

Coordinated Hybrid Precoding Design in Millimeter Wave Fog-RAN

Wanming Hao, Gangcan Sun, Osamu Muta, Jiankang Zhang, Shouyi Yang

Abstract—In this paper, we propose a femto base station cluster (FBSC)-based fog radio access network (F-RAN) structure, where multiple FBSCs coordinately serve users in each FBSC with mmWave band. In addition, the FBS head (FBSH) in each FBSC receives data from the central processor (CP) via fronthaul link with microwave band. Based on this, we formulate a transmission delay minimization problem by optimizing coordinated hybrid analog/digital precoding under transmit power constraints of each FBSC in the FBSCs and the BS in the CP. Then, we propose a coordinated analog precoding scheme for multiple FBSCs within the same FBSC. Meanwhile, we equivalently divide the original problem into two independent optimization problems, and propose an effective iterative algorithm to solve them. Simulation results show the effectiveness of our proposed algorithms.

Index Terms—Hybrid precoding, transmission delay, F-RAN

I. INTRODUCTION

To satisfy the rapid increasing rate demand in future wireless communications, advanced communication techniques have been proposed such as multiple point coordinated transmission, millimeter wave (mmWave) communication, massive multiple-input multiple-output (mMIMO), and cloud radio access network (C-RAN) [1]. However, there still exist challenges to apply the above techniques. For example, although the C-RAN can effectively coordinate the large-scale resource scheduling and allocation by baseband unit (BBU) pool for reducing the interference and improving the transmission efficiency, the fronthaul link is still insufficient to provide huge capacity requirement [2], [3]. In addition, mmWave has also obtained the great attention owing to its wide bandwidth. However, mmWave signals suffer from the high propagation loss, and thus the large-scale antenna array is equipped at the base station (BS) [4]. Meanwhile, to reduce the hardware cost and energy consumption, the sparse radio frequency (RF) chain antenna array structure is usually adopted, which results in more challenge in mmWave communications.

Recently, there have been several works to tackle the above challenges. The fog RAN (F-RAN), where the conventional BS is equipped with signal processing and data storage functionalities, has been proposed [5]. The BS can pre-pick the

most frequently requested files to the local caches, reducing the fronthaul overhead and lowering the latency. In [6], the authors propose a cache-based two-level transmission scheme. Then, they formulate a sum rate maximization problem in each level transmission, and develop two centralized and decentralized optimization algorithms to solve the formulated problem. In order to lower the latency, a loosely coupled architecture and beamforming optimization algorithm is proposed in [7]. The authors in [8] investigate the dynamic BS clustering and multi-cast beamforming problem according to the caching status, and a joint BS clustering and beamforming optimization scheme is proposed to minimize the total network cost. In [9], the authors study the optimization problem of hybrid cache placement of data in a coordinated relaying network. In [10], the authors investigate the fundamental limits of the higher signal-to-interference-plus-noise-ratio (SINR) metric and termed normalized delivery time from an information theoretic viewpoint, and propose a placement and delivery strategy that effectively reduces the latency. The authors in [11] study the joint optimal resource allocation and probabilistic caching design for device-to-device communication in a heterogeneous network with full-duplex relays, and the advanced cache protocol and low-complexity algorithm are proposed. In addition, hybrid analog/digital precoding design in mmWave communication is also a challenge. Several advanced design schemes have been proposed, like [12]-[14], and they first design the analog precoding according to the predefined codebook. Then, the equivalent channel is obtained according to designed analog precoding, and digital precoding is designed by zero-forcing (ZF) technique or convex optimization theory.

However, most of the related work are based on the single BS model, like [7], [10]-[13]. In fact, coordinative transmission among BSs will be necessary due to the dense deployment of BSs in mmWave communication. [5], [6], [8] only consider the spectrum efficiency (SE) or energy efficiency (EE) optimization problem, the minimization delay is not investigated. In addition, the wireless fronthaul link is also not considered in [11]. Unlike the previous system structure and works, in this paper, we propose a new femto BS cluster (FBSC)-based F-RAN structure with hybrid mmWave/microwave communications, and the main contributions include:

- A new FBSC-based F-RAN structure is proposed. The users are coordinately served by all FBSCs in each FBSC with mmWave band, and the FBS head (FBSH) receives data from the central processor (CP) with microwave band. Meanwhile, FBSH takes charge of the resource allocation and scheduling in each FBSC.
- We formulate a minimization transmission delay problem by optimizing hybrid analog/digital precoding, while the

This work was supported by the National Natural Science Foundation of China under Grant U1604159. (*Corresponding author: Gangcan Sun.*)

W. Hao, G. Sun, and S. Yang are with the School of Information Engineering, Zhengzhou University, Zhengzhou 450001, China, and W. Hao is also with the 5G Innovation Centre, Institute of Communication Systems, University of Surrey, Guildford GU2 7XH, U.K. (E-mail: wnhao@hotmail.com, {iegsun, iesyyang}@zsu.edu.cn)

O. Muta is with Center for Japan-Egypt Cooperation in Science and Technology, Kyushu University, Fukuoka 819-0395, Japan. (E-mails: muta@ieee.org, ait.kyushu-u.ac.jp)

J. Zhang is with the School of Electronics and Computer Science, University of Southampton, Southampton SO17 1BJ, U.K. (e-mail: jz09v@ecs.soton.ac.uk).

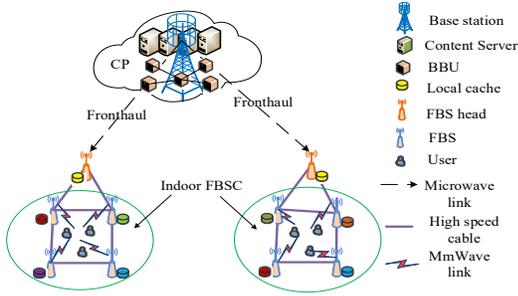


Fig. 1: System model of the FBSC-based mmWave F-RAN.

limited size of cached files in each FBSC and maximum transmit power for the BS in the CP and each FBS are imposed. We first transform the original problem into two independent ones, i.e., the FBSs' hybrid precoding design problem and the BS's precoding design problem.

- For the former, we propose a coordinated analog precoding design scheme and then, a successive convex approximation (SCA)-based iterative algorithm is proposed to optimize digital precoding. For the latter, we apply the similar iterative algorithm to solve it.

The remainder of this paper is organized as follows. Section II presents the system model and optimization problem. In Section III, we give the proposed algorithm to solve the formulated problem. Simulation results are given in Section IV. We conclude this paper in Section V.

Notations: Following notations are used throughout this paper: $(\cdot)^H$ and $(\cdot)^T$ denote Hermitian transpose and transpose, respectively, $\|\cdot\|$ represents the Frobenius norm. $\text{Re}(\cdot)$ and $\text{Tr}(\cdot)$ denote real number and trace operations, respectively. $\text{Diag}(\mathbf{f}_1, \dots, \mathbf{f}_M)$ is a diagonal matrix.

II. SYSTEM MODEL AND PROBLEM FORMULATION

We consider a cache-enable F-RAN¹ consisting of one CP and L ($\mathcal{L} = \{1, \dots, L\}$) FBSCs as shown in Fig. 1. The FBSC is suitable to apply to a small area which owns ultra-dense user distribution, e.g., supermarket, stadium, station and so on. Due to the short distance among FBSs, it is convenient to connect all FBSs by high speed cables within each FBSC. In general, for the deployment of ultra-dense FBSs, there must exist a lot of FBSHs. It is unrealistic with wired connected between all FBSHs and CP due to the high cost. In addition, the capacity provided via wired connection is fixed and inflexible, and it is difficult to dynamically adjust the traffic according to the users' demands. Therefore, we adopt the microwave as the fronthaul link carrier. There are one microwave-band BS equipped with S ($S \geq L$) antennas, one BBU pool and several large-scale content servers in the CP. Each FBSC includes one FBSH equipped with single antenna, M ($\mathcal{M} = \{1, \dots, M\}$) mmWave FBSs equipped with N ($\mathcal{N} = \{1, \dots, N\}$) antennas connected to a dedicated RF chain, and K ($\mathcal{K} = \{1, \dots, K\}$) users. We focus

¹In this paper, we assume that the BS and FBSs are based on 3GPP standard, so that users can receive data and telephone voice services. In addition, we only consider single F-RANs. For the multiple F-RANs in a system, to further reduce the interference and improve the SE, adjacent F-RANs can operate at the different frequencies, and no-adjacent F-RANs operate at the same frequency.

on an indoor scenario, and the FBSH is installed in outdoor and connected to indoor FBSs with high speed cable. In addition, we assume that the perfect channel state information (CSI) are available at the CP and FBSCs.

Assume that the contents or files required by users are cached in content servers of the CP, denoted by $\mathcal{F} = \{1, \dots, F\}$. Without loss of generality, we assume that all files are of the same size². We assume that all access point clusters (FBSCs) are relatively independent, and they do not share the cached contents. However, the contents can be shared among FBSs within each FBSC via high speed cable, namely multiple FBSs cooperatively serve users at each FBSC [17]. We denote binary variable $a_{l,m}^f$ as the state of cache file f ($f \in \mathcal{F}$) in the m th FBS of the l th FBSC, which can be expressed as

$$a_{l,m}^f = \begin{cases} 1, & \text{if } f \text{ is cached in FBS } (l, m), \\ 0, & \text{otherwise.} \end{cases} \quad (1)$$

Meanwhile, $\sum_{m=1}^M a_{l,m}^f = 1$ or $\sum_{m=1}^M a_{l,m}^f = 0$ ($l \in \mathcal{L}$) denotes that different FBSCs cache different files.

If a file required by user (l, k) is cached in the l th FBSC, all FBSs coordinately transmit the file to this user, and the received signal can be expressed as

$$y_{l,k}^{\text{FBS}} = \mathbf{h}_{l,k} \mathbf{F}_l \mathbf{v}_{l,k} x_{l,k}^{\text{FBS}} + \sum_{i \neq k} \mathbf{h}_{l,k} \mathbf{F}_l \mathbf{v}_{l,i} x_{l,i}^{\text{FBS}} + n_{l,k}, \quad (2)$$

where $\mathbf{h}_{l,k} = [(\mathbf{h}_{l,k}^1)^T, (\mathbf{h}_{l,k}^2)^T, \dots, (\mathbf{h}_{l,k}^M)^T]^T$, and $\mathbf{h}_{l,k}^m \in \mathbb{C}^{1 \times N}$ denotes the channel coefficient from the FBS (l, m) to user (l, k) . $\mathbf{v}_{l,k} \in \mathbb{C}^{M \times 1}$ and $x_{l,k}^{\text{FBS}}$ represent the digital precoding and transmit signal for user (l, k) , respectively. $n_{l,k}$ is an independent and identically distributed (i.i.d.) additive white Gaussian noise (AWGN) defined as $\mathcal{CN}(0, \delta^2)$. In addition, the analog precoding matrix \mathbf{F}_l can be written as $\mathbf{F}_l = \text{Diag}(\mathbf{f}_{l,1}, \mathbf{f}_{l,2}, \dots, \mathbf{f}_{l,M})$, where $\mathbf{f}_{l,m} \in \mathbb{C}^{N \times 1}$ represents the analog precoding at FBS (l, m) with the same amplitude $1/\sqrt{N}$ but different phases. In (2), the first term stands for the desired signal, and the second term represents the intra-cluster interference. Here, we do not consider the inter-cluster interference since mmWave communication is usually noise-limited. Moreover, for our indoor scenario, signal interference between two different indoor sites is negligible [18]. Thus, the achievable rate of user (l, k) can be calculated by $R_{l,k}^{\text{FBS}} = \log_2(1 + \gamma_{l,k}^{\text{FBS}})$, where $\gamma_{l,k}^{\text{FBS}}$ is

$$\gamma_{l,k}^{\text{FBS}} = \frac{|\mathbf{h}_{l,k} \mathbf{F}_l \mathbf{v}_{l,k}|^2}{\sum_{i \neq k} |\mathbf{h}_{l,k} \mathbf{F}_l \mathbf{v}_{l,i}|^2 + \delta^2}. \quad (3)$$

On the other hand, if a file required by user (l, k) is not cached in the l th FBSC, this file first needs to be transformed to the l th FBSH from the CP via fronthaul link at microwave-band. The received signal at l th FBSH can be expressed as

$$y_l^{\text{CP}} = \mathbf{g}_l \mathbf{w}_l x_l^{\text{CP}} + \sum_{i \neq l} \mathbf{g}_l \mathbf{w}_i x_i^{\text{CP}} + n_l, \quad (4)$$

where $\mathbf{g}_l \in \mathbb{C}^{1 \times S}$, $\mathbf{w}_l \in \mathbb{C}^{S \times 1}$ and x_l^{CP} , respectively, mean the channel coefficient, precoding vector and transmit signal from the CP to the l th FBSH. Then, the achievable rate of the l th FBSH can be calculated by $R_l^{\text{CP}} = \log_2(1 + \gamma_l^{\text{CP}})$, where γ_l^{CP} is

²In general, how to cache contents depends on the user behavior, the delay requirement and the long-term information popularity distribution, which can be realized by learning-based methods, such as based on deep Q-learning [15] and artificial intelligence-based multi-timescale framework [16]. In this paper, we mainly focus on the resource allocation for a given caching placement. Note that the proposed method can be applied to other caching placement scenarios, and the research of caching placement strategy is out scope of this paper.

Algorithm 1: Analog Precoding Design for the l th FBSC

```

1 Initialize  $\mathbf{F}_l = \mathbf{0}_{NM \times M}$ ,  $m = 1$ .
2 while  $m \leq M$  do
3   for  $k = 1 : K$  do
4     for  $n = 1 : N$  do
5       Computer (10) and obtain  $\mathbf{f}_{l,m}(n) = \frac{1}{\sqrt{N}} \exp(j2\pi\hat{\alpha}/2^B)$ .
6     end for
7      $\mathbf{F}_l((m-1)N + 1 : mN, m) = \mathbf{f}_{l,m}$ ,  $m \leftarrow m + 1$ .
8     If  $m > M$  then Break.
9   end for
10 end while
  
```

the SINR and can be expressed as

$$\gamma_l^{\text{CP}} = \frac{|\mathbf{g}_l \mathbf{w}_l|^2}{\sum_{i \neq l} |\mathbf{g}_l \mathbf{w}_i|^2 + \delta^2}. \quad (5)$$

In this paper, we mainly focus on the transmission delay, and it includes the fronthaul link transmission (from CP to FBSHs) delay and access link transmission (from FBSs to users) delay. Here, we assume that the FBSC must wait for the arrival of the uncached requested files before transmitting all requested files to the users. Therefore, the transmission delay for user (l, k) can be calculated as

$$T_{l,k} = \frac{\sum_{f \in \mathcal{F}_{l,k}} b}{R_{l,k}^{\text{FBS}}} + \frac{\sum_{f \in \mathcal{F}_{l,k}} (1 - \sum_{m=1}^M a_{l,m}^f) b}{R_l^{\text{CP}}}, \quad (6)$$

where $\mathcal{F}_{l,k}$ denotes the files required by user (l, k) , and b is the size of each file. We assume that the state information $a_{l,m}^f$ and the required files $\mathcal{F}_{l,k}$ have been predetermined. Then, we aim to minimize the total transmission delay by optimizing the analog/digital precoding, which can be formulated as

$$\min_{\{\mathbf{F}_l, \{\mathbf{v}_{l,k}\}, \{\mathbf{w}_l\}\}} \sum_{l=1}^L \sum_{k=1}^K T_{l,k} \quad (7a)$$

$$\text{s.t.} \quad \sum_{k=1}^K \|\mathbf{v}_{l,k}(m)\|^2 \leq P_{l,m}^{\max}, m \in \mathcal{M}, l \in \mathcal{L}, \quad (7b)$$

$$\sum_{l=1}^L \|\mathbf{w}_l\|^2 \leq P_{\max}, \quad (7c)$$

where $\mathbf{v}_{l,k}(m)$ denotes m th element of $\mathbf{v}_{l,k}$, (7b) represents each FBS's power constraint, and (7c) is the BS's power constraint in the CP. Problem (7) is intractable to directly solve in term of its original form.

III. THE SOLUTION OF THE FORMULATED PROBLEM

One can observe that the precoding design of each FBSC and the BS in the PC is relatively independent, thus we can divide (7) into the following two optimization problems

$$\mathbb{P1} \quad \min_{\{\mathbf{F}_l, \{\mathbf{v}_{l,k}\}\}} \sum_{k=1}^K \frac{\sum_{f \in \mathcal{F}_{l,k}} b}{R_{l,k}^{\text{FBS}}}, \quad \text{s.t.} \quad (7b). \quad (8)$$

$$\mathbb{P2} \quad \min_{\{\mathbf{w}_l\}} \sum_{l=1}^L \sum_{k=1}^K \frac{\sum_{f \in \mathcal{F}_{l,k}} (1 - \sum_{m=1}^M a_{l,m}^f) b}{R_l^{\text{CP}}}, \quad \text{s.t.} \quad (7c). \quad (9)$$

We first solve $\mathbb{P1}$ for each FBSC. For the analog precoding \mathbf{F}_l in (8), only quantized phase can be applied in practice. Assume B bits quantized phase shifters, the non-zero entries of \mathbf{F}_l should belong to $\exp(j\frac{2\pi\hat{\alpha}}{2^B})/\sqrt{N}$ ($\alpha = 0, 1, \dots, 2^B - 1$). In addition, since $\mathbf{h}_{l,k} \mathbf{F}_l = [\mathbf{h}_{l,k}^1 \mathbf{f}_{l,1}, \dots, \mathbf{h}_{l,k}^M \mathbf{f}_{l,M}]$, we can design analog precoding to maximize the array gain $|\mathbf{h}_{l,k}^m \mathbf{f}_{l,m}|^2$, namely

$$\hat{\alpha} = \arg \min_{\alpha \in \{0, \dots, 2^B - 1\}} |\angle(\mathbf{h}_{l,k}^m(n)) - 2\pi\alpha/2^B|, n \in \mathcal{N}. \quad (10)$$

Then, the n -th element of the analog precoding vector $\mathbf{f}_{l,m}(n) = \exp(j\frac{2\pi\hat{\alpha}}{2^B})/\sqrt{N}$. Meanwhile, to guarantee the users' fairness, each user must be allocated at least one FBS to maximize the array gain, and we summarize the scheme in Algorithm 1.

Next, we introduce new variable $\tau_{l,k} \geq \sum_{f \in \mathcal{F}_{l,k}} b/R_{l,k}^{\text{FBS}}$, and (10) can be transformed into the following optimization problem by intermediate variable $\{\beta_{l,k}\}$

$$\min_{\{\mathbf{v}_{l,k}, \{\beta_{l,k}\}, \{\tau_{l,k}\}\}} \sum_{k=1}^K \tau_{l,k} \quad (11a)$$

$$\text{s.t.} \quad \log_2(1 + \beta_{l,k}) \geq \frac{\sum_{f \in \mathcal{F}_{l,k}} b}{\tau_{l,k}}, k \in \mathcal{K}, \quad (11b)$$

$$|\mathbf{h}_{l,k} \mathbf{F}_l \mathbf{v}_{l,k}|^2 \geq \beta_{l,k} (\sum_{i \neq k} |\mathbf{h}_{l,k} \mathbf{F}_l \mathbf{v}_{l,i}|^2 + \delta^2), k \in \mathcal{K}, \quad (11c)$$

$$(7b). \quad (11d)$$

However, (11) is still difficult to solve due to the non-convex constraint (11c). By bringing the new variables $\{\theta_{l,k}\}$, we have

$$\min_{\{\mathbf{v}_{l,k}, \{\beta_{l,k}\}, \{\theta_{l,k}\}, \{\tau_{l,k}\}\}} \sum_{k=1}^K \tau_{l,k} \quad (12a)$$

$$\text{s.t.} \quad |\mathbf{h}_{l,k} \mathbf{F}_l \mathbf{v}_{l,k}|^2 \geq \beta_{l,k} \theta_{l,k}, k \in \mathcal{K}, \quad (12b)$$

$$\theta_{l,k} \geq \sum_{i \neq k} |\mathbf{h}_{l,k} \mathbf{F}_l \mathbf{v}_{l,i}|^2 + \delta^2, k \in \mathcal{K}, \quad (12c)$$

$$(7b), (11b). \quad (12d)$$

Now, (12b) is still a non-convex constraint. Next, we assume that $\{\tilde{\mathbf{v}}_{l,k}\}$ is a feasible solution and then, define a small constant as $\mathbf{v}_{l,k} \triangleq \tilde{\mathbf{v}}_{l,k} + \Delta \mathbf{v}_{l,k}$, thus we have

$$\begin{aligned} |\mathbf{h}_{l,k} \mathbf{F}_l \mathbf{v}_{l,k}|^2 &= (\tilde{\mathbf{v}}_{l,k} + \Delta \mathbf{v}_{l,k})^H \mathbf{\Omega}_{l,k} (\tilde{\mathbf{v}}_{l,k} + \Delta \mathbf{v}_{l,k}) \\ &\geq 2\text{Re}\{\tilde{\mathbf{v}}_{l,k}^H \mathbf{\Omega}_{l,k} \Delta \mathbf{v}_{l,k}\} + \mathbf{v}_{l,k}^H \mathbf{\Omega}_{l,k} \tilde{\mathbf{v}}_{l,k}, \end{aligned} \quad (13)$$

where $\mathbf{\Omega}_{l,k} = \mathbf{h}_{l,k} \mathbf{F}_l (\mathbf{h}_{l,k} \mathbf{F}_l)^H$. In addition, the upper bound of $\beta_{l,k} \theta_{l,k}$ can be obtained as

$$\theta_{l,k}^{(i)} \beta_{l,k}^2 / 2\beta_{l,k}^{(i)} + \beta_{l,k}^{(i)} \theta_{l,k}^2 / 2\theta_{l,k}^{(i)} \geq \beta_{l,k} \theta_{l,k}, \quad (14)$$

where $\theta_{l,k}^{(i)}$ and $\beta_{l,k}^{(i)}$, respectively, are the value of $\theta_{l,k}$ and $\beta_{l,k}$ at the i th iteration. To this end, (12b) can be transformed into the following convex constraint

$$2\text{Re}\{\tilde{\mathbf{v}}_{l,k}^H \mathbf{\Omega}_{l,k} \Delta \mathbf{v}_{l,k}\} + \mathbf{v}_{l,k}^H \mathbf{\Omega}_{l,k} \tilde{\mathbf{v}}_{l,k} \geq \theta_{l,k}^{(i)} \beta_{l,k}^2 / 2\beta_{l,k}^{(i)} + \beta_{l,k}^{(i)} \theta_{l,k}^2 / 2\theta_{l,k}^{(i)}. \quad (15)$$

Finally, we only need to solve the following convex optimization problem

$$\min_{\{\mathbf{v}_{l,k}, \{\beta_{l,k}\}, \{\theta_{l,k}\}, \{\tau_{l,k}\}\}} \sum_{k=1}^K \tau_{l,k} \quad (16a)$$

$$\text{s.t.} \quad (7b), (11b), (12c), (15). \quad (16b)$$

The convex solvers (e.g., CVX) can be used to solve (16). Summarily, solving the original problem (12), we need to iteratively solve the optimal values of $\{\mathbf{v}_{l,k}\}, \{\beta_{l,k}\}, \{\theta_{l,k}\}, \{\tau_{l,k}\}$ via (17). In addition, since obtained $\{\mathbf{v}_{l,k}\}, \{\beta_{l,k}\}, \{\theta_{l,k}\}, \{\tau_{l,k}\}$ are the optimal solutions at each iteration by solving (17), iteratively updating these variables will decrease or maintain the value of the objective function (12). Due to the limited transmit power, the proposed iterative algorithm will converge to at least a local optimal solution. In addition, we can adopt the similar algorithm to solve $\mathbb{P2}$, and the process is omitted due to the limited space.

We summarize the process of the resource allocation and scheduling in the proposed FBSC-based F-RAN as shown in Fig. 2. One can observe that the CP does not need to optimize the global resource allocation, and thus the CP's overhead can be decreased. On the other hand, each FBSH only needs to

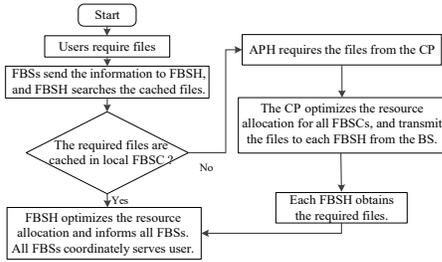


Fig. 2: The flow chart of the scheduling in FBSC-based F-RAN.

obtain the local CSI for optimizing and scheduling the local resource. Therefore, the proposed network structure is effective in not only decreasing the low transmission delay but also reducing the overhead of the system.

IV. SIMULATION RESULTS

In this section, simulation results are presented to evaluate the performance of our proposed algorithm. We consider $L = 3$ FBSCs and $M = 4$ FBSs in each FBSC. The BS is equipped with $S = 32$ antennas, and each FBS is equipped with $N = 16$ antennas. We assume there are $K = 4$ users in each FBSC. The mmWave channel model refers [18]. The noise power is set to -100 dBm. The maximum transmit power of each FBS is set as 40 dBm. For simplicity, the size of each file is normalized to unity, and each user requires one file. We assume that the cache time to live (TTL) of each content is enough long such that users can obtain this cached content before it is removed [8].

Fig. 3(a) shows the convergence performance of the proposed Algorithm 2. Here, we assume that the files required by users are cached in the local FBSCs. One can observe that the proposed iterative algorithm can achieve convergence after several iterations, which guarantees the effectiveness of Algorithm 2. In Fig. 3(b), we plot the transmission delay versus transmit power of the BS under different schemes. “Number of cached files” means the number of files cached in each FBSC. When the number of cached files is 4, it means that the required files by users are all cached in FBSC. Therefore, there is no fronthaul link delay, and only the access link delay exists. One can observe that the delay increases as number of cached files decreases. It is easy to understand that several files must be obtained from the CP, which causes the extra fronthaul link delay. In addition, we compare the conventional RAN structure (i.e., no cache) [3], and it is obvious that the delay is the largest. To perform our proposed algorithm, we also adopt the ZF beamforming scheme [10] to solve problem 2, we find that the delay is higher than that of our proposed algorithm.

V. CONCLUSIONS

In this paper, we investigated the transmission delay problem in FBSC-based F-RAN with mmWave communication. We designed an analog precoding scheme for each FBS. Then, we proposed an iterative digital precoding algorithm to minimize the system transmission delay. Simulation results shown that our proposed algorithm can speedily converge. Meanwhile, the transmission delay of our proposed algorithm is smaller

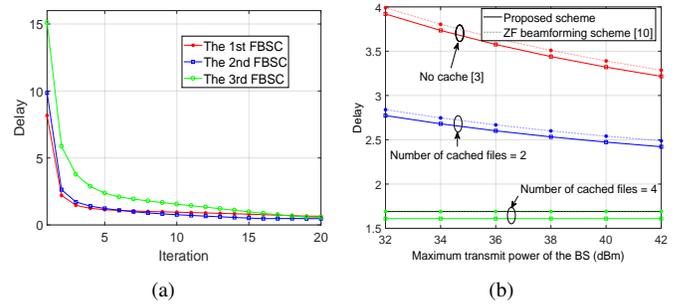


Fig. 3: (a) Delay versus number of iterations. (b) Delay versus maximum transmit power of the BS.

than that of the ZF beamforming scheme, which shows the effectiveness of the our proposed algorithm.

REFERENCES

- [1] J. G. Andrews, et al., “What will 5G be?” *IEEE J. Sel. Areas Commun.*, vol. 32, no. 6, pp. 1065-1082, Jun. 2014.
- [2] J. Liu, et al., “Cache placement in Fog-RANs: From centralized to distributed algorithms,” *IEEE Trans. Wireless Commun.*, vol. 16, no. 11, pp. 7039-7051, Nov. 2017.
- [3] J. Tang, et al., “System cost minimization in cloud RAN with limited fronthaul capacity,” *IEEE Trans. Wireless Commun.*, vol. 16, no. 5, pp. 3371-3384, May 2017.
- [4] X. Gao, et al., “Low RF-complexity technologies to enable millimeter-wave MIMO with large antenna array for 5G wireless communications,” *IEEE Commun. Mag.*, vol. 56, no. 4, pp. 211-217, Apr. 2018.
- [5] M. Peng, et al., “Recent advances in fog radio access networks: performance analysis and radio resource allocation,” *IEEE Access*, vol. 4, pp. 5003-5009, 2016.
- [6] S. He, et al., “Two-level transmission scheme for cache-enabled fog radio access Networks,” *IEEE Trans. Commun.*, vol. 67, no. 1, pp. 445-456, Jan. 2019.
- [7] G. M. S. Rahman, et al., “Radio resource allocation for achieving ultra-low latency in fog radio access networks,” *IEEE Access*, vol. 6, pp. 17442-17454, 2018.
- [8] M. Tao, et al., “Content-centric sparse multicast beamforming for cache-enabled cloud RAN,” *IEEE Trans. Wireless Commun.*, vol. 15, no. 9, pp. 6118-6131, Sep. 2016.
- [9] G. Zheng, et al., “Optimization of hybrid cache placement for collaborative relaying,” *IEEE Commun. Letters*, vol. 21, no. 2, pp. 442-445, Feb. 2017.
- [10] A. Sengupta, et al., “Fog-aided wireless networks for content delivery: Fundamental latency tradeoffs,” *IEEE Trans. Inf. Theory*, vol. 63, no. 10, pp. 6650-6678, Oct. 2017.
- [11] L. T. Tan, et al “D2D communications in heterogeneous networks with full-duplex relays and edge caching,” *IEEE Trans. Ind. Inform.*, vol. 14, no. 10, pp. 4557-4567, Oct. 2018.
- [12] W. Hao, et al., “Energy-efficient power allocation in millimeter wave massive MIMO with non-orthogonal multiple access,” *IEEE Wireless Commun. Lett.*, vol. 6, no. 6, pp. 782-785, Dec. 2017.
- [13] Y. Qiu, et al., “Energy-efficient power allocation with interference mitigation in mmWave-based fog radio access networks,” *IEEE Wireless Commun.*, vol. 25, no. 4, pp. 25-31, Aug. 2018.
- [14] C. Lin et al., “Energy-efficient design of indoor mmWave and sub-THz systems with antenna arrays,” *IEEE Trans. Wireless Commun.*, vol. 15, no. 7, pp. 4660-4672, Jul. 2016.
- [15] T. Tan, et al, “Mobility-aware edge caching and computing in vehicle networks: A Deep reinforcement learning,” *IEEE Trans. Veh. Technol.*, vol. 67, no. 11, pp. 10190-10203, Nov. 2018.
- [16] L. T. Tan, et al., “Twin-Timescale Artificial Intelligence Aided Mobility-Aware Edge Caching and Computing in Vehicular Networks,” *IEEE Trans. Veh. Technol.*, vol. 68, no. 4, pp. 3086-3099, Apr. 2019.
- [17] B. Dai, et al., “Optimized base-station cache allocation for cloud radio access network with multicast backhaul,” *IEEE J. Sel. Areas Commun.*, vol. 36, no. 8, pp. 1737-1750, Aug. 2018.
- [18] J. G. Andrews, et al., “Modeling and analyzing millimeter wave cellular systems,” *IEEE Trans. Wireless Commun.*, vol. 65, no. 1, pp. 403-430, Jan. 2017.