IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received October 20, 2017, accepted November 20, 2017, date of publication December 4, 2017, date of current version February 14, 2018.

*Digital Object Identifier 10.1109/ACCESS.2017.2779123*

# A Cache-Aided Communication Scheme for Downlink Coordinated Multipoint Transmission

**FASHENG ZHOU[1], (Member, IEEE), LISHENG FAN[2], NING WANG[3], (Member, IEEE), GAOYONG LUO[1], JIE TANG[4], (Member, IEEE), AND WEI CHEN[5], (Senior Member, IEEE)**

[1]School of Physics and Electronic Engineering, Guangzhou University , Guangzhou 510006, China
[2]School of Computer Science and Educational Software, Guangzhou University, Guangzhou 510006, China
[3]School of Information Engineering, Zhengzhou University, Zhengzhou 450001, China
[4]School of Electronic and Information Engineering, South China University of Technology, Guangzhou 510630, China
[5]Department of Electronic Engineering, Tsinghua University, Beijing 100084, China

Corresponding author: Fasheng Zhou (zhoufs@gzhu.edu.cn)

This work was supported in part by the Innovation Team Project of Guangdong Province University under Grant 2016KCXTD017, in part by the National Science Foundation of China under Grant 61401402 and Grant 61601186, in part by the Natural Science Foundation of Guangdong Province under Grant 1714050004216, in part by the Guangzhou Science Technology and Innovation Commission under Grant 201707010159, and in part by the Open Research Fund of National Mobile Communications Research Laboratory, Southeast University under Grant 2015D04.

**ABSTRACT** Content caching at the base stations (BSs) is a promising strategy that can provide improved quality-of-service (QoS) provisioning for content-centric services, and at the same time reduce instantaneous backhaul payload. This paper focuses on the applications of the caching mechanism in the downlink of coordinated multipoint (CoMP) transmission, where the collaborating BSs cooperatively serve the user equipments (UEs). A cache-aided transmission scheme for downlink CoMP with limited local cache resources at the BSs is proposed to improve QoS provisioning in terms of outage performance. In order to cache files more efficiently while guaranteeing the QoS, the contents are divided into two sets, a popular set and a less popular set, according to their popularity profile following the Zipf model. The popular files are cached at all the BSs, while the less popular files are cached only at part of the BSs in a cache cluster subject to the strategy design. Furthermore, based on the popularity of the requested files, UEs are categorized into two groups and the notion of request competition between the two groups is introduced. The corresponding request competition model and admission probabilities for transmission requests are analyzed. Based on the request competition model, the best cache placement strategy for the popular and less popular sets that minimizes the average outage probability is derived. Numerical results show that the proposed cache-aided CoMP scheme outperforms the baseline traditional caching strategies in terms of the average transmission outage.

**INDEX TERMS** Cache-aided communication, CoMP, content popularity, edge caching, multiuser request competition.

## I. INTRODUCTION

Over the past five years, the global mobile IP traffic has increased more than fivefold, and there is strong evidence that the growth is not going to slow down in the upcoming 5G era [1] considering the huge traffic demand from more and more diverse service types such as interactive multimedia, Internet of Everything (IoE), and cloud-based services. Coordinated multipoint (CoMP) is a promising technology that has been recognized as an effective approach to mitigate intercell interference (ICI) and as such better accommodate the ever growing need for data-traffic by the interference limited 5G heterogeneous networks (HetNets) [2]. The concept of CoMP in general focuses on multi-cell signal processing for ICI mitigation, which include, but not limited to, distributed beamforming in the downlink and base station cooperative receiving in the uplink. With the paradigm shift of wireless communications from connection-centric, e.g., voice telephony, to content-centric such as video-on-demand and push media [3], more and more traffic is delivered in the downlink through multicasting instead of unicasting to allow simultaneous transmission of the same requested content to multiple user equipments (UEs) [4]. As a consequence, user


2169-3536 © 2017 IEEE. Translations and content mining are permitted for academic research only.
Personal use is also permitted, but republication/redistribution requires IEEE permission.
See http://www.ieee.org/publications_standards/publications/rights/index.html for more information.

association will be fundamentally different from the connection-centric scenario, which in turn also greatly impacts the CoMP design. More sophisticated technologies must be exploited in downlink CoMP to address both the need for wireless networking and the content-centric nature of new services.

One promising solution is to combine cooperative processing with edge caching, which caches the content files at local base stations (BSs) for future transmission purposes. By doing so, better interference management can be achieved with lower cost, which helps in improving multicast throughput [4]. Specifically, it is widely accepted that capacity and delay are the key restrictions in the backhaul that prevent delivery of all the benefits of CoMP. Existing work in the literature [5]–[7] has been focusing on reducing the backhaul payload to alleviate the backhaul crunch, which is critical to densely deployed 5G small cell networks (SCNs) [8]. If the files to be transmitted are cached at the edge BSs in advance, the demand for backhaul resources by CoMP can be significantly reduced [9]. However, because the BS local caches are in general limited, and it is resource consuming and unrealistic to cache all the files at all the BSs, cache placement strategy must be carefully designed to balance between multicast throughput improvement and resource consumption. Note that popular contents are more likely to be requested by different UEs over a relatively large time span, it is reasonable to cache more popular files when available cache resource is limited. Making use of content popularity information in the design of caching strategy would therefore lead to good overall system performance [3], [10].

Due to the bursty nature of content-centric data traffic, caching can be achieved through file pre-fetching during off-peak periods [11], possibly in conjunction with prediction-based content pushing [12], and consequently reduce the instantaneous backhaul payload during downlink transmissions. A joint precoding and cache control problem was investigated by mixed-timescale joint stochastic optimization for a multi-cell multi-user networks [13]. The optimal control problem was decomposed into a long-term and a short-term sub-problems. The control variables of the caching strategy were designated as long-term variables, which were determined to adapt to popularity of media files such that the opportunity of using CoMP to achieve ICI mitigation is increased. The short-term variables for multiple-input-multiple-output (MIMO) precoding, on the other hand, were determined to guarantee the quality-of-service (QoS) requirements given the long-term cache control. Tao *et al.* [14] further considered a cloud radio access network (C-RAN) architecture under backhaul constraints, and investigated caching and user association mechanism that took both the channel station information (CSI) and the cache status into consideration. By minimizing the weighted sum of backhaul cost and transmit power in a cache-enabled C-RAN, subject to constraints on the user received signal quality, a user-centric cell clustering scheme was proposed, in which any BS that caches a user's requested files is always included in this user's serving cluster. However, it would be cumbersome to extend the framework to cache-enabled CoMP transmissions due to complexity issues.

Recently, Chen *et al.* studied a hybrid caching mechanism for cluster-centric SCNs where the small cells were grouped into disjoint clusters [15]. The contents to be cached were first classified into two groups according to their popularity. A dedicated cache space was reserved at every small-cell base station (SBS) to cache the most popular contents. The less popular contents, on the other hand, were partitioned and placed at different SBSs strategically using the remaining limited cache space. Each SBS then caches all the popular contents and part of the less popular contents. Through cooperation, CoMP transmission in the form of either joint transmission or parallel transmission can be achieved according to placement of the requested content. For edge caching, the most popular content (MPC) strategy and the largest content diversity (LCD) strategy are the most commonly used strategies to achieve the content and network diversity gains [16], [17]. Nevertheless, caching with the MPC or LCD criterion is not always the best choice, especially when overall QoS provisioning comes into play. Targeting at attaining the optimal outage performance, the outage minimization problem for hybrid cache placement in a cooperative multi-relay system was studied, and a hybrid caching scheme was proposed [18]. A trade-off between signal cooperation gain and content diversity gain was investigated through threshold-based cache placement.

Different from the above mentioned works, in this paper we consider downlink CoMP transmission with UE request competition, and propose a cache-aided scheme to improve QoS provisioning of the system in terms of outage probability. As in [15], the content files are divided into two categories (popular versus less popular), according to their popularity characteristics. A caching strategy based on the best placement of the popular and less popular files is proposed. Popular files are cached at all BSs, while less popular files are cached at part of BSs in the proposed scheme. The best cache placement for CoMP would attain the minimum average outage probability performance. The main contributions of this paper are summarized in the following.

- We propose a cache-aided CoMP transmission scheme, which works well in the system having two coordination modes, two file sets, and two UE groups.
- The notion of *request competition* is introduced. Based on request scheduling, we respectively derive the request successful probabilities for the two UE groups. The relationship between the request successful probabilities and the average transmission outage is then established.
- An optimization problem is formulated to minimize the average transmission outage through optimal cache placement. Numerical results show that the proposed cache placement scheme outperforms the traditional MPC and LCD caching strategies in terms of the average outage for the cache-aided CoMP system.

The remainder of this paper is organized as follows. In Section II, we illustrate our system and cache models for the cache-aided CoMP system. Outage performance of the cache-aided CoMP transmission is analyzed Section III, followed by the optimization problem formulation and the corresponding optimal cache placement scheme in Section IV. Numerical results are provided in Section V to validate the proposed scheme. Finally, some concluding remarks are made in Section VI.

*Notation:* The notation $\mathbb{E}[\cdot]$ denotes statistical expectation. $Z \sim \mathcal{CN}(0, a)$ is a circular symmetric Gaussian random variable (RV) with zero mean and variance $a$. We use $||\cdot||$ for the 2-norm operation. $\mathbb{C}$, $\mathbb{Z}_+$, and $\mathbb{R}_+$ represent the set of complex numbers, the set of nonnegative integers, and the set of nonnegative real numbers, respectively. $\Pr\{\cdot\}$ denotes the probability of the event specified in the argument.

## II. SYSTEM MODEL

### A. WIRELESS SYSTEM MODEL FOR DOWNLINK CoMP

In this work we consider a CoMP system that comprises $M$ BSs and $N$ UEs. Backhaul links for BS inter-connection are available such that the BSs can exchange their CSI and transmitted data. Each BS is equipped with $N_t$ antennas, while each UE has only one antenna due to the size limitation. The transmit power of each BS is $P_b$. For convenience, we denote such a system by CoMP$(M, N)$.

Consider quasi-static wireless channels where the channel coefficients are constant over one symbol period. Circularly symmetric Gaussian distribution with zero mean and unit variance is assumed for the channel coefficients. Let BS$_j$ denote the $j$th BS for $1 \le j \le M$, UE$_i$ denote the $i$th UE for $1 \le i \le N$, $\mathcal{S}_B$ denote the set of the $M$ BSs, and $\mathcal{U}_j$ denote the set of UEs served by BS$_j$. The transmitted signal of UE$_i$ is denoted by $x_i \in \mathbb{C}$, and we further assume that $\mathbb{E}\{|x_i|^2\} = 1$. The large-scale fading channel gain between BS$_j$ and UE$_i$ is $\mu_{ji} \in \mathbb{R}$, and the corresponding small-scale fading channel vector is $\mathbf{h}_{ji} = [h_{ji}^1 \ldots h_{ji}^{N_t}]^T$, where the $h_{ji}^k \sim \mathcal{CN}(0, \sigma_1^2)$ for $1 \le k \le N_t$ is the complex channel gain from the $k$th antenna of BS$_j$ to UE$_i$.

### B. CONTENT AND CACHE MODEL

Let $\mathcal{K} = \{1, 2, \ldots, K\}$ denote the set of $K$ files to be requested by the UEs. The contents' popularity follows a Zipf distribution [19]. the popularity profile of the $c$-th file in the database is characterized by $f_c$, which is defined as

$$f_c = \frac{c^{-\alpha}}{\sum\limits_{k=1}^{C} k^{-\alpha}}, \tag{1}$$

where $0 < \alpha < 1$ is the parameter that characterizes the Zipf distribution. The larger $\alpha$ is, the more concentrated the request distribution is around the most popular file(s). Note that $f_c$ can also be interpreted as the request possibility for the $c$-th file.

Assume each BS has a local storage (cache) capacity that can cache up to $C$ files, where we require $C \le K$ to reflect the limitation in cache resources. As in [15], the content files are divided into two sets according to content popularity. The first set, which is named the *popular* set, comprises the most popular $F$ files. A notation $\mathcal{F}$ is used to denote the *popular* set, and we must have $|\mathcal{F}| = F$. The second set comprises the remaining less popular $K - F$ files. The term *less popular* has a relative meaning, relative to *popular*. In other words, the $K - F$ least popular files are determined according to the cache placement policy for the *less popular* set. $F$ is an adjustable parameter and it can be interpreted as a threshold for selecting the most popular files to be cached at all BSs. Increasing $F$ would result in decrease of the remaining less popular files, and vice versa.

The popular set is cached at all the local BSs because they are most likely to be requested by multiple UEs at different locations. For files in the less popular set, we do not cache them at all BSs due to limitations in cache and backhaul resources. Instead, each of them will be cached at only part of the BSs. Thereby, we divide the BSs into $L$ clusters, where $L = \left\lceil \frac{K-F}{C-F} \right\rceil$, and $\lceil\cdot\rceil$ is the ceiling operator that returns the smallest integer greater than its argument. We use a notation $\mathcal{S}_l^c$ to denote the set of BSs in the $l$-th cluster CL$_l$ for $1 \le l \le L$. Files cached in the $l$-th cluster form the set $\{\mathcal{K} - \mathcal{F}\}_l$. The less popular files should be completely covered by the $L$ less popular content subsets, which requires

$$\mathcal{K} = \mathcal{F} \cup \bigcup_{l=1}^{L} \{\mathcal{K} - \mathcal{F}\}_l. \tag{2}$$

Each user requests a content at the beginning of the transmission time slot. The UEs are also divided into two groups according to popularity of their requested contents. The first group's requests fall into the popular set $\mathcal{F}$. The second UE group requests contents from the less popular set and is further divided into $L$ clusters (subgroups). Those UEs in the $l$-th cluster of the second group ($1 \le l \le L$) request the less popular files belonging to $\{\mathcal{K} - \mathcal{F}\}_l$ and are served by the cluster $\mathcal{S}_l^c$ of BSs. Note that if $l \neq k$, $\{\mathcal{K} - \mathcal{F}\}_l$ does not overlap with $\{\mathcal{K} - \mathcal{F}\}_k$. However, the overall popularity profiles of $\{\mathcal{K} - \mathcal{F}\}_l$ and $\{\mathcal{K} - \mathcal{F}\}_k$ should be made (almost) equal for $l \neq k$, i.e.,

$$\sum_{c \in \{\mathcal{K}-\mathcal{F}\}_l} f_c = \sum_{c \in \{\mathcal{K}-\mathcal{F}\}_k} f_c, \quad \forall\, l, k \in [1, L],\ l \neq k. \tag{3}$$

This BS clustering and cache placement strategy is illustrated by Fig. 1, where the different areas formed by different cells, *e.g.*, Cell 1, Cell 2... as referred to C$_1$, C$_2$... request the same popular contents, but quite distinct less popular contents. This follows from the fact that UEs from different areas (cell clusters) may request the same popular contents, but quite distinct less popular contents.

## III. CACHE-AIDED CoMP TRANSMISSION

### A. CACHE-AIDED CoMP TRANSMISSION

A typical CoMP scheme requires multiple BSs to serve UE(s) in a collaborative fashion. Intra-cell interference and inter-cell interference, or collectively multiuser interference (MUI), requires much attention in CoMP design.

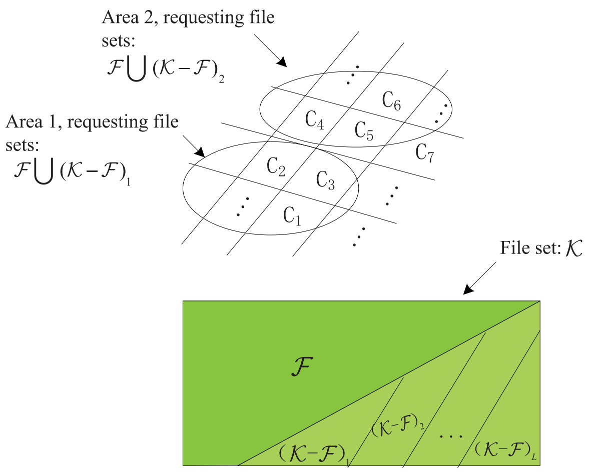



**FIGURE 1.** **Illustration of the BS clustering and cache placement based on grouping of requested files.**

To address the interference problem, coordinated zero-forcing beamforming (ZFBF) can be employed [20], [21]. Denote by $\mathbf{h}_{ji} \sim \mathcal{CN}(\mathbf{0}, \sigma_1^2\mathbf{I})$ the complex small-scale channel coefficient vector from the $j$-th BS to UE$_i$, with $\mathbf{I}$ an identical matrix, and $n_i \sim \mathcal{CN}(0, \sigma_2^2)$ the additive white Gaussian noise (AWGN) at UE$_i$. For transmission from BS$_j$ to UE$_i$, the ZFBF beamformer $\mathbf{w}_{ji}$ is constructed as [7]

$$\mathbf{w}_{ji} = \frac{\sqrt{P_b}\left[\mathbf{I} - \hat{\mathbf{H}}_{ji}^H(\hat{\mathbf{H}}_{ji}\hat{\mathbf{H}}_{ji}^H)^{-1}\hat{\mathbf{H}}_{ji}\right]\hat{\mathbf{h}}_{ji}}{\left\|\left(\mathbf{I} - \hat{\mathbf{H}}_{ji}^H(\hat{\mathbf{H}}_{ji}\hat{\mathbf{H}}_{ji}^H)^{-1}\hat{\mathbf{H}}_{ji}\right)\hat{\mathbf{h}}_{ji}\right\|}, \tag{4}$$

where $\hat{\mathbf{h}}_{ji} = [\hat{h}_{ji}^1 \ldots \hat{h}_{ji}^{N_t}]$ denotes the unbiased estimate of $\mathbf{h}_{ji}$, and $\hat{\mathbf{H}}_{ji} = [\hat{\mathbf{h}}_{j1} \ldots \hat{\mathbf{h}}_{j(i-1)} \ \hat{\mathbf{h}}_{j(i+1)} \ldots \hat{\mathbf{h}}_{jN}]$. Then $\left[\mathbf{I} - \hat{\mathbf{H}}_{ji}^H(\hat{\mathbf{H}}_{ji}\hat{\mathbf{H}}_{ji}^H)^{-1}\hat{\mathbf{H}}_{ji}\right]$ is the null space spanned by the small-scale channel vectors of all the wireless channels except for that of UE$_i$.

Suppose UE$_i$ is served by all the BSs, which is the case when it requests files from the *popular* set $\mathcal{F}$, then the received signal $y_i$ at UE$_i$ can be expressed as

$$y_i = \underbrace{\sum_{j\in\mathcal{S}_B} \mu_{ji}\mathbf{h}_{ji}^H\mathbf{w}_{ji}x_i}_{\text{Desired signals}} + \underbrace{\sum_{j\in\mathcal{S}_B}\sum_{k\in\mathcal{U}_j,k\neq i} \mu_{ji}\mathbf{h}_{ji}^H\mathbf{w}_{jk}x_k}_{\text{Multiuser interference}} + n_i, \tag{5}$$

where $\mu_{ji}$ represents the large-scale channel coefficient between the $j$-th BS and the $i$-th UE. The first and second terms on the right-hand side of (5) represent the desired signal and the MUI, respectively. The received signal-to-interference-plus-noise ratio (SINR) at UE$_i$ is then given as

$$\text{SINR}_i = \frac{\sum_{j\in\mathcal{S}_B} {\mu_{ji}}^2||\mathbf{h}_{ji}^H\mathbf{w}_{ji}||^2}{\sum_{j\in\mathcal{S}_B}\sum_{k\in\mathcal{U}_j,k\neq i} {\mu_{ji}}^2||\mathbf{h}_{ji}^H\mathbf{w}_{jk}||^2 + \sigma_2^2}. \tag{6}$$

If the transmit power of every cooperating BS is finite and the BSs are sufficiently separated apart in the space, the residual MUI after ZFBF in eqn. (5) is negligible. Consequently, the interference term $\sum_{j\in\mathcal{S}_B,k\neq i} {\mu_{ji}}^2||\mathbf{h}_{ji}^H\mathbf{w}_{jk}||^2$ in eqn. (6) approaches zero [7].

According to cache placement of the less popular files, the BSs are divided into $L$ clusters, as illustrated by Fig. 1 in Section II-B. A UE, say UE$_i$, requesting less popular files from $\{\mathcal{K}-\mathcal{F}\}_l$ therefore is served by the $l$-th cluster of BSs. The corresponding received signal $y_i$ is given by

$$y_i = \underbrace{\sum_{j\in\mathcal{S}_l^c} \mu_{ji}\mathbf{h}_{ji}^H\mathbf{w}_{ji}x_i}_{\text{Desired signal}} + \underbrace{\sum_{j\in\mathcal{S}_l^c}\sum_{k\in\mathcal{U}_j,k\neq i} \mu_{ji}\mathbf{h}_{ji}^H\mathbf{w}_{jk}x_k}_{\text{Intra-cell interference}} + \underbrace{\sum_{j\in\mathcal{S}_n^c,n\neq l}\sum_{k\in\mathcal{U}_j} \mu_{ji}\mathbf{h}_{ji}^H\mathbf{w}_{jk}x_k}_{\text{Inter-cell interference}} + n_i. \tag{7}$$

The SINR corresponding to eqn. (7) is given by eqn. (8) on the bottom of the next page. Similarly, if the residual intra- and inter-cell interference after ZFBF in eqn. (7) is negligible, the interference terms $\sum_{j\in\mathcal{S}_l^c}\sum_{k\in\mathcal{U}_j,k\neq i} {\mu_{ji}}^2||\mathbf{h}_{ji}^H\mathbf{w}_{jk}||^2$ and $\sum_{j\in\mathcal{S}_n^c,n\neq l}\sum_{k\in\mathcal{U}_j} {\mu_{ji}}^2||\mathbf{h}_{ji}^H\mathbf{w}_{jk}||^2$ in eqn. (8) are nulled.

### B. OUTAGE PERFORMANCE ANALYSES

For CoMP with multiple UEs, the optimal scheduling to attain the minimum information-theoretic outage probability is to admit the request from the UE with the best channel condition [22]. Then, the requested file is transmitted to the corresponding admitted UE. In this work, we are interested in the following two scenarios, corresponding to whether or not the requested files are in the popular set $\mathcal{F}$.

*Scenario 1:* When files in the popular set $\mathcal{F}$ are requested, since in this case the files are cached at all BSs, all the BSs can cooperatively transmit to the requesting UE.

*Scenario 2:* When files in the less popular set $\{\mathcal{K}-\mathcal{F}\}$ are requested, those BSs (not all) in the cluster that caches the corresponding files transmit cooperatively to the requesting UE.

Consider requests for different files in the same set can be accommodated through broadcasting the data stream by the cooperating BSs that have the files cached. Therefore, UEs requesting files from the same set do not conflict. However, if the UE requested files are in different sets, accommodating those requests require different CoMP transmission modes, i.e., full CoMP transmission of all BSs versus partial CoMP transmission of a BS cluster. We define such situation as *request competition*. Note that the transmission model for Scenario 1 and Scenario 2 are expressed by eqns. (5) and (7), respectively. Observe that SINR$_i$ given by eqn. (6) constrained by eqn. (5) is obviously greater than SINR$_{i,l}$ given by eqn. (8) constrained by eqn. (7), the received signal quality for Scenario 1 is in general better than that in Scenario 2. Therefore, in order to achieve the minimum information-theoretic outage, the best transmission channel is chosen among multi-requirement for transmission of contents. That is to say, once the file in the popular set is requested by a UE, the other UEs have to wait until the transmission for the file finishes. When a file in the less popular set is requested, the request is admitted and the requested file is sent only if currently there is no request for the popular files.

In our analysis, the UEs randomly request the files based on file popularity. Once requested, a popular files cached at all the BSs can be jointly transmitted by all the BSs, while a less popular file can only be transmitted by a subset (cluster) of the BSs. The outage performance associated with a content file is then dependent on the popularity profile. The average outage of the system is therefore the sum of the outage values of each file weighted by the popularity profile. Suppose that $\mathrm{UE}_i$ belongs to the first UE group and it requests files in the first set. All the BSs serve $\mathrm{UE}_i$ upon its requests, and the transmission outage probability, denoted by $P_1^{out}$, is given by

$$P_1^{out} = P_e^{FC}(e_o)\sum_{c=1}^{F} f_c \tag{9}$$

where $P_e^{FC}(e_o)$ is the outage probability regarding a bit error rate (BER) requirement $e_o$ for full coordination of all BSs, and $\sum_{c=1}^{F} f_c$ is a weighting coefficient corresponding to the popularity profile of the first (popular) file set. Based on eqn. (6), $P_e(e_o)$ can be well approximated by

$$\begin{aligned} P_e^{FC}(e_o) &\triangleq \Pr\left\{\mathbf{Q}\left(\mathrm{c}_1\sqrt{\mathrm{c}_2\frac{E_b}{N_0}}\right) > e_o\right\} \\ &= \Pr\left\{\mathbf{Q}\left(\mathrm{c}_1\sqrt{\mathrm{c}_2\frac{W}{R}\mathrm{SINR}_i}\right) > e_o\right\}, \end{aligned} \tag{10}$$

where

$$\mathbf{Q}(x) = \int_x^{\infty} \frac{1}{\sqrt{2\pi}}\exp(-\frac{y^2}{2})dy \tag{11}$$

is the Gaussian Q-function, $\mathrm{c}_1$ and $\mathrm{c}_2$ are constants depending on the modulation and coding scheme [23], $E_b$, $N_0$, $W$, and $R$ are the bit energy, noise spectral density, bandwidth, and code rate, respectively. By using approximation of the Q-function [24, eq.(10)] can be transformed into a more explicit and simpler form

$$\begin{aligned} P_e^{FC}(e_o) \approx \Pr\Bigg\{&\frac{1}{12}\exp\left(-\frac{\mathrm{c}_1^2\mathrm{c}_2}{2}\frac{W}{R}\mathrm{SINR}_i\right) \\ &+ \frac{1}{4}\exp\left(-\frac{2\mathrm{c}_1^2\mathrm{c}_2}{3}\frac{W}{R}\mathrm{SINR}_i\right) > e_o\Bigg\}. \end{aligned} \tag{12}$$

We next derive outage performance for the second file set, i.e., the less popular set. Suppose $\mathrm{UE}_i$ belongs to the $l$-th cluster in the second file set. Note that the less popular files $\{\mathcal{K}-\mathcal{F}\}_l$ are cached at BSs in $l$-th cluster, $\mathcal{S}_l^c$. When $\mathrm{UE}_i$ requests for files in $\{\mathcal{K}-\mathcal{F}\}_l$, the BSs in the $l$-th cluster jointly transmit the files to $\mathrm{UE}_i$. The corresponding outage probability, denoted by $P_{2,l}^{out}$, is given by

$$P_{2,l}^{out} = P_{e,l}^{PC}(e_o)\sum\nolimits_{c=F+1}^{C} f_c \tag{13}$$

where $P_{e,l}^{PC}(e_o)$ denotes the BER-based outage probability at a predetermined BER threshold $e_o$ with partial coordination implemented by the BSs in $\mathcal{S}_l^c$, and $\sum_{c=F+1}^{C} f_c$ is a weighting term corresponding to the popularity of the less popular files. Similarly, based on eqn. (8), $P_{e,l}^{PC}(e_o)$ can be approximated by

$$\begin{aligned} P_{e,l}^{PC}(e_o) \approx \Pr\Bigg\{&\frac{1}{12}\exp\left(-\frac{\mathrm{c}_1^2\mathrm{c}_2}{2}\frac{W}{R}\mathrm{SINR}_{i,l}\right) \\ &+ \frac{1}{4}\exp\left(-\frac{2\mathrm{c}_1^2\mathrm{c}_2}{3}\frac{W}{R}\mathrm{SINR}_{i,l}\right) > e_o\Bigg\}. \end{aligned} \tag{14}$$

Let $\bar{P}_e^{PC}(e_o)$, and $\bar{P}_2^{out}$ denote the average outage probability of transmitting less popular files with partial CoMP transmission by BSs in one cluster, and the average of $P_{2,l}^{out}$ for $1 \le l \le L$, respectively. Then, $\bar{P}_e^{PC}(e_o)$ and $\bar{P}_2^{out}$ can be expressed by

$$\bar{P}_e^{PC}(e_o) = \frac{1}{L}\sum_{l=1}^{L} P_{e,l}^{PC}(e_o) \tag{15}$$

and

$$\bar{P}_2^{out} = \bar{P}_e^{PC}(e_o)\sum\nolimits_{c=F+1}^{C} f_c, \tag{16}$$

respectively.

## IV. OPTIMAL CACHE PLACEMENT FOR CoMP

UEs may request files from different sets. As aforementioned, choosing the UE with the best channel condition for transmission can achieve minimum average outage probability. Therefore, the requests for popular files are assigned higher priorities than that for the less popular files. This means when competition between content file requests occurs, the popular files shall be transmitted. In other words, if a UE requests the popular files, its request is always admitted. The probability of admitting requests for the popular files is always equal to one, no matter competition occurs or not. However, if a UE requests the less popular files, its request is only admitted if no competition with popular file requests occur. We have the following Lemma 1 for the admission probability of requests for the less popular files.

$$\mathrm{SINR}_{i,l} = \frac{\sum\limits_{j\in\mathcal{S}_l^c} {\mu_{ji}}^2||\mathbf{h}_{ji}^H\mathbf{w}_{ji}||^2}{\sum\limits_{j\in\mathcal{S}_l^c}\sum\limits_{k\in\mathcal{U}_j,k\neq i} {\mu_{ji}}^2||\mathbf{h}_{ji}^H\mathbf{w}_{jk}||^2 + \sum\limits_{j\in\mathcal{S}_n^c,n\neq l}\sum\limits_{k\in\mathcal{U}_j} {\mu_{ji}}^2||\mathbf{h}_{ji}^H\mathbf{w}_{jk}||^2 + \sigma_2^2}. \tag{8}$$

*Lemma 1:* The admission probability of the request for the less popular files in a unit time interval is

$$P_s = e^{-\sum_{c=1}^{F} f_c}. \tag{17}$$

*Proof:* See Appendix A. ■

We next formulate the following optimization problem for outage optimal cache placement.

*Problem 1:* Denote the average outage probability of the downlink CoMP transmission system by $\bar{P}_{out}$. Considering the request admission probability of the less popular files $P_s$ given in Lemma 1, we formulate the problem for optimal cache placement of $F$ files that minimizes $\bar{P}_{out}$.

$$\min_{F} \bar{P}_{out} = P_1^{out} + P_s \bar{P}_2^{out}$$
$$\text{s.t. } 0 \leq F \leq K;$$
$$F \in \mathbb{Z}_+. \tag{18}$$

By approximating the sum of the Zipf probabilities as in [18], we obtain

$$\sum_{c=1}^{F} f_c \approx \frac{F^{(1-\alpha)} - 1}{K^{(1-\alpha)} - 1}. \tag{19}$$

Substituting (19) into (17) and (18), the optimization objective $\bar{P}_{out}$ of Problem 1 is rewritten as

$$\bar{P}_{out} = P_e^{FC}(e_o) \cdot \frac{F^{(1-\alpha)} - 1}{K^{(1-\alpha)} - 1} + \mathrm{e}^{-\frac{F^{(1-\alpha)}-1}{K^{(1-\alpha)}-1}} \bar{P}_e^{PC}(e_o) \cdot \left[1 - \frac{F^{(1-\alpha)} - 1}{K^{(1-\alpha)} - 1}\right], \tag{20}$$

Note that the feasible set of Problem 1 is not convex because $F$ is an integer. The problem can be relaxed by replacing $F$ with a continuous variable $F'$ on a convex set $[0, K] \subset \mathbb{R}_+$. The following Lemma 2 and Theorem 1 are obtained based on the above relaxation.

*Lemma 2:* The objective function of Problem 1 is a convex function of the relaxed continuous variable $F'$.

*Proof:* The second order derivative of $\bar{P}_{out}(F')$ with respect to the continuous variable $F'$ is obtained as

$$\frac{\partial^2 \bar{P}_{out}(F')}{\partial F'^2} = \left[2\bar{P}_e^{PC}(e_o) \cdot \mathrm{e}^{-\frac{F'^{(1-\alpha)}-1}{K^{(1-\alpha)}-1}} + \bar{P}_e^{PC}(e_o) \cdot \mathrm{e}^{-\frac{F'^{(1-\alpha)}-1}{K^{(1-\alpha)}-1}} \cdot \left(1 - \frac{F'^{(1-\alpha)} - 1}{K^{(1-\alpha)} - 1}\right)\right] \times \frac{(1-\alpha)}{F'^{\alpha}(K^{(1-\alpha)} - 1)}. \tag{21}$$

Because $0 \leq F' \leq K$ with $K > 1$, and $0 < \alpha < 1$, it is straightforward to show that the above second order derivative is strictly greater than 0. As a consequence, the optimization objective function $\bar{P}_{out}(F')$ is a convex function of the relaxed variable $F'$ over a convex set. ■

*Theorem 1:* The optimal value of $F$, denoted by $F^*$, for the cache placement in the downlink of the cache-aided CoMP system is given by

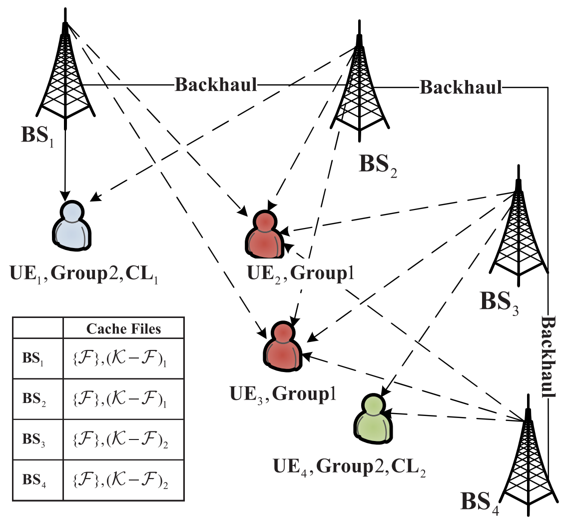
$$F^* = \begin{cases} \lceil (F')^* \rceil, & (F')^* \in (0, K), \\ K, & (F')^* > K, \end{cases} \tag{22}$$

where $(F')^*$ is the optimal solution to the relaxed continuous problem given by eqn. (23) on the bottom of this page. The function form Lambertw$(y)$ is the Lambert-W function [26] that returns the solution $x$ to the equation $xe^x = y$.

*Proof:* See Appendix B. ■

We can therefore achieve the minimum average outage probability if we set $F = F^*$ as specified by eqn. (22) in Theorem 1.

Up to this point, we have proposed a cache-aided downlink CoMP transmission scheme and the corresponding optimal cache placement strategy. In the following we summarize and briefly illustrate the proposed scheme in a 4-BS 4-UE scenario as shown in Fig. 2.




| | Cache Files |
|---|---|
| BS1 | $\{\mathcal{F}\}, (\mathcal{K}-\mathcal{F})_1$ |
| BS2 | $\{\mathcal{F}\}, (\mathcal{K}-\mathcal{F})_1$ |
| BS3 | $\{\mathcal{F}\}, (\mathcal{K}-\mathcal{F})_2$ |
| BS4 | $\{\mathcal{F}\}, (\mathcal{K}-\mathcal{F})_2$ |

**FIGURE 2. A demonstration of the caching model with $M = 4$ and $N = 4$. The UEs are divided to two groups, where Group1 has UE2 and UE3, and Group2 comprises UE1 and UE4 each served by a cluster of BSs.**

The four UEs are divided into two groups, namely Group1 and Group2, corresponding to popularity of their requested contents. As analyzed above, requests raised by the UEs in Group1 are given higher priority than that raised by

$$(F')^* = \left[\left(K^{(1-\alpha)} - 1\right)\left(2 - \text{Lambertw}\left(\exp(2)\left(\frac{P_e^{FC}(e_o)}{\bar{P}_e^{PC}(e_o)}\right)\right)\right) + 1\right]^{\frac{1}{(1-\alpha)}}. \tag{23}$$

Group2 UEs. The proposed scheme schedules the requests so that the admission probability for Group1 requests (for files in the popular set) is equal to 1, while that for Group2 requests is given by eqn. (17). The threshold parameter $F$ for caching the most popular files is set to $F^*$ given in eqn. (23). The most popular $F$ files are cached at all the four BSs. The less popular files of $\{\mathcal{K}-\mathcal{F}\}_1$ and $\{\mathcal{K}-\mathcal{F}\}_2$ are cached at the two BS clusters, $\mathcal{S}_1^c$ (or $CL_1$) and $\mathcal{S}_2^c$ (or $CL_2$), respectively. In this example, $\mathcal{S}_1^c$ consists of $BS_1$ and $BS_2$, and $\mathcal{S}_2^c$ consists of $BS_3$ and $BS_4$. Assume $UE_2$ requests files in the popular set, and therefore it belongs to Group1. Once the request is admitted, all four BSs cooperatively transmit to $UE_2$ the requested file using their own cached data and global CSI. In comparison, $UE_1$ and $UE_4$ request the files in $\{\mathcal{K}-\mathcal{F}\}_1$, and $\{\mathcal{K}-\mathcal{F}\}_2$, respectively. Hence, they respectively belong to $\mathcal{S}_1^c$ and $\mathcal{S}_2^c$. When $UE_1$'s request is admitted, $BS_1$ and $BS_2$ cooperatively serve $UE_1$ with cached data. Similarly, when $UE_4$'s request is admitted, $BS_3$ and $BS_4$ conduct cache-aided cooperative transmission to $UE_4$. Backhaul links connecting BSs are available such that collaborative ZFBF can be achieved by sharing the CSI. However, because of the file caching mechanism employed, there is no need to conduct instantaneous file data exchange through backhaul for CoMP transmission.

## V. NUMERICAL RESULTS AND DISCUSSIONS

In this section, numerical studies are conducted to evaluate effectiveness of the proposed cache-aided CoMP transmission scheme. We consider two baselines schemes in the numerical analysis for comparison: the MPC strategy in [16] and the LCD strategy in [17]. We also present performance for the scenario with no BS cooperation or caching. The requesting UE is therefore served by only one BS.

In the numerical studies, we consider a scenario with four small cell BSs and four UEs. The topology of the example network is demonstrated in Fig. 2. The UEs are grouped into two clusters according to the proposed cache placement strategy, as explained in the last paragraph of Section IV. By referring to [27], the maximum transmit power of each small cell BS is set to 30 dBm. The following large-scale pathloss model is adopted

$$PL(\text{dB}) = 140.7 + 36.7\log\left(\frac{d}{10^3}\right), \tag{24}$$

where $d$ is the distance between the transmitter and receiver. Because CoMP receiving is conducted mainly for cell-edge communication scenarios, we consider a cell radius value of 50m and also set the all the BS-UE distances to 50m, such that the UEs under consideration are at cell-edge. Assume each UE has AWGN noise power of $-100$ dBm. With uncoded binary phase-shift keying (BPSK) modulation, the coefficients for the outage probability approximations (12) and (14) are given by $c_1 = 1$, $c_2 = \frac{1}{2}$ [25], and $\frac{W}{R} = 1$. The variance parameters $\sigma_1^2 = \sigma_2^2 = 1$ are assumed, and the outage threshold $e_o = 10^{-4}$ is adopted. Simulation setting parameters are summarized in Table 1. The impacts of changing the total number of content files $K$ and the Zipf distribution parameter $\alpha$, as well as the choice of $F$, are investigated through numerical studies.

**TABLE 1. System settings for simulation.**

| | |
|---|---|
| Number of BSs | 4 |
| Number of UEs | 4 |
| Number of BS clusters | 2 |
| Maximum transmit power of each BS | 30 dBm |
| Noise power at each UE | -100 dBm |
| Antennas at each BS | $N_t = 2$ |
| Radius of each cell | 50 m |
| Path loss from BS to UE | $PL(\text{dB}) = 140.7 + 36.7\log\left(\frac{d}{10^3}\right)$ |
| Distance between each BS and its serving UE(s) | $d$ = 50 m |
| Modulation scheme | uncoded BPSK |
| Outage threshold | $10^{-4}$ |
| Other system parameters | $c_1 = 1, c_2 = \frac{1}{2}$<br>$\sigma_1^2 = \sigma_2^2 = 1$<br>$\frac{W}{R} = 1$ |

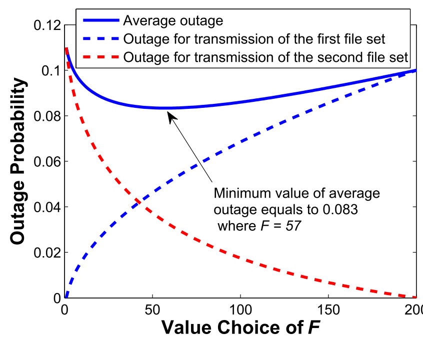



**FIGURE 3. Average outage probability, outage probability for transmissions of files in the first (popular) set, and the outage probability for transmissions of files in the second (less popular) set, with respect to the value of $F$. The system setting parameters are set to $K = 200$, and the $\alpha = 0.5$, $P_e^{FC}(e_o) = 0.1$, and $\bar{P}_e^{PC}(e_o) = 0.11$.**

The change of outage probabilities with the value of $F$ is presented in Fig. 3. Here the total number of files is set to $K = 200$, and the Zipf distribution parameter is equal to 0.5. As shown, the outage probability for transmission of files in the first (popular) set increases as we increase the value of $F$. This is because, with the increase of $F$, the number of files in the first set becomes larger, which results in higher transmission probability for the files in the first set, *i.e.*, a greater value of $\sum_{c=1}^{F} f_c$. Then, the value of $P_1^{out}$ increases accordingly as specified in eqn. (9). In contrast, the outage probability for transmission of files in the second (less popular) set decreases due to the reduced probability of transmission. When $F$ increases to the maximum value of 200, all files are in the first set. In this case, $\bar{P}_{out} = P_1^{out} = P_e^{FC}(e_o)$ and $P_2^{out} = 0$. Similarly, when $F$ decreases to zero, all transmission files are in the second set, and therefore, $\bar{P}_{out} = P_2^{out} = \bar{P}_e^{PC}(e_o)$ and $P_1^{out} = 0$. Since the outages for transmissions of files in two sets compete with each other, i.e., if one decreases, the other one shall increase, and vice versa, a tradeoff must be made to attain the minimum average outage. In Fig. 3, the optimal tradeoff point can be found, which exactly equals to the outcome of our scheme, where $F = 57$, and $\bar{P}_{out} = 0.083 < P_e^{FC}(e_o) = 0.1$. The outcome reveals a not so intuitive phenomenon that full coordination of all BSs does not always achieve the minimum average outage probability in CoMP transmission. This can be explained as full coordination does not schedule the request competition raised by multiple UEs, while our scheme does.

**TABLE 2.** Outage performance gain $\epsilon$ and optimal cache placement $F^*$ for different values of $\theta = \frac{\bar{P}_e^{PC}(e_o)}{P_e^{FC}(e_o)}$. $K = 200$ and $\alpha = 0.5$ are used for the numerical results.

| $\theta$ | 1.00 | 1.25 | 1.50 | 1.75 | 2.00 | 2.25 | 2.50 | 2.75 |
|---|---|---|---|---|---|---|---|---|
| $F^*$ | 47 | 74 | 99 | 123 | 145 | 166 | 185 | 200 |
| $\epsilon$ | 1.25 | 1.14 | 1.08 | 1.05 | 1.01 | 1.00 | 1.00 | 1.00 |

| $\theta = 2.75$ | | | | |
|---|---|---|---|---|
| $F$ | 50 | 100 | 150 | 200 |
| $\epsilon$ | 0.717 | 0.891 | 0.976 | 1.000 |

Define the outage performance gain as $\epsilon = \frac{P_e^{FC}(e_o)}{P^*(e_o)}$, where $P^*(e_o)$ is the outage probability achieved by optimal cache placement with respect to the outage threshold $e_o$. We present in Table 2 the values $F^*$ and $\epsilon$, as functions of the outage ratio $\theta = \frac{\bar{P}_e^{PC}(e_o)}{P_e^{FC}(e_o)}$. The system setting parameters are also set to $K = 200$ and $\alpha = 0.5$. It can be observed that when the ratio $\theta$ increases, the value of the optimal $F$ also increases. This is intuitive since when the outage of full coordination is lower than that of partial coordination, it is preferred to cache files in the first set, such that all the transmissions are conducted by full coordination to achieve better outage performance. However, it can also be observed that for a certain $\theta$, with the increase of $F$, the outage performance gain $\epsilon$ becomes more and more non-evident. For example, in the case of $\theta = 2.75$, the $\epsilon$ values for $F = 150$ and $F = 200$ are almost identical. This observation reveals a trade-off between fulfillment of the desired outage performance and content diversity.

By performing Monte Carlo simulation, Fig. 4 compares the theoretical outages among the proposed scheme and the baselines, where the outages values are obtained according to eqns. (12) and (14). It can be seen that the proposed scheme outperforms the other schemes in general. For the MPC strategy, since the method inclines to cache the most popular contents with all storage capacity, its performance is equal to $P_1^{out}$, which has a worse outage performance than the proposed scheme which optimizes $P_1^{out}$ and $P_2^{out}$ through optimal cache placement. Observe that caching with LCD is unfavorable compared with both the proposed scheme and the MPC strategy. This is due to the fact that according to LCD, half of the cache storage is cached with files in the first set and half is cached with files in the second set. As a consequence, the overall performance is the average of $P_1^{out}$ and $P_2^{out}$, which results in worse performance than both our scheme and the

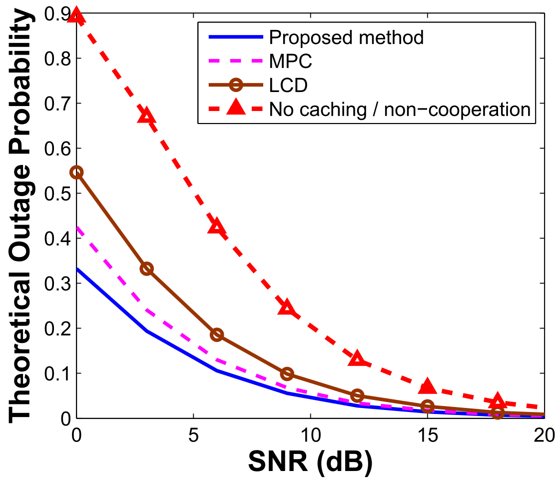



**FIGURE 4.** Comparison of the theoretical outage performance between the proposed scheme and the baseline schemes, where $K = 200$, and $\alpha = 0.5$ are used in the numerical studies.

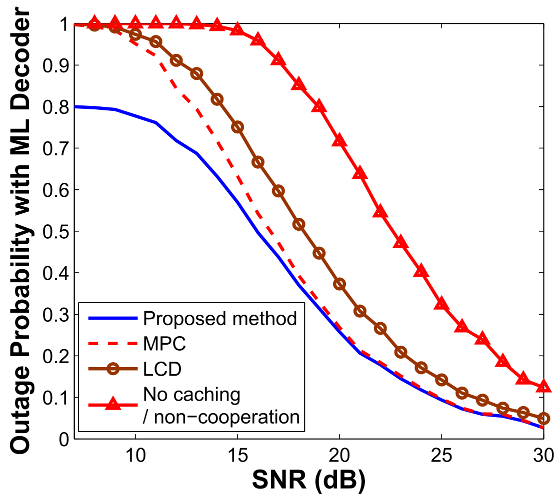



**FIGURE 5.** Comparison of the outage performance between the proposed scheme and the baseline schemes. In the simulation we assumed BPSK modulation at the transmitter, ML decoder at the receiver, $K = 200$, and $\alpha = 0.5$.

MPC scheme. In the case of no caching/cooperation, because each UE is served by only one BS, the performance is the worst in the comparison and therefore serves as a outage upper bound. In Fig. 5, outage performance comparison with maximum likelihood (ML) decoder at the receiver is given. The block BER exceeding the threshold of $10^{-4}$ is considered as a block in outage. Similar to Fig. 4, the results show that the proposed scheme has the best performance comparing to the baseline schemes. However, when the SNR goes over 20 dB, the performance gap between our proposed scheme and the MPC scheme diminishes. This can be explained as that in the high SNR regime, the difference between $P_1^{out}$ and $P_2^{out}$ is small and the optimization for cache placement achieves very little gain.

In Fig. 6, we present the optimal choice for $F$ for different values of the Zipf distribution parameter $\alpha$. It is shown

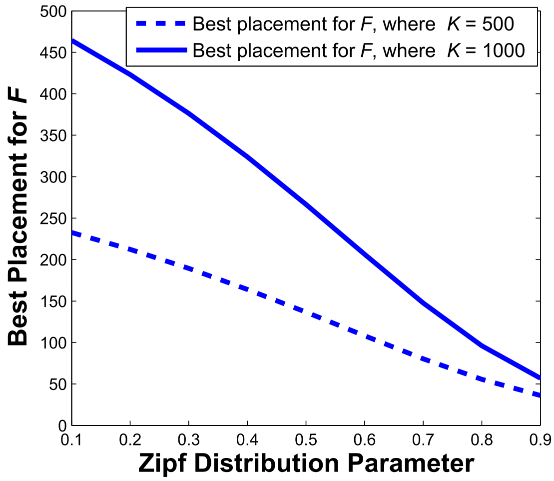



**FIGURE 6.** Optimal cache placement threshold $F^*$ for the popular files with respect to different values of the Zipf distribution parameter $\alpha$. where $P_e^{FC}(e_o) = 0.1$, and $\bar{P}_e^{PC}(e_o) = 1.1P_e^{FC}(e_o)$.

that, with the increase of the Zipf parameter $\alpha$, the value of $F^*$ decreases in the proposed scheme. This is to achieve a balance in transmission opportunities between the popular and less popular file sets. Specifically, when $\alpha$ increases, the distribution is more concentrated around the most popular contents. As a result, the request probability for the popular files increases, and the value of $F$ should be reduced to offer more transmission opportunities for the less popular files. Conversely, when $\alpha$ is getting small, the request probability for the less popular files increases. Larger $F$ is required to assign less transmission opportunity to the less popular files, such that the higher outage probability caused by the transmissions of the less popular files can be avoided.

## VI. CONCLUSIONS

In this paper, we have studied content-centric wireless communications and proposed a cache-aided scheme for downlink coordinated multi-point transmission. Based on content popularity, files are sorted and classified into two sets, a *popular* set and a *less popular* set. In order to efficiently utilize the limited cache resources to cache more files, we have proposed that files in the popular set are cached at all BSs, whereas the less popular files are cached only at part of the BSs strategically such that all the files can be cached at the local caches for CoMP transmission. As such, request for the popular files is cooperatively served by all the BSs, while request for the less popular files is served by part of the BSs collaboratively within a cache cluster. To assign transmission to requests having the best transmission channel condition, the request for popular files is given higher priority than that for less popular files. An optimal cache placement problem for two file sets has been formulated, and the optimal strategy for popular file selection and placement has been obtained, which minimizes the average outage probability by adjusting the popular file placement parameter. Through simulation studies, the proposed cache placement and cache-aided CoMP transmission scheme has been validated. It can be observed that the proposed scheme outperforms the traditional MPC, LCD, and no-caching schemes in terms of the average outage performance.

## APPENDIX A

*Proof of Lemma 1:*

*Proof:* Let $\Pr\{R_1(t) = n\}$ denote the probability of having $n$ requests for the popular files in a time period of length $t$. The file requests are assumed to be independent. A simplified notation $P_1(t, n)$ is used to represent $\Pr\{R_1(t) = n\}$ for convenience. Denote by $\delta_T$ a small time interval, we have

$$\begin{aligned} P_1(t+\delta_T, 0) &= \Pr\{R_1(t+\delta_T) = 0\} \\ &= \Pr\{R_1(t) = 0, R_1(t+\delta_T) - R_1(t) = 0\} \\ &= \Pr\{R_1(t) = 0\} \cdot \Pr\{R_1(t+\delta_T) - R_1(t) = 0\} \\ &= P_1(t, 0)\,[1 - \Pr\{R_1(t+\delta_T) - R_1(t) > 0\}] \\ &= P_1(t, 0)\,[1 - \lambda_1\delta_T + \mathcal{O}(\delta_T)]\,, \end{aligned} \tag{A.1}$$

in which $\mathcal{O}(\delta_T)$ is an infinitesimal of higher order than $\delta_T$, and $\lambda_1$ is the request possibility for the popular files. Based on definition, $\lambda_1 = \sum_{c=1}^{F} f_c$. Taking the limit of the change of request probability as $\delta_T$ approaches zero, we obtain

$$\lim_{\delta_T \to 0} \frac{P_1(t+\delta_T, 0) - P_1(t, 0)}{\delta_T} = \lim_{\delta_T \to 0} \frac{-\lambda_1 P_1(t, 0)\delta_T + \mathcal{O}(\delta_T)}{\delta_T}. \tag{A.2}$$

Therefore, we formulate a differential equation

$$\frac{\partial P_1(t, 0)}{\partial t} = -\left(\sum_{c=1}^{F} f_c\right) P_1(t, 0). \tag{A.3}$$

By considering the initial condition of $P_1(0, 0) = 1$ and solving eqn. (A.3), we obtain

$$P_1(t, 0) = \exp\left(-\sum_{c=1}^{F} f_c t\right). \tag{A.4}$$

For a unit time interval, the admission probability of the request for less popular files is equal to the probability of no request for the popular files, i.e., $P_1(1, 0)$. Therefore, we have Lemma 1 proved. ■

## APPENDIX B

*Proof of Theorem 1:*

*Proof:* Let $\phi = \frac{F'^{(1-\alpha)}-1}{K^{(1-\alpha)}-1}$. The first order derivative of $\bar{P}_{out}$ with respect to $F'$ is derived as

$$\begin{aligned} \frac{\partial \bar{P}_{out}(F')}{\partial F'} &= \frac{\partial \bar{P}_{out}(\phi)}{\partial \phi}\frac{\partial \phi(F')}{\partial F'} \\ &= \left(P_e^{FC} + \exp(-\phi) \cdot \bar{P}_e^{PC} \cdot (\phi - 2)\right) \frac{(1-\alpha)F'^{-\alpha}}{K^{(1-\alpha)}-1}. \end{aligned} \tag{B.1}$$

To obtain the optimal value of the relaxed variable $F'$, we first consider that the optimal value is in the interior of the feasible set, i.e., $(F')^* \in (0, K)$. Then we must have

$$(F')^* = \arg_{F'}\left(\frac{\partial \bar{P}_{out}(F')}{\partial F'} = 0\right). \tag{B.2}$$

It follows from eqn. (B.1) that $\frac{\partial \bar{P}_{out}(F')}{\partial F'} = 0$ if and only if

$$P_e^{FC} + \exp(-\phi)\bar{P}_e^{PC}(\phi - 2) = 0. \tag{B.3}$$

Substituting $\phi = \frac{F'^{(1-\alpha)}-1}{K^{(1-\alpha)}-1}$ into eqn. (B.3), and from eqn. (B.2), it can be shown that the optimum $(F')^*$ has the form as expressed by (23). By letting $F^* = \lceil (F')^* \rceil$, the solution to the original problem can be obtained approximately by rounding up the solution to the relaxed continuous problem.

Then we assume that the optimal $F'$ is on the boundary of the feasible set, i.e., $(F')^* = K$ or $(F')^* = 0$. In the case $(F')^* = K$, from the optimality condition we must have [28]

$$\left[\frac{\partial \bar{P}_{out}(F')}{\partial F'}\bigg|_{F'=K}\right]\underbrace{(F' - K)}_{①} \geq 0, \quad \forall F' \in [0, K]. \tag{B.4}$$

By substituting $(F')^* = K$ into eqn. (B.1), we obtain

$$\frac{\partial \bar{P}_{out}(F')}{\partial F'}\bigg|_{F'=K} = \underbrace{\left[P_e^{FC} - \exp(-1)\cdot \bar{P}_e^{PC}\right]}_{②}\underbrace{\frac{(1-\alpha)K^{(-\alpha)}}{K^{(1-\alpha)}-1}}_{③}. \tag{B.5}$$

Note that in eqns. (B.4) and (B.5), ① is less than zero, ③ is greater than zero, while the sign of ② is uncertain. The assumption of eqn. (B.4) cannot be true. Therefore, the boundary point $K$ is obviously not the optimal solution to the relaxed continuous problem. However, we can prove that $K$ obtains the minimum value for the objective function in eqn. (20) on $[0, K]$ if the feasible set is relaxed to $\mathbb{R}_+$. Assume $K < (F')^* \in \mathbb{R}+$. As an interior point of the feasible set $\mathbb{R}_+$, $(F')^*$ can be obtained by eqn. (23). Recall that according to Lemma 2, the objective function in eqn. (20) is convex for the relaxed continuous problem. Based on the first order condition of a convex problem, we obtain that for any $x_0 \in \mathbb{R}_+$,

$$\underbrace{\bar{P}_{out}\left((F')^*\right)}_{④} \geq \underbrace{\bar{P}_{out}(x_0)}_{⑤} + \underbrace{\left[\frac{\partial \bar{P}_{out}(F')}{\partial F'}\bigg|_{F'=x_0}\right]}_{⑥}\cdot\underbrace{\left[(F')^* - x_0\right]}_{⑦}. \tag{B.6}$$

Observe that ⑤ > ④ because $(F')^*$ is the optimal solution, and ⑦ > 0 if $x_0 < (F')^*$. Then it follows from (B.6) that ⑥ < 0. It implies that the objective function given by eqn. (20) is monotonically decreasing on $[0, K] \subset \mathbb{R}_+$. Therefore, we have proved that, if the feasible set is relaxed to $\mathbb{R}_+$, then $K$ attains the minimum for the objective function (20) on $[0, K]$, or equivalently, $K$ is the optimal solution to Problem 1 if $(F')^*$ obtained by eqn. (23) is greater than $K$. It can be shown in a similar way that the case of $(F')^* = 0$ is not optimal. Theorem 1 is then proved. ■

## REFERENCES


[1] Cisco. *Visual Networking Index: Global Mobile Data Traffic Forecast Update 2016–2021*. Accessed: Mar. 28, 2017. [Online]. Available: http://www.cisco.com/c/en/us/solutions/collateral/service-provider/vni-service-adoption-forecast/Cisco_VNI_SA_Forecast_WP.pdf

[2] V. Jungnickel *et al.*, "The role of small cells, coordinated multipoint, and massive MIMO in 5G," *IEEE Commun. Mag.*, vol. 52, no. 5, pp. 44–51, May 2014.

[3] Y. Cui, D. Jiang, and Y. Wu, "Analysis and optimization of caching and multicasting in large-scale cache-enabled wireless networks," *IEEE Trans. Wireless Commun.*, vol. 15, no. 7, pp. 5101–5112, Jul. 2016.

[4] F. Xu, K. Liu, and M. Tao, "Cooperative Tx/Rx caching in interference channels: A storage-latency tradeoff study," in *Proc. IEEE Int. Symp. Inf. Theory (ISIT)*, Barcelona, Spain, Jul. 2016, pp. 2034–2038.

[5] Y. Yang, B. Bai, W. Chen, and L. Hanzo, "A low-complexity cross-layer algorithm for coordinated downlink scheduling and robust beamforming under a limited feedback constraint," *IEEE Trans. Veh. Technol.*, vol. 63, no. 1, pp. 107–118, Jan. 2014.

[6] H. Zhang, H. Liu, C. Jiang, X. Chu, A. Nallanathan, and X. Wen, "A practical semidynamic clustering scheme using affinity propagation in cooperative picocells," *IEEE Trans. Veh. Technol.*, vol. 64, no. 9, pp. 4372–4377, Sep. 2015.

[7] L. Su, C. Yang, and S. Han, "The value of channel prediction in CoMP systems with large backhaul latency," *IEEE Trans. Commun.*, vol. 61, no. 11, pp. 4577–4590, Nov. 2013.

[8] N. Wang, E. Hossain, and V. K. Bhargava, "Backhauling 5G small cells: A radio resource management perspective," *IEEE Wireless Commun.*, vol. 22, no. 5, pp. 41–49, Oct. 2015.

[9] C. Liang, Y. He, F. R. Yu, and N. Zhao, "Enhancing QoE-aware wireless edge caching with software-defined wireless networks," *IEEE Trans. Wireless Commun.*, vol. 16, no. 10, pp. 6912–6925, Oct. 2017.

[10] K. Shanmugam, N. Golrezaei, A. G. Dimakis, A. F. Molisch, and G. Caire, "FemtoCaching: Wireless content delivery through distributed caching helpers," *IEEE Trans. Inf. Theory*, vol. 59, no. 12, pp. 8402–8413, Dec. 2013.

[11] E. Bastug, M. Bennis, and M. Debbah, "Living on the edge: The role of proactive caching in 5G wireless networks," *IEEE Commun. Mag.*, vol. 52, no. 8, pp. 82–89, Aug. 2014.

[12] W. Chen and H. V. Poor, "Joint pushing and caching with a finite receiver buffer: Optimal policies and throughput analysis," in *Proc. IEEE Int. Conf. Commun. (ICC)*, Kuala Lumpur, Malaysia, May 2016, pp. 1–6.

[13] A. Liu and V. K. N. Lau, "Mixed-timescale precoding and cache control in cached MIMO interference network," *IEEE Trans. Signal Process.*, vol. 61, no. 24, pp. 6320–6332, Dec. 2013.

[14] M. Tao, E. Chen, H. Zhou, and W. Yu, "Content-centric sparse multicast beamforming for cache-enabled Cloud RAN," *IEEE Trans. Wireless Commun.*, vol. 15, no. 9, pp. 6118–6131, Sep. 2016.

[15] Z. Chen, J. Lee, T. Q. S. Quek, and M. Kountouris, "Cooperative caching and transmission design in cluster-centric small cell networks," *IEEE Trans. Wireless Commun.*, vol. 16, no. 5, pp. 3401–3415, May 2017.

[16] B. Zhou, Y. Cui, and M. Tao, "Stochastic content-centric multicast scheduling for cache-enabled heterogeneous cellular networks," *IEEE Trans. Wireless Commun.*, vol. 15, no. 9, pp. 6284–6297, Sep. 2016.

[17] S. H. Chae, T. Q. S. Quek, and W. Choi, "Content placement for wireless cooperative caching helpers: A tradeoff between cooperative gain and content diversity gain," *IEEE Trans. Wireless Commun.*, vol. 16, no. 10, pp. 6795–6807, Oct. 2017.

[18] G. Zheng, H. A. Suraweera, and I. Krikidis, "Optimization of hybrid cache placement for collaborative relaying," *IEEE Commun. Lett.*, vol. 21, no. 2, pp. 442–445, Feb. 2017.

[19] M. Taghizadeh, K. Micinski, S. Biswas, C. Ofria, and E. Torng, "Distributed cooperative caching in social wireless networks," *IEEE Trans. Mobile Comput.*, vol. 12, no. 6, pp. 1037–1053, Jun. 2013.

[20] M. Li, I. B. Collings, S. V. Hanly, C. Liu, and P. Whiting, "Multicell coordinated scheduling with multiuser zero-forcing beamforming," *IEEE Trans. Wireless Commun.*, vol. 15, no. 2, pp. 827–842, Feb. 2016.

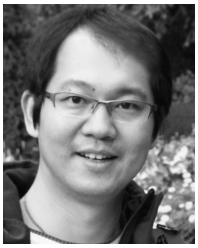
[21] H. Zhang, C. Jiang, and J. Cheng, "Cooperative interference mitigation and handover management for heterogeneous cloud small cell networks," *IEEE Wireless Commun.*, vol. 22, no. 3, pp. 92–99, Jun. 2015.

[22] P. Viswanath, D. N. C. Tse, and R. Laroia, "Opportunistic beamforming using dumb antennas," *IEEE Trans. Inf. Theory*, vol. 48, no. 6, pp. 1277–1294, Jun. 2002.

[23] A. Mahmood and R. Jäntti, "Packet error rate analysis of uncoded schemes in block-fading channels using extreme value theory," *IEEE Commun. Lett.*, vol. 21, no. 1, pp. 208–211, Jan. 2017.
[24] M. Chiani, D. Dardari, and M. K. Simon, "New exponential bounds and approximations for the computation of error probability in fading channels," *IEEE Trans. Wireless Commun.*, vol. 2, no. 4, pp. 840–845, Jul. 2003.
[25] R. G. Gallager, "Error probability for orthogonal signal sets," in *Principles of Digital Communication*, 1st ed. New York, NY, USA: Cambridge Univ. Press, 2008, ch. 8, sec. 5, pp. 294–298.
[26] M. S. Motlagh, M. B. Khuzani, and P. Mitran, "On lossy joint source-channel coding in energy harvesting communication systems," *IEEE Trans. Commun.*, vol. 63, no. 11, pp. 4433–4447, Nov. 2015.
[27] Z. Xu, C. Yang, G. Y. Li, Y. Liu, and S. Xu, "Energy-efficient CoMP precoding in heterogeneous networks," *IEEE Trans. Signal Process.*, vol. 62, no. 4, pp. 1005–1017, Feb. 2014.
[28] D. P. Bertsekas, "Optimization over a convex set," in *Nonlinear Programming*, 3rd ed. Belmont, MA, USA: Athena Scientific, 2016, ch. 3, sec. 1, pp. 236–237.

**FASHENG ZHOU** (M'16) received the Ph.D. degree in information and communication engineering from the South China University of Technology, Guangzhou, China. He is currently with Guangzhou University, Guangzhou. Before joining Guangzhou University, he was with Huawei Company for nine years. His current research interests are edge caching, coordinated communications, and heterogeneous small cell networks.

**LISHENG FAN** received the bachelor's degree from the Department of Electronic Engineering, Fudan University, in 2002, the master's degree from the Department of Electronic Engineering, Tsinghua University, China, in 2005, and the Ph.D. degree from the Department of Communications and Integrated Systems, Tokyo Institute of Technology, Japan, in 2008. He is currently a Professor with Guangzhou University. His research interests span the areas of wireless cooperative communications, physical-layer secure communications, interference modeling, and system performance evaluation. He has authored or coauthored many papers in international journals, such as the IEEE Transactions on Wireless Communications, the IEEE Transactions on Communications, and the IEEE Transactions on Information Theory, and papers in conferences, such as the IEEE ICC, the IEEE Globecom, and the IEEE WCNC. He has served as a member of Technical Program Committees for IEEE conferences, such as Globecom, ICC, WCNC, and VTC. He is a Guest Editor of *EURASIP Journal on Wireless Communications and Networking*, and served as the Chair of the Wireless Communications and Networking Symposium for Chinacom 2014.



**NING WANG** (S'09–M'13) received the B.E. degree in communication engineering from Tianjin University in 2004, the M.A.Sc. degree in electrical engineering from The University of British Columbia in 2010, and the Ph.D. degree in electrical engineering from the University of Victoria in 2013. He was a Post-Doctoral Research Fellow with the Department of Electrical and Computer Engineering, The University of British Columbia. From 2004 to 2008, he was with the China Information Technology Design and Consulting Institute, as a Mobile Communication System Engineer, specializing in large-scale commercial mobile communication system design, network traffic analysis, and radio network optimization. He is currently an Associate Professor with the School of Information Engineering, Zhengzhou University, Zhengzhou, China. His research interests include the resource allocation and security design of future cellular networks, channel modeling for wireless communications, statistical signal processing, and cooperative wireless communications.

**GAOYONG LUO** received the Ph.D. degree in electrical engineering from Brunel University, U.K. Since 1998, he has been with the Buckinghamshire Chilterns University College and Buckinghamshire New University. He is currently a Professor and has been the Field Chair of electronics information and communication engineering at Guangzhou University. He has taught thousands of students and supervised the work of master's and Ph.D. students. He has authored or coauthored over 100 refereed journal/conference papers and patents. He is the author of the book *Wavelets in Engineering Applications* (Science Press, 2014). His main research interests are in the field of wavelets, spread spectrum communications, wireless positioning, remote sensing, audio coding, and power line communications, with expertise in coding and modulation theory and applications to communication systems. He is an IET member, and serves as a technical program committee member or chair for a number of international conferences. His current research work has been mainly focused on wavelet multicarrier multiplexing, high precision wireless indoor positioning, and low-latency audio communication over 5G network.

**JIE TANG** (S'10–M'13) received the B.Eng. degree from the South China University of Technology, Guangzhou, China, in 2008, the M.Sc. degree (Hons.) from the University of Bristol, U.K., in 2009, and the Ph.D. degree from Loughborough University, Leicestershire, U.K., in 2012. From 2013 to 2015, he held a post-doctoral research position at the School of Electrical and Electronic Engineering, The University of Manchester, U.K. He is currently an Associate Professor with the School of Electronic and Information Engineering, South China University of Technology. His research interests include optimization techniques and the analysis of wireless communication networks, with particular focus on 5G networks, green communications, SWIPT systems, heterogeneous networks, cognitive radio, and MIMO systems. He is currently serving in the editorial boards of *EURASIP Journal on Wireless Communications and Networking*, *Physical Communications*, and *Ad Hoc and Sensor Wireless Networks*. He is also a Symposium Co-Chair for VTC Spring 2018.

**WEI CHEN** (S'05–M'07–SM'13) received the B.S. degree (Hons.) in operations research and the Ph.D. degree (Hons.) in electronic engineering from Tsinghua University, Beijing, China, in 2002 and 2007, respectively. From 2005 to 2007, he was also a Visiting Research Staff with the Hong Kong University of Science and Technology, Hong Kong. Since 2007, he has been with the Department of Electronic Engineering, Tsinghua University, where he is currently a Full Professor and the Vice Department Head. Since 2012, he has been the youngest Full Professor with Tsinghua University. Since 2010, he visited the University of Southampton, Southampton, U.K. Since 2014, he has been with Telecom ParisTech, Paris, France. In 2008, 2009, 2012, and 2014, he has also visited the Hong Kong University of Science and Technology and the Chinese University of Hong Kong, Hong Kong. He also holds Guest Professor positions with several universities in China. He is a National 973 Youth Project Chief Scientist and a National Young and Middle-Aged Leader for science and technology innovation. He is supported by the National Natural Science Foundation of China Excellent Young Investigator Project, the New Century Talent Program of the Ministry of Education, the Beijing Nova Program, and the 100 Fundamental Research Talents Program of Tsinghua University (also known as 221 Talents Program). His research interests include wireless communications, information theory, and applied mathematics. He is a member of the 2013 Education Working Group of the IEEE Communication Society and the Vice Director of the Youth Committee of China Institute of Communications and of the Youth Board of the Education Committee of China Institute of Electronics. He was a recipient of the First Prize of the 14th Henry Fok Ying-Tung Young Faculty Award; the Yi-Sheng Mao Beijing Youth Science and Technology Award; the 2010 IEEE Comsoc Asia Pacific Board Best Young Researcher Award; the 2009 IEEE Marconi Prize Paper Award; the Best Paper Awards at the 2006 IEEE International Conference on Communications (IEEE ICC), the 2007 IEEE International Workshop on Cross Layer Design, and the 2012 IEEE International Conference on Smart Grid Communications; the First Prize in the Seventh Young Faculty Teaching Competition in Beijing; and the 2011 Tsinghua Raising Academic Star Award and the 2012 Tsinghua Teaching Excellence Award, which are, respectively, the highest research and education honors for young faculties (age $\leq$ 41) in Tsinghua University. He holds the honorary titles of Beijing Outstanding Teacher and Beijing Outstanding Young Talent. He is the Champion of the First National Young Faculty Teaching Competition and a winner of the National May 1st Medal. He currently serves as an Editor for the IEEE TRANSACTIONS ON EDUCATION, the IEEE WIRELESS COMMUNICATIONS LETTERS, and *China Communications*. He served as a Tutorial Co-Chair for the 2013 IEEE International Conference on Communications, as a Technical Program Committee Co-Chair for the 2011 Spring IEEE Vehicular Technology Conference, and as a Symposium Co-Chair for the IEEE ICC, the IEEE/CIC International Conference on Communications in China, and the IEEE Consumer Communications and Networking Conference.