(\kappa)}$ defined in (75) and

$$\tilde{\theta}^{(\kappa)} \triangleq KL \frac{\max_{(k',\ell') \in \mathcal{K} \times \mathcal{L}} \tilde{r}_{k',\ell'}(V^{(\kappa)})}{\tilde{\pi}_e(V^{(\kappa)})}. \tag{90}$$

Recalling the function $\Phi^{(\kappa)}(\mathbf{V})$ from (80), we solve the following convex problem to generate $V^{(\kappa+1)}$:

$$\max_{\mathbf{V}} \ \tilde{F}^{(\kappa)}(\mathbf{V}) \triangleq \Phi^{(\kappa)}(\mathbf{V}) - \tilde{\theta}^{(\kappa)} \tilde{\pi}_e(\mathbf{V}) \quad \text{s.t.} \quad (71b), \tag{91}$$

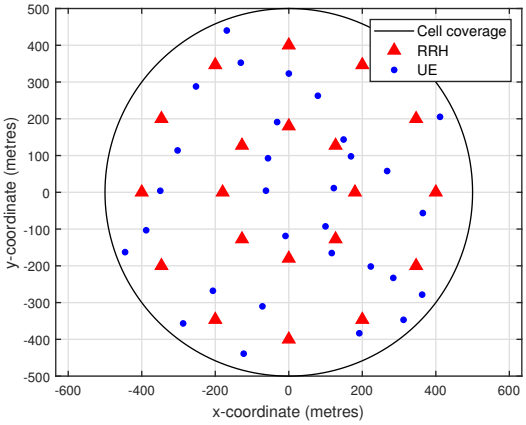

Fig. 2. The simulation scenario of 20 RRHs and 30 UEs distributed in the cell

TABLE II
PARAMETER SETTINGS

Parameter	Value
Carrier frequency / Bandwidth(Bw)	2 GHz / 10 MHz
Shadowing standard deviation	8 dB
Noise power density	-174 dBm/Hz
$(\alpha, P_{sc}, P_0, P_{bt})$ in (37)	$(1/0.4, 0.2\text{W}, 0.825\text{W}, 0.25\text{W})$

which admits the following closed-form solution:

$$V_{k,\ell}^{(\kappa+1)} = \begin{cases} \tilde{\gamma}_{k,\ell}^{(\kappa)} \left(\tilde{\mathcal{C}}_{k,\ell}^{(\kappa)} + \tilde{\theta}^{(\kappa)} \alpha I_{2N_{k,\ell}}\right)^{-1} \tilde{\mathcal{G}}_{k\ell}^{(\kappa)} \\ \text{if } \sum_{k=1}^{K}\sum_{\ell=1}^{L} ||\tilde{\gamma}_{k,\ell}^{(\kappa)} \left(\tilde{\mathcal{C}}_{k,\ell}^{(\kappa)} + \tilde{\theta}^{(\kappa)} \alpha I_{2N_{k,\ell}}\right)^{-1} \tilde{\mathcal{G}}_{k\ell}^{(\kappa)}||^2 \\ \qquad \leq 2P, \\ \tilde{\gamma}_{k,\ell}^{(\kappa)} \left(\tilde{\mathcal{C}}_{k,\ell}^{(\kappa)} + \tilde{\theta}^{(\kappa)} \alpha I_{2N_{k,\ell}} + \lambda I_{2N_{k,\ell}}\right)^{-1} \tilde{\mathcal{G}}_{k\ell}^{(\kappa)} \\ \text{otherwise}, \end{cases} \tag{92}$$

where λ is found from bisection search, so that we have

$$\begin{aligned} &\tilde{g}(\lambda) \\ &\triangleq \sum_{k=1}^{K}\sum_{\ell=1}^{L} ||\tilde{\gamma}_{k,\ell}^{(\kappa)} \left(\tilde{\mathcal{C}}_{k,\ell}^{(\kappa)} + \theta^{(\kappa)} \alpha I_{2N_{k,\ell}} + \lambda I_{2N_{k,\ell}}\right)^{-1} \tilde{\mathcal{G}}_{k\ell}^{(\kappa)}||^2 \\ &= 2P. \end{aligned} \tag{93}$$

Similar to (47), we can show $F^{(\kappa)}(V^{(\kappa+1)}) > F^{(\kappa)}(V^{(\kappa)})$ as far as $F^{(\kappa)}(V^{(\kappa+1)}) \neq F^{(\kappa)}(V^{(\kappa)})$, i.e. $V^{(\kappa+1)}$ is a better feasible point for the problem (89). Algorithm 4 thus provides steep descent iterations for the problem (85).

The computational complexity of each iteration of Algorithm 4 is given by (36).

IV. SIMULATION RESULTS

The efficiency of the proposed algorithms along with their convergence is demonstrated through the numerical results of this section. All these algorithms are centralised.

There are 200 files in the library ($F = 200$), which are labelled in the order of their popularity obeying Zipfs distribution having the popularity exponent of $\gamma_z = 0.6$. Each file is split into 4 subfiles ($L = 4$). The caching capacity of each RRH is $\mu = 1/2$. The RRHs are uniformly distributed in the circular area with radius $R = 500$ m, while the UEs are randomly distributed in the area, as illustrated in Fig. 2. Unless stated otherwise, we use the following settings: the sum transmit power is $P = 40$ dBm; the number of RRHs (UEs, resp.) is $N_R = 20$ ($K = 30$, resp.); each RRH is equipped with a three-element antenna array ($N_t = 3$) and each UE is served by the three nearest RRHs ($N_F = 3$).

The channel vector $h_{k,i}$ between the RRH i and UE k at the distance $d_{k,i}$ in km is modeled as $h_{k,i} = \sqrt{10^{-\rho_{k,i}/10}}\tilde{h}_{k,i}$, where $\rho_{k,i} = 148.1 + 37.6 \log_{10}(d_{k,i})$ (dB) is the pathloss and $\tilde{h}_{k,i}$ having independent and identically distributed complex entries is the small-scale fading [45]. The error tolerance for declaring convergence is set to $\varepsilon =$ 1e-4. The other parameters are given by Table II.

The ultimate objective of our simulations is to show that the CBF design based on the GM-throughput maximization problems (18) and (71) and their energy-efficient extension (38) and (85) help to achieve the rate target (21) for the successful delivery of all subfiles under the fronthaul capacity constrained by (19) with moderate $C < 10$ bps/Hz, under the following caching strategies:

- Caching the most popular files (CMP): Each RRH stores the $\mathcal{F}_i$ most popular files, so $|\mathcal{F}_i| = \lfloor \mu F \rfloor$ and $c_{f,\ell}^i = 1$ if and only if $f \leq |\mathcal{F}_i|$.
- Caching fractions of distinct files (CFD): Each RRH stores up to $\lfloor \mu L \rfloor$ fragments of each file that are randomly chosen, so $c_{f,\ell}^i = 1$ if and only if $\ell \in \mathcal{L}_f^i$, where $\mathcal{L}_f^i$ is a set of $\lfloor \mu L \rfloor$ random numbers from $\mathcal{L} \triangleq \{1, \ldots, L\}$, which are independent across the file f and the RRH index i.
- Random caching (RanC): Each RRH stores a set $\mathcal{F}_i$ of distinct files, which are arbitrarily selected from the F files, so $|\mathcal{F}_i| = \lfloor \mu F \rfloor$ and $c_{f,l}^i = 1$ if and only if $f \in \mathcal{F}_i$.

Note that under the above setting, these design problems involve at least $KLN_FN_t = 1080$ decision variables, ruling out any chance of using convex solver based iterative algorithms for solution.

A. Efficacy of GM-throughput maximization in limiting the fronthaul burden and meeting the rate-constraint

Fig. 3 shows that the individual subfile throughputs are quite similar, which demonstrates the benefit of the dense deployment of RRHs in the cell and as such the rate requirement (21) for the successful delivery of all subfiles is automatically satisfied. Compared to the PGS CBF Algorithm 1, the IGS CBF Algorithm 3 improves both the GM-throughput and also the individual subfile throughput.

Fig. 4 and Fig. 5 show that the fronthaul capacities required by the RRHs are more or less similar under each of the three aforementioned caching schemes, and they are the lowest under the CMP scheme.

Fig. 6 shows that the transmit powers of the RRHs are evenly distributed among twelve of the 20 RRHs located in an outer circle of the cell depicted in Fig. 2, and among the remaining eight RRHs located in an inner circle. Those in the inner circle experience better service conditions and thus

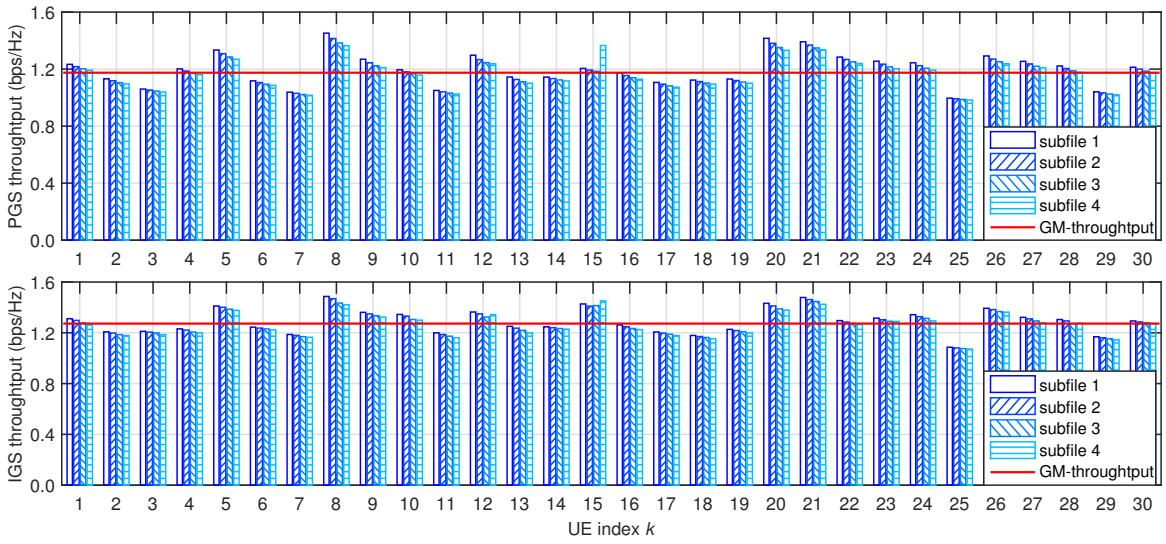

Fig. 3. The distribution of the individual subfile throughput by Algorithm 1 and 3 using the parameters of Table II

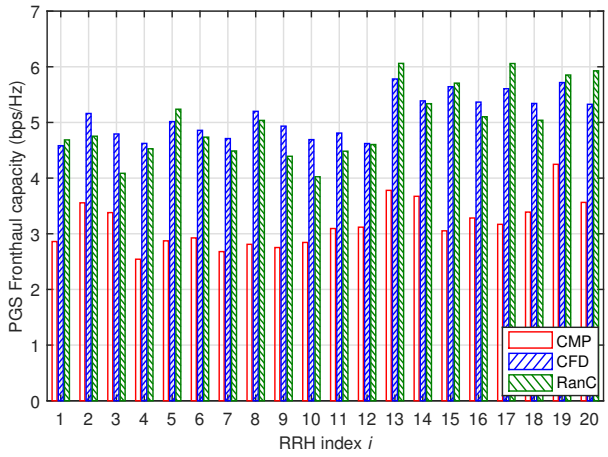

Fig. 4. The distribution of the fronthaul capacity by Algorithm 1 using the parameters of Table II

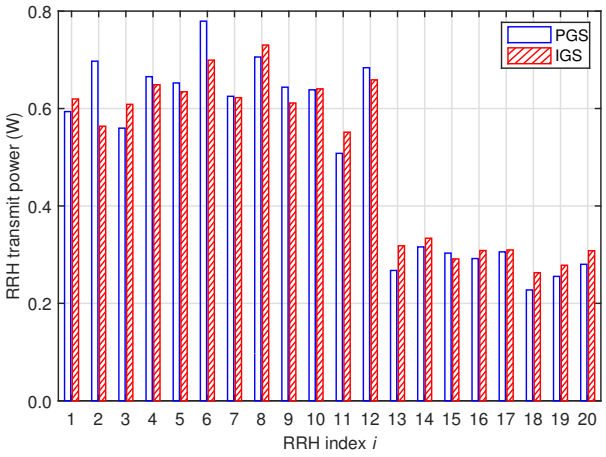

Fig. 6. The distribution of the RRH transmit power by Algorithm 1 and 3 using the parameters of Table II

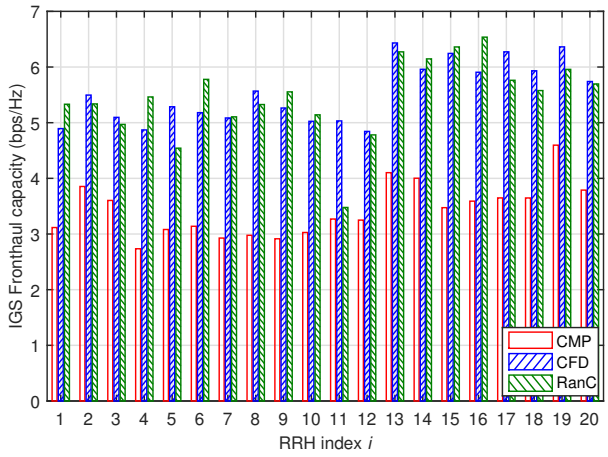

Fig. 5. The distribution of the fronthaul capacity by Algorithm 3 using the parameters of Table II

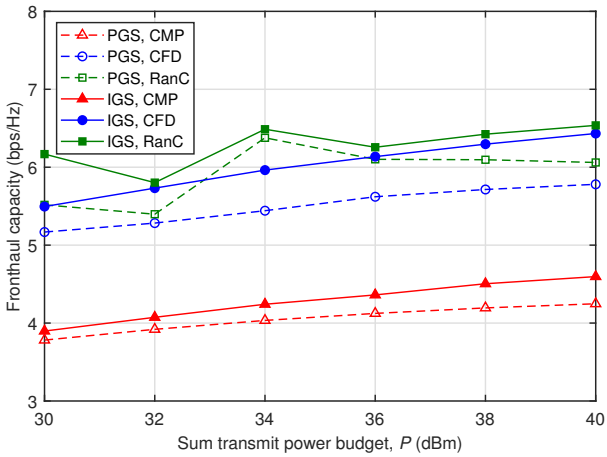

Fig. 7. The fronthaul capacity by Algorithm 1 and 3 versus the sum transmit power budget P using the parameters of Table II

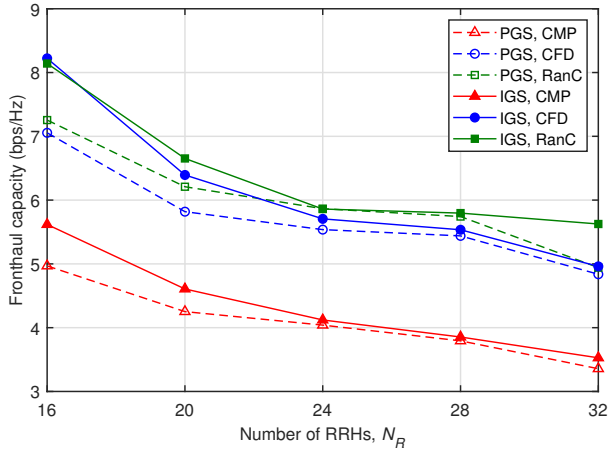

Fig. 8. The fronthaul capacity by Algorithm 1 and 3 versus the number N_R of RRHs using the parameters of Table II

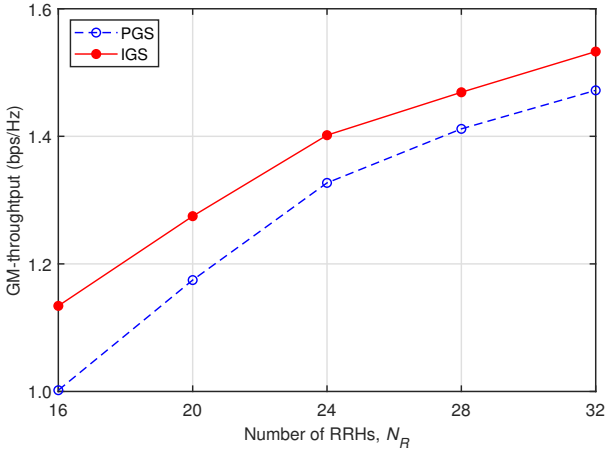

Fig. 11. The GM-throughput versus the number N_R of RRHs using the parameters of Table II

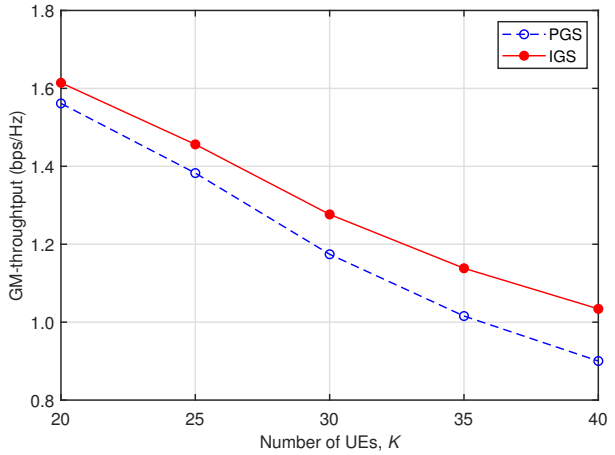

Fig. 9. The GM-throughput versus the number K of UEs using the parameters of Table II

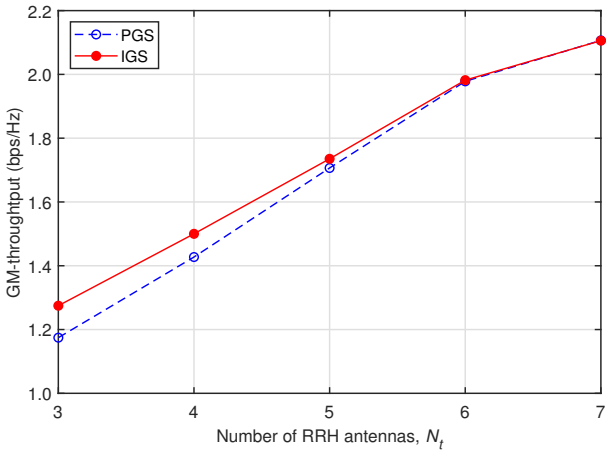

Fig. 12. The GM-throughput versus the number N_t of RRH antennas using the parameters of Table II

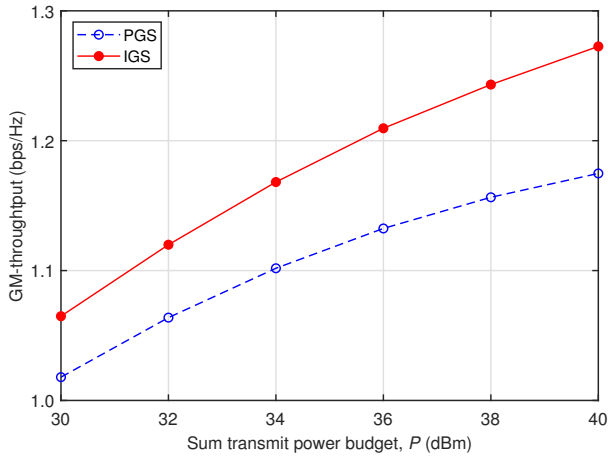

Fig. 10. The GM-throughput versus sum transmit power budget P using the parameters of Table II

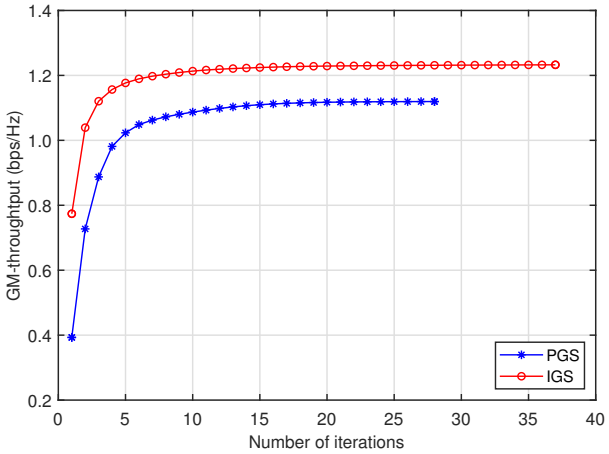

Fig. 13. The convergence behaviors of the GM-throughput maximization schemes

need lower transmit powers from the RRHs to achieve the throughput requirements.

Fig. 7 plots the fronthaul capacity defined by (20)/(74) for the PGS and IGS scenarios versus the sum transmit power budget P under the three aforementioned caching schemes. The IGS CBF Algorithm 3 requires higher fronthaul capacity than the PGS CBF Algorithm 1. Observe that the CMP scheme requires the lowest fronthaul capacity. This result

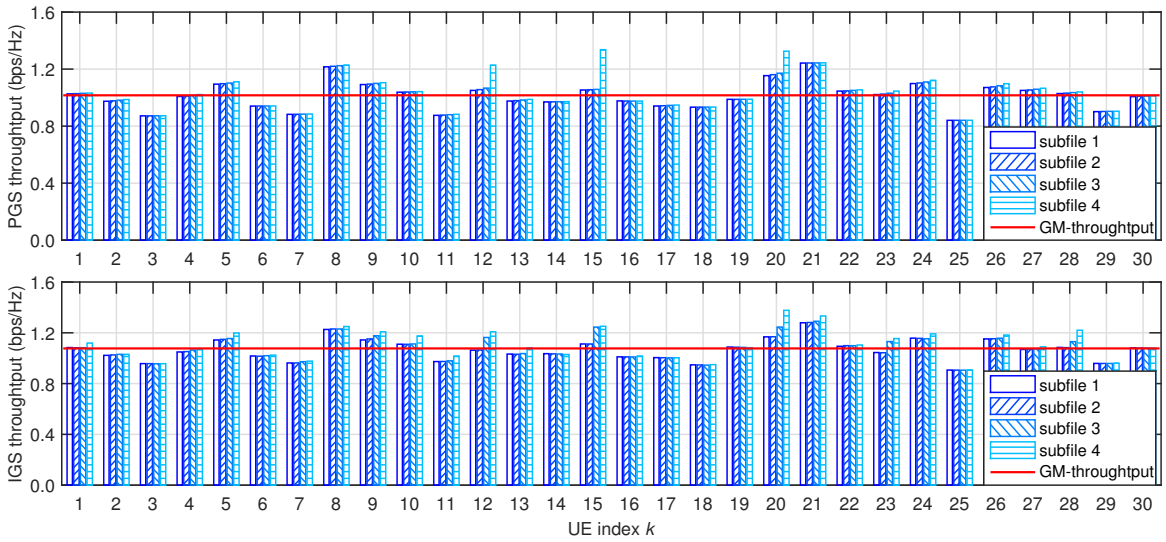

Fig. 14. The distribution of the individual subfile throughput by Algorithm 2 and 4 using the parameters of Table II

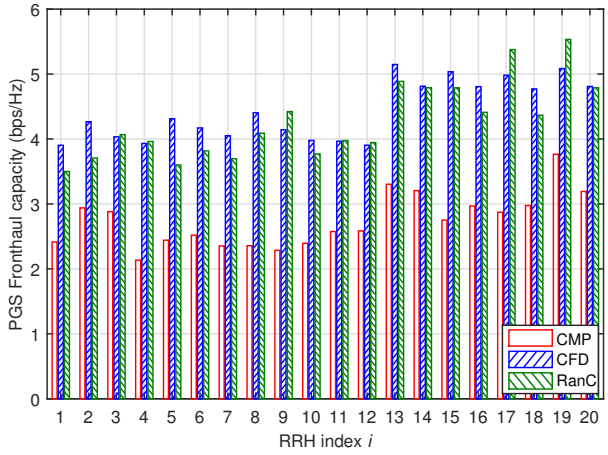

Fig. 15. The distribution of the fronthaul capacity by Algorithm 2 using the parameters of Table II

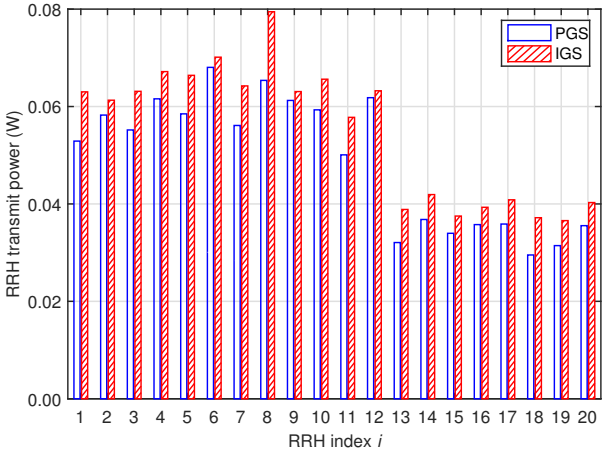

Fig. 17. The distribution of the RRH transmit power by Algorithm 2 and 4 using the parameters of Table II

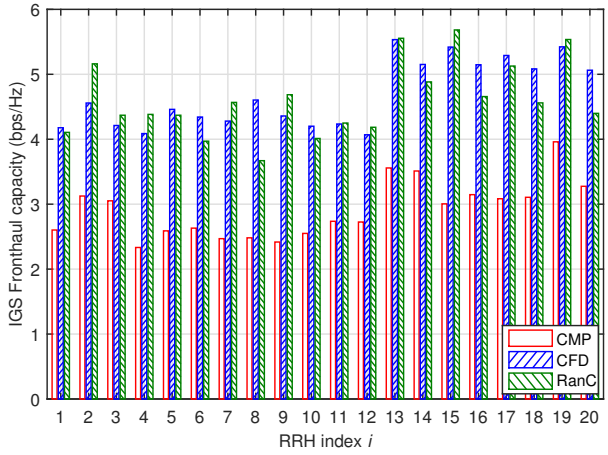

Fig. 16. The distribution of the fronthaul capacity by Algorithm 4 using the parameters of Table II

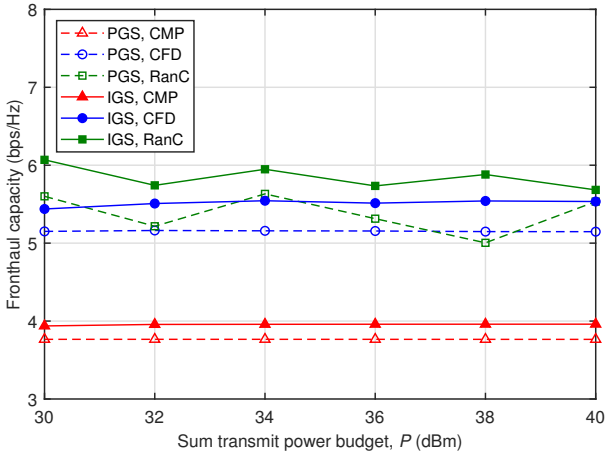

Fig. 18. The fronthaul capacity by Algorithm 2 and 4 versus the sum transmit power budget P using the parameters of Table II

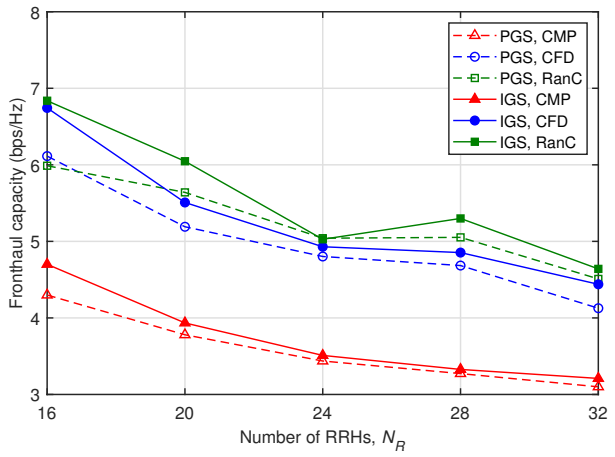

Fig. 19. The fronthaul capacity by Algorithm 2 and 4 versus the number N_R of RRHs using the parameters of Table II

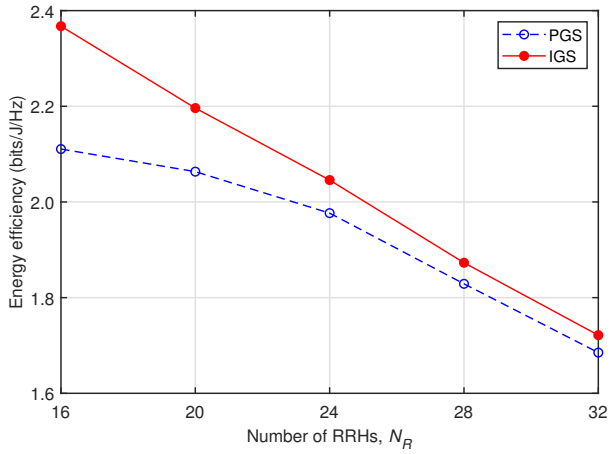

Fig. 20. Energy efficiency versus the number N_R of RRHs using the parameters of Table II

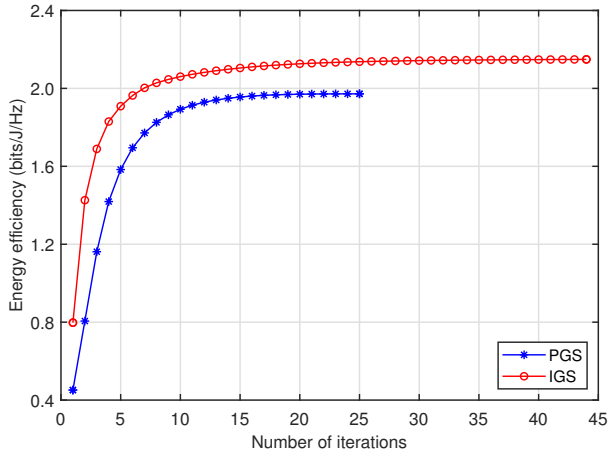

Fig. 21. The convergence behaviors of the energy-efficiency maximization schemes

clearly shows that caching the most popular files at the RRHs indeed helps to mitigate the fronthauling burden. Naturally, the fronthaul requirement increases monotonically with P. However, under the RanC scheme it fluctuates owing to the

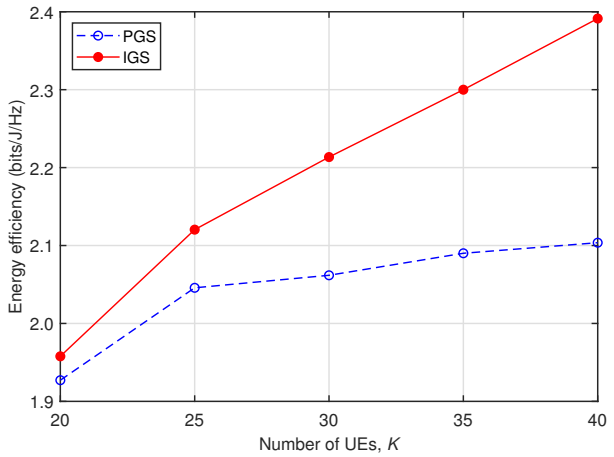

Fig. 22. Energy efficiency versus the number K of UEs using the parameters of Table II

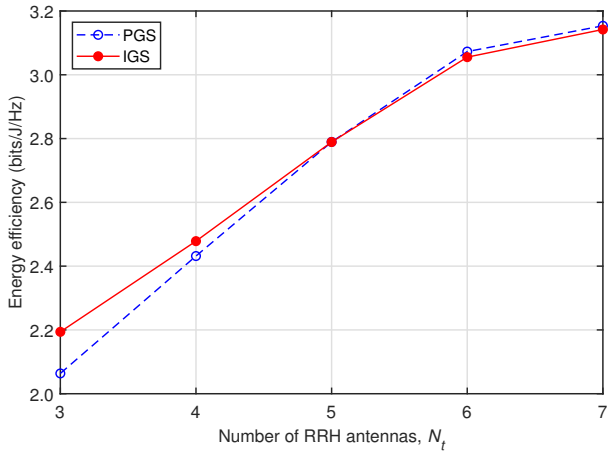

Fig. 23. Energy efficiency versus the number N_t of RRH antennas using the parameters of Table II

random selection of the cached files from the file library. This trend is consistent with that observed in Fig. 4 and Fig. 5.

Fig. 8 shows that the fronthaul capacity requirement is monotonically decreasing with the number N_R of RRHs, as expected. The capacity gap between IGS and PGS fluctuates under the RanC scheme due to its random nature. Again, caching the most popular files at the RRHs is the most appropriate require for alleviating the fronthaul load.

Fig. 9 shows that the GM-throughput is monotonically decreasing with the number K of UEs. This is because the same resources must be shared among the increased numbers of UEs, hence dropping the subfile throughput. The performance gap between the IGS CBF Algorithm 3 and the PGS CBF Algorithm 1 becomes wider under more UEs, showing that the IGS CBF Algorithm 3 has the edge, since it is capable of more UEs.

Fig. 10, Fig. 11, and Fig. 12 show that the GM-throughput is monotonically increasing both with the power budget P, as well as with the number N_R of RRHs - which tend to be closer to the UEs - and with the number N_t of RRH antennas. Again, the IGS CBF Algorithm 3 outperforms the PGS CBF Algorithm 1 in all these figures. However, there is almost no

difference between the two performances in Fig. 12 for $N_t > 5$, because in general PGS also performs well with the aid of more transmit antennas. Fig. 9, Fig. 11, and Fig. 12 also demonstrate that the IGS CBF Algorithm 3 has an advantage over its PGS counterpart Algorithm 1 for less RRHs, or smaller numbers of RRH antennas, and larger numbers of UEs.

Fig. 13 shows the convergence behaviors of the GM-throughput maximization schemes in simulating Fig. 10 with $P = 40$ dBm. The GM-throughput improves after each iteration.

B. EE optimization

This subsection demonstrates the efficacy of the PGS EE Algorithm 2 and of the IGS EE Algorithm 4 in achieving the energy-efficient delivery of all subfiles at moderate fronthaul rates, while meeting the individual user-rate requirements.

Fig. 14 shows that the individual subfile throughputs automatically become similar both without imposing constraints.

Fig. 15 and Fig. 16 plot the fronthaul capacities required by the RRHs for satisfying the fronthaul link capacity constraint (19) with C being substantially below 10 bps/Hz.

Further, Fig. 17 plots the transmit powers of the RRHs calculated by solving the problems (39) and (85) by the PGS EE Algorithm 2 and the IGS EE Algorithm 2. The trends are similar to those in Fig. 6. Those RRHs that are located in the outer circle of Fig. 2 use more transmit power to reach their served UEs.

Fig. 18 reveals that the CMP requires lower fronthaul capacities than the other two. The IGS EE Algorithm 4 requires higher fronthaul capacity than its PGS counterpart Algorithm 2. As a further trend, the fronthaul capacity is insensitive to the power budget P under the CMP and CFD schemes, but slightly fluctuates under the RanC scheme.

Fig. 19 plots the fronthaul capacity versus the number N_R of RRHs, which tends to decay with N_R, similarly to Fig. 8.

Fig. 20 shows that the EE drops when the number of RRHs increases. Thus, the increase of the total subfile throughput does not keep up with the growth of the total power consumption when more RRHs are deployed. Fig. 20 also shows that the advantage of IGS diminishes with more RRHs deployed.

Fig. 21 shows the convergence behaviors of the EE maximization schemes in simulating Fig. 20 with $N_R = 20$. The EE improves after each iteration. The PGS algorithm converges faster than the IGS algorithm, since the latter has to handle more decision variables.

Fig. 22 shows that the IGS EE Algorithm 4 handles the EE requirement much better than its counterpart PGS EE Algorithm 2. Finally, Fig. 23 shows that the EE of the PGS EE Algorithm 2 and of the IGS EE Algorithm 4 by the number N_t of RRH antennas is quite similar, which is consistent with what observed from Fig. 12.

V. Conclusions

The problem of CBF design for F-RANs equipped with capacity-constrained fronthaul links and RRH caches is computationally challenging due to the large numbers of beamforming variables involved. We tackled this problem by conceiving the novel concept of the GM-throughput maximization subject to total transmit power constraints. We derived new low-complexity closed-form expressions to be evaluated at each iteration of our optimization procedure. The simulations conducted for large-scale F-RANs showed that indeed GM-through maximization only requires moderate fronthaul throughput and limited RRH caches, while meeting the content throughput requirements. An extension of the proposed approach to the problem of three-dimensional (3D) beamforming design is under current study.

References

[1] M. Chiang and T. Zhang, "Fog and IoT: An overview of research opportunities," *IEEE Internet Things J.*, vol. 3, pp. 854–864, Jun. 2016.

[2] M. Peng, S. Yan, K. Zhang, and C. Wang, "Fog-computing-based radio access networks: issues and challenges," *IEEE Network*, vol. 30, pp. 46–53, Apr. 2016.

[3] Y.-J. Ku, D.-Y. Lin, C.-F. Lee, P.-J. Hsieh, H.-Y. Wei, C.-T. Chou, and A.-C. Pang, "5G radio access network design with the fog paradigm: Confluence of communications and computing," *IEEE Commun. Mag.*, vol. 55, pp. 46–52, Apr. 2017.

[4] X. Zhu et al., "Improving video performance with edge servers in the fog computing architecture," *J. Intel Technol.*, vol. 19, pp. 202–224, Apr. 2015.

[5] C. Mouradian, D. Naboulsi, S. Yangui, R. H. Glitho, M. J. Morrow, and P. A. Polakos, "A comprehensive survey on fog computing: State-of-the-art and research challenges," *IEEE Commun. Surv. Tut.*, vol. 20, no. 1, pp. 416–464, 2018.

[6] T. Dang and M. Peng, "Joint radio communication caching and computing design for mobile virtual reality delivery in fog radio access networks," *IEEE J. Select. Areas Commun.*, vol. 37, pp. 1594–1607, Jul. 2019.

[7] R. Karasik, O. Simeone, and S. S. Shitz, "How much can D2D communication reduce content delivery latency in fog networks with edge caching?," *IEEE Trans. Commun.*, vol. 68, pp. 2308–2323, Apr. 2020.

[8] H. Freeman and T. Zhang, "The emerging era of fog computing and networking [the president's page]," *IEEE Commun. Mag.*, vol. 54, pp. 4–5, Jun. 2016.

[9] G. S. Paschos, G. Iosifidis, M. Tao, D. Towsley, and G. Caire, "The role of caching in future communication systems and networks," *IEEE J. Select. Areas Commun.*, vol. 36, pp. 1111–1125, Jun. 2018.

[10] S. M. Azimi, O. Simeone, A. Sengupta, and R. Tandon, "Online edge caching and wireless delivery in fog-aided networks with dynamic content popularity," *IEEE J. Select. Areas Commun.*, vol. 36, pp. 1189–1202, Jun. 2018.

[11] S. He, J. Ren, J. Wang, Y. Huang, Y. Zhang, W. Zhuang, and S. Shen, "Cloud-edge coordinated processing: Low-latency multicasting transmission," *IEEE J. Select. Areas Commun.*, vol. 37, pp. 1144–1158, May 2019.

[12] X. Wang, M. Chen, T. Taleb, A. Ksentini, and V. Leung, "Cache in the air: exploiting content caching and delivery techniques for 5G systems," *IEEE Commun. Mag.*, vol. 52, pp. 131–139, Feb. 2014.

[13] D. Liu and C. Yang, "Energy efficiency of downlink networks with caching at base stations," *IEEE J. Sel. Areas Commun.*, vol. 34, pp. 907–922, Apr. 2016.

[14] H. Ochiai, P. Mitran, H. V. Poor, and V. Tarokh, "Collaborative beamforming for distributed wireless ad hoc sensor networks," *IEEE Trans. Signal Process.*, vol. 53, pp. 4110–4124, Nov. 2005.

[15] L. Dong, A. P. Petropulu, and H. V. Poor, "Weighted cross-layer cooperative beamforming for wireless networks," *IEEE Trans. Signal Process.*, vol. 57, pp. 3240–3251, Aug. 2009.

[16] U. Rashid, H. D. Tuan, P. Apkarian, and H. H. Kha, "Globally optimized power allocation in multiple sensor fusion for linear and nonlinear networks," *IEEE Trans. Signal Process.*, vol. 60, pp. 903–915, Feb. 2012.

[17] N. Chatzipanagiotis, Y. Liu, A. Petropulu, and M. M. Zavlanos, "Distributed cooperative beamforming in multi-source multi-destination clustered systems," *IEEE Trans. Signal Process.*, vol. 62, pp. 6105–6117, Dec. 2014.

[18] S. Zaidi and S. Affes, "Distributed collaborative beamforming design for maximized throughput in interfered and scattered environments," *IEEE Trans. Commun.*, vol. 63, pp. 4905–4918, Dec. 2015.

[19] S. Jayaprakasam, S. K. A. Rahim, and C. Y. Leow, "Distributed and collaborative beamforming in wireless sensor networks: Classifications trends and research directions," *IEEE Commun. Surv. Tuts.*, vol. 19, no. 4, pp. 2092–2116, 2017.

[20] J. A. Bengua, H. D. Tuan, T. Q. Duong, and H. V. Poor, "Joint sensor and relay power control in tracking Gaussian mixture targets by wireless sensor networks," *IEEE Trans. Signal Process.*, vol. 66, pp. 492–506, Feb. 2018.

[21] M. Tao, E. Chen, H. Zhou, and W. Yu, "Content-centric sparse multicast beamforming for cache-enabled cloud RAN," *IEEE Trans. Wirel. Commun.*, vol. 15, pp. 6118–6131, Sept. 2016.

[22] S.-H. Park, O. Simeone, and S. Shamai, "Joint optimization of cloud and edge processing for fog radio access networks," *IEEE Trans. Wirel. Commun*, vol. 15, pp. 7621–7632, Nov. 2016.

[23] J. Kang, O. Simeone, J. Kang, and S. Shamai, "Fronthaul compression and precoding design for C-RANs over ergodic fading channel," *IEEE Trans. Veh. Technol.*, vol. 65, pp. 5022–5032, Jul. 2016.

[24] J. Kang, O. Simeone, J. Kang, and S. Shamai, "Layered downlink precoding for C-RAN systems with full dimensional MIMO," *IEEE Trans. Vehic. Technol.*, vol. 66, pp. 2170–2182, Mar. 2017.

[25] M. Tao, D. Gunduz, F. Xu, and J. S. P. Roig, "Content caching and delivery in wireless radio access networks," *IEEE Trans. Commun.*, vol. 67, pp. 4724–4749, Jul. 2019.

[26] H. T. Nguyen, H. D. Tuan, T. Q. Duong, H. V. Poor, and W.-J. Hwang, "Nonsmooth optimization algorithms for multicast beamforming in content-centric fog radio access networks," *IEEE Trans. Signal Process.*, vol. 68, pp. 1455–1469, 2020.

[27] K. Wang, J. Li, Y. Yang, W. Chen, and L. Hanzo, "Content-centric heterogeneous fog networks relying on energy efficiency optimization," *IEEE Trans. Vehic. Techn.*, vol. 69, pp. 13579–13592, Nov. 2020.

[28] W. Zhu, H. D. Tuan, E. Dutkiewicz, Y. Fang, and L. Hanzo, "A new class of structured beamforming for content-centric fog radio access networks," *IEEE Trans. Commun.*, vol. 69, pp. 7269–7282, Nov. 2021.

[29] H. Shin, S. Park, H. Park, and I. Lee, "A new approach of interference alignment through asymmetric complex signaling and multiuser diversity," *IEEE Trans. Wirel. Commun.*, vol. 11, pp. 880–884, Mar. 2012.

[30] C. Hellings, M. Joham, and W. Utschick, "QoS feasibility in MIMO broadcast channels with widely linear transceivers," *IEEE Signal Process. Letts.*, vol. 20, pp. 1134–1137, Nov. 2013.

[31] S. Lagen, A. Agustin, and J. Vidal, "Coexisting linear and widely linear transceivers in the MIMO interference channel," *IEEE Trans. Signal Process.*, vol. 64, pp. 652–664, Feb. 2016.

[32] A. A. Nasir, H. D. Tuan, T. Q. Duong, and H. V. Poor, "Improper Gaussian signaling for broadcast interference networks," *IEEE Signal Process. Lett.*, vol. 26, pp. 808–812, Jun. 2019.

[33] H. D. Tuan, A. A. Nasir, H. H. Nguyen, T. Q. Duong, and H. V. Poor, "Non-orthogonal multiple access with improper Gaussian signaling," *IEEE J. Selec. Topics Signal Process.*, vol. 13, pp. 496–507, Mar. 2019.

[34] H. Yu, H. D. Tuan, T. Q. Duong, Y. Fang, and L. Hanzo, "Improper Gaussian signaling for integrated data and energy networking," *IEEE Trans. Commun.*, vol. 68, pp. 3922–3934, Jun. 2020.

[35] H. Yu, H. D. Tuan, A. A. Nasir, T. Q. Duong, and L. Hanzo, "Improper Gaussian signaling for computationally tractable energy and information beamforming," *IEEE Trans. Veh. Techn.*, vol. 69, pp. 13990–13995, Nov. 2020.

[36] P. J. Schreier and L. L. Scharf, *Statistical Signal Processing of Complex-Valued Data: The Theory of Improper and Noncircular Signals*. Cambridge Univ. Press, 2010.

[37] H. Yu, H. D. Tuan, E. Dutkiewicz, H. V. Poor, and L. Hanzo, "Maximizing the geometric mean of user-rates to improve rate-fairness: Proper vs. improper Gaussian signaling," *IEEE Trans. Wirel. Commun.*, vol. 21, pp. 295–309, Jan. 2022.

[38] H. H. M. Tam, H. D. Tuan, and D. T. Ngo, "Successive convex quadratic programming for quality-of-service management in full-duplex MU-MIMO multicell networks," *IEEE Trans. Commun.*, vol. 64, pp. 2340–2353, Jun. 2016.

[39] A. Sengupta, R. Tandon, and O. Simeone, "Fog-aided wireless networks for content delivery: Fundamental latency tradeoffs," *IEEE Trans. Info. Theory*, vol. 63, pp. 6650–6678, Oct. 2017.

[40] U. Rashid, H. D. Tuan, and H. H. Nguyen, "Relay beamforming designs in multi-user wireless relay networks based on throughput maximin optimization," *IEEE Trans. Commun.*, vol. 61, pp. 1739–1749, May 2013.

[41] H. Q. Ngo, A. Ashikhmin, H. Yang, E. G. Larsson, and T. L. Marzetta, "Cell-free massive MIMO versus small cells," *IEEE Trans. Wirel. Commun.*, vol. 16, pp. 1834–1849, Mar. 2017.

[42] A. A. Nasir, H. D. Tuan, H. Q. Ngo, T. Q. Duong, and H. V. Poor, "Cell-free massive MIMO in the short blocklength regime for URLLC," *IEEE Trans. Wirel. Commun.*, vol. 20, pp. 5861–5871, Sept. 2021.

[43] G. H. Golub and C. F. Van Loan, *Matrix Computations (fourth edition)*. Johns Hopkins Univ. Press, 2012.

[44] B. Dai and W. Yu, "Energy efficiency of downlink transmission strategies for cloud radio access networks," *IEEE J. Sel. Areas Commun.*, vol. 34, pp. 1037–1050, Apr. 2016.

[45] E. Bjornson, M. Kountouris, and M. Debbah, "Massive MIMO and small cells: Improving energy efficiency by optimal soft-cell coordination," in *Proc. IEEE ICT*, pp. 1–5, May 2013.