Analysis and Optimization of Probabilistic Caching in Multi-Antenna Small-Cell Networks

Xianzhe Xu and Meixia Tao
Dept. of Electronic Engineering, Shanghai Jiao Tong University, Shanghai, China
Emails: {august.xxz, mxtao}@sjtu.edu.cn

***Abstract*—Previous works on cache-enabled small-cell networks (SCNs) with probabilistic caching often assume that each user is connected to the nearest small base station (SBS) among all that have cached its desired content. The user may, however, suffer strong interference from other SBSs which do not cache the desired content but are geographically closer than the serving SBS. In this work, we investigate this issue by deploying multiple antennas at each SBS. We first propose a user-centric SBS clustering model where each user chooses its serving SBS from a cluster of closest SBSs with a fixed cluster size only. Two beamforming schemes are considered. One is coordinated beamforming, where each SBS uses zero-forcing (ZF) beamformer to null out the interference within its coordination cluster. The other is uncoordinated beamforming, where each SBS simply applies matched-filter (MF) beamformer. Using tools from stochastic geometry, we obtain tractable expressions for the successful transmission probability of a typical user for both schemes in the high signal-to-noise ratio (SNR) region. Tight approximations in closed-form expressions are also obtained. We then formulate and solve the optimal probabilistic caching problem to maximize the successful transmission probability. Numerical results reveal interesting insights on the choices of ZF and MF beamforming to maximize the performance in cache-enabled SCNs.**

I. Introduction

In recent decades, mobile data traffic has grown rapidly due to the proliferation of mobile devices, which imposes heavy pressure on the limited backhaul link in cellular networks. Caching popular contents in small base stations (SBSs) at off-peak time is a promising way to alleviate the backhaul pressure by avoiding repetitive transmissions in the backhaul link. It can also improve user-perceived experience due to the reduced transmission latency.

Many prior works have studied cache-enabled small-cell networks (SCNs) with various caching strategies and optimization objectives. Thus far, there are two commonly used caching strategies. One is for each SBS to cache the most popular contents as in [1], [2], which is suitable for networks where each user can only be associated with its nearest SBS. The other is for each SBS to cache contents randomly with probabilities to be optimized as in [3]–[6]. In particular, the work [3] formulates and solves the probabilistic cache placement problem to maximize the cache hit rate using stochastic geometry. It shows that caching the contents randomly with the optimized probabilities has significant gain over caching the most popular contents since each user can be covered by multiple SBSs. However, when a user is not served by the nearest SBS, the strong interference from the closer SBSs will severely harm the received signal-to-interference ratio (SIR). It is thus crucial to study interference management schemes with probabilistic caching in SCNs. Many works have studied the transmission schemes to solve this problem to enhance the performance in cache-enabled networks. The authors in [7] make an initial attempt to apply joint transmission and successive interference cancelation (SIC) in cache-enabled SCNs and study the tradeoff between transmission diversity and content diversity. The authors in [8] employ zero-forcing (ZF) beamforming to null interference and optimize the cache placement for maximizing the success probability and area spectral efficiency.

In this paper, we aim to exploit the benefits of multiple antennas in cache-enabled SCNs when probabilistic caching is adopted. Multiple antennas can be used to cancel interference from neighboring cells with SBS coordination, or only to strengthen the effective channel gain of desired signals without SBS coordination, both yielding significant increase in the received SINR. The study in [7] is limited to disjoint SBS clustering and the analysis only focuses on a user located at the cluster center. In this work, we consider a typical user and employ user-centric dynamic SBS clustering for user association in cache-enabled SCNs. Each user is allowed to associate with its nearest multiple SBSs and each SBS caches files independently with identical probabilities. We study the performance in cache-enabled SCNs with ZF and matched-filter (MF) beamformings. The main contributions are summarized as follows.

- We propose a user-centric dynamic SBS clustering model in cache-enabled multi-antenna SCNs with probabilistic caching, where each user is allowed to access one of the SBSs within its cluster. Each SBS adopts ZF-based coordinated beamforming when transmission coordination is allowed within the SBS cluster or MF beamforming otherwise.
- We obtain the closed-form (approximate) upper and lower bounds of the coverage probability of a typical user in both MF and ZF beamforming scenarios in the high signal-to-noise ratio (SNR) region using tools from stochastic geometry. Our analysis extends that in [9] by allowing each user to be associated with the k-th nearest SBS, with k being any integer not larger than the cluster size.
- We formulate an optimal probabilistic caching problem to maximize the probability that the typical user successfully receives its requested file locally from the SBSs within its cluster. The problem is shown to be convex and solved by

Lagrangian method. Closed-form optimal cache solutions are also obtained.

• Extensive numerical results reveal that the optimal probabilistic caching strategy has a gain over the most popular caching. More importantly, it is found that when the number of antennas is the same as the cluster size, MF outperforms ZF. While when the number of antennas is larger than the cluster size, ZF performs better.

Notations: This paper uses bold-face lower-case $\mathbf{h}$ for vectors and bold-face uppercase $\mathbf{H}$ for matrices. $\mathbf{H}^H$ is the conjugate transpose of $\mathbf{H}$ and $\mathbf{H}^\dagger$ is the left pseudo-inverse of $\mathbf{H}$, defined as $\mathbf{H}^\dagger = (\mathbf{H}^H\mathbf{H})^{-1}\mathbf{H}^H$. We use $\mathbf{I}_m$ to denote a $m \times m$ identity matrix and $\mathbf{0}_{1\times m}$ denotes a $1 \times m$ zero vector.

II. System Model

We consider a cache-enabled multi-antenna SCN, where the locations of SBSs are modeled as a homogenous Poisson point process (HPPP) $\Phi_b = \{d_i \in \mathbb{R}^2, \forall i \in \mathbb{N}^+\}$ with intensity λ_b. Each SBS is equipped with L antennas. The intensity of users is λ_u and each user is equipped with a single antenna. We assume that $\lambda_u \gg \lambda_b$ so that the network is fully loaded with each SBS serving one user at a time.

We consider a file library $\mathcal{F} = \{f_1, f_2, \cdots, f_N\}$, where N is the total number of files. All files are assumed to have the same normalized size of 1 for analytical tractability. The popularity of files follows the Zipf distribution with the popularity of file f_n given by:

$$p_n = \frac{1/n^\delta}{\sum_{j=1}^N 1/j^\delta}, \quad n = 1, 2, \ldots, N, \tag{1}$$

where $\delta \geq 0$ is the skewnesss parameter. Each SBS has a local cache that can store up to M files with $M < N$. We adopt the *probabilistic caching* strategy, where each SBS caches file f_n with probability b_n independently. Due to the cache size constraint and probability property, we have the constraints: $\sum_{n=1}^N b_n \leq M$ and $0 \leq b_n \leq 1$ for $n = 1, 2, \ldots, N$. Note that a practical cache placement strategy with given probabilities can be found in [3].

A. User Association Strategy

In this work, each user is allowed to choose its serving SBS from a cluster of K SBSs that are closest to the user, where $K \geq 2$ is a positive integer. We shall refer to the K closest SBSs of each user as the user-centric SBS cluster with size of K. When a user submits a file request, the nearest SBS within the cluster that have cached the requested file will serve the user. If none of the K SBSs in the cluster caches the requested file, a macro base station (MBS) will download the file from the core network via backhaul link and then transmit it to the user. By such user-centric SBS clustering, the plane is tessellated into K-th order Voronoi regions, denoted as $\mathcal{V}_K(d_1, \cdots, d_K)$. The K-th order Voronoi cell associated with a set of K points $\{d_1, \cdots, d_K\}$ is the region that all the points in this region are closer to these K points than to any other point of Φ_b, i.e., $\mathcal{V}_K(d_1, \cdots, d_K) = \{d \in \mathbb{R}^2 | \cap_{k=1}^K \{\|d - d_k\| \leq \|d - d_i\|\}, d_i \in \Phi_b \backslash \{d_1, d_2, \cdots, d_K\}\}$.

Without loss of generality, we focus on a typical user u_0 which is located at the origin, and can choose to connect to any of the SBSs in the K-th order Voronoi cell that it belongs to, denoted as $\mathcal{C} = \{d_1, d_2, \cdots, d_K\}$. The distance between u_0 and the k-th nearest SBS d_k is r_k. As assumed earlier, the user intensity is much larger than SBS intensity, and hence the network is fully loaded with all the SBSs being active. We consider an interference-limited network where the noise can be neglected. Hence, the received signal of u_0 when associated with the k-th nearest SBS d_k, for $k = 1, 2, \ldots, K$ is given by:

$$y_0 = r_k^{-\frac{\alpha}{2}} \mathbf{h}_{k0} \mathbf{w}_k x_k + \sum_{j \in \Phi_b \backslash \{d_k\}} r_j^{-\frac{\alpha}{2}} \mathbf{h}_{j0} \mathbf{w}_j x_j, \tag{2}$$

where the channel is assumed to have both small-scale fading, denoted as $\mathbf{h}_{i0}$, and large-scale fading $r_i^{-\frac{\alpha}{2}}$. The small-scale fading is modeled as Rayleigh fading, i.e., $\mathbf{h}_{i0} \sim \mathcal{CN}(\mathbf{0}_{1\times L}, \mathbf{I}_L)$ and the large-scale fading $r_i^{-\frac{\alpha}{2}}$ follows the distance-dependent power law model with $\alpha > 2$ being the path loss exponent. x_i and $\mathbf{w}_i \in \mathbb{C}^{L\times 1}$ denote the transmit signal and the associated beamformer vector, respectively.

We consider two types of beamforming design in each SBS. One is uncoordinated, where each SBS applies an MF based beamforming independently to maximize the effective channel gain of its own user. The other is coordinated, where the K SBSs in each K-th order Voronoi cell apply ZF beamforming coordinately so that the intra-cluster interference can be nulled out completely. This requires that the number of antennas at each SBS should not be smaller than the cluster size, i.e. $L \geq K$. Note that since we consider a user-centric SBS clustering for coordinated beamforming, all the other $K - 1$ users involved in the coordination must also locate in the same K-th order Voronoi cell as the typical user. This requirement is enforced to justify the *typicality* of the typical user in our analysis and it can be easily satisfied given that the user intensity is much larger than the SBS intensity. However, this requirement is not needed for the uncoordinated beamforming.

B. Matched-Filter Beamforming

When the typical user is associated with SBS $d_k \in \mathcal{C}$, the MF beamforming vector $\mathbf{w}_k$ is given by:

$$\mathbf{w}_{k,\text{mf}} = \frac{\mathbf{h}_{k0}^H}{\|\mathbf{h}_{k0}\|}. \tag{3}$$

Since each SBS serves its own user independently, the interference for the typical user comes from all SBSs except the serving SBS d_k in the network. Thus, based on (2), the SIR for u_0 is given by:

$$\text{SIR}_{k,\text{mf}} = \frac{g_{k,\text{mf}} \cdot r_k^{-\alpha}}{\sum_{j \in \Phi_b \backslash \{d_k\}} g_{j,\text{mf}} \cdot r_j^{-\alpha}}, \tag{4}$$

where $g_{k,\text{mf}} = \|\mathbf{h}_{k0}\|^2$ is the effective channel gain of the desired signal and it follows the Gamma distribution with shape parameter L and scale parameter 1, denoted as $g_{k,\text{mf}} \sim \Gamma(L, 1)$, and $g_{j,\text{mf}} = |\mathbf{h}_{j0}\mathbf{w}_{j,\text{mf}}|^2$ is the effective channel gain of the undesired signal and follows the exponential distribution with parameter 1, denoted as $g_{j,\text{mf}} \sim \exp(1)$ [10].

C. Zero-Forcing Beamforming

In the ZF beamforming vector design, the $K-1$ intra-cluster interferences of the typical user are completely nulled [9]. To ensure the feasibility, we assume $L \geq K$ as mentioned earlier. The beamforming vector is given by:

$$\mathbf{w}_{k,\text{zf}} = \frac{(\mathbf{I}_L - \mathbf{H}\mathbf{H}^{\dagger})\mathbf{h}_{k0}^T}{\|(\mathbf{I}_L - \mathbf{H}\mathbf{H}^{\dagger})\mathbf{h}_{k0}^T\|}, \tag{5}$$

where $\mathbf{H} = [\mathbf{h}_{k1}^T, \mathbf{h}_{k2}^T, \cdots, \mathbf{h}_{k(K-1)}^T]$ is the channel between the serving SBS d_k of the typical user and the $K-1$ users served by the other $K-1$ SBSs in the coordination group. By (5), the interference caused by SBS d_k to all other $K-1$ users in the cluster is canceled. The other $K-1$ SBSs from $\mathcal{C}$ adopt the same ZF beamforming strategy. Thus, the interference for the typical user only comes from the SBSs out of the cluster. Therefore, the SIR for the typical user is given by:

$$\text{SIR}_{k,\text{zf}} = \frac{g_{k,\text{zf}} \cdot r_k^{-\alpha}}{\sum_{j\in\Phi_b\setminus\mathcal{C}} g_{j,\text{zf}} \cdot r_j^{-\alpha}}, \tag{6}$$

where $g_{k,\text{zf}} = |\mathbf{h}_{k0}\mathbf{w}_{k,\text{zf}}|^2$ is the effective channel gain of the desired signal and follows $g_{k,\text{zf}} \sim \Gamma(L-K+1, 1)$, and $g_{j,\text{zf}} = |\mathbf{h}_{j0}\mathbf{w}_{j,\text{zf}}|^2$ is the effective channel gain of the undesired signal and follows $g_{j,\text{zf}} \sim \exp(1)$ [10].

Before concluding this section, we would like to remark that ZF beamforming requires each SBS to know the channel state information (CSI) of all the K users in the cluster, therefore extra signalling overhead is needed. On the other hand, MF beamforming only require each SBS to know the CSI of its only user.

III. PERFORMANCE ANALYSIS

In this paper, we use the successful transmission probability (STP) of the typical user as the performance metric.

The successful transmission probability, denoted as $P_{\text{suc}}(K)$, is defined as the probability that the typical user can successfully receive its requested file locally from the cluster of K closest SBSs without resorting to the core network. A file can be successful transmitted locally if and only if it is cached and the received SIR exceeds a given SIR target. The probability that the received SIR exceeds a SIR target is called the coverage probability. Denote $P_{\text{cov}}^k(K)$ as the coverage probability of the typical user served by the k-th $(k = 1, 2 \ldots K)$ nearest SBS in the cluster and it is given by:

$$P_{\text{cov}}^k(K) = P[\text{SIR}_k \geq \gamma], \tag{7}$$

where γ is the given SIR threshold.

Therefore, the successful transmission probability is given by:

$$P_{\text{suc}}(K) = \sum_{n=1}^{N} p_n \sum_{k=1}^{K} b_n(1-b_n)^{k-1} P_{\text{cov}}^k(K), \tag{8}$$

where p_n is the request probability of file f_n and b_n is the cache probability of file f_n.

Our goal is to analyze $P_{\text{suc}}(K)$ with MF and ZF beamformings, respectively, then optimize the cache probabilities for both cases and finally compare their performances.

In the rest of this section, we analyze the coverage probability $P_{\text{cov}}^k(K)$ in both MF and ZF beamforming cases. Note that the works [9], [11], [12] have studied the SIR distribution in multi-antenna networks with coordinated beamforming using stochastic geometry tools. In particular, [11] obtains the characteristic function of other-cell interference when both base stations and users are equipped with multiple antennas. The work [12] obtains a tractable and approximation expression of the complementary cumulative distribution function (CCDF) of SIR when the user is served by the nearest SBS only. The work [9] further obtains closed-form upper and lower bounds of the CCDF of SIR. Our analysis differs from the above references in that each user is allowed to be served by the k-th nearest SBS, for $k = 1, 2, \ldots, K$, which (to our best knowledge, has not been studied before in the literature) makes our problem more challenging and more general.

A. Coverage Probability with MF Beamforming

Lemma 1: The coverage probability of the typical user served by the k-th nearest SBS with MF beamforming is:

$$P_{\text{cov,mf}}^k(K) = \mathbb{E}_{r_k}\left[\sum_{i=0}^{L-1} \frac{(-\gamma {r_k}^{\alpha})^i}{i!} \mathcal{L}_{I_r}^{(i)}(\gamma {r_k}^{\alpha})\right], \tag{9}$$

where $I_r = \sum_{j\in\Phi_b\setminus\{d_k\}} g_{j,\text{mf}} \cdot r_j^{-\alpha}$, $\mathcal{L}_{I_r}(s) = \mathbb{E}[e^{-sI_r}]$ is the Laplace transform of I_r given by:

$$\mathcal{L}_{I_r}(s) = \left(\int_0^{r_k} \frac{1}{1+sr^{-\alpha}} \frac{2r}{r_k^2} dr\right)^{k-1} \times \exp\left(-2\pi\lambda_b \int_{r_k}^{\infty} \frac{sr^{-\alpha}}{1+sr^{-\alpha}} r dr\right), \tag{10}$$

and $\mathcal{L}_{I_r}^{(i)}(s)$ is the i-th order derivative of $\mathcal{L}_{I_r}(s)$.

Proof: See Appendix A. ■

With the expression (9), we still need to obtain the expectation over r_k. The pdf of r_k is given in Lemma 2.

Lemma 2 ([13]): Given a HPPP in the 2-dimensional plane with intensity λ_b, the distance R_k of the k-th nearest point to the origin follows the generalized Gamma distribution:

$$f_{R_k}(r) = \frac{2\left(\lambda_b \pi r^2\right)^k}{r\Gamma(k)} \exp\left(-\lambda_b \pi r^2\right). \tag{11}$$

With Lemma 1 and Lemma 2, we can get a tractable expression of the coverage probability. In the following, we provide more compact forms to bound the coverage probability.

Theorem 1: The coverage probability of the typical user served by the k-th nearest SBS using MF beamforming is bounded as:

$$P_{\text{cov,mf}}^{k,\text{l}}(K) \leq P_{\text{cov,mf}}^k(K) \leq P_{\text{cov,mf}}^{k,\text{u}}(K), \tag{12}$$

with

$$P_{\text{cov,mf}}^{k,\text{u}}(K) = \sum_{l=1}^{L} \beta_1(\eta, \gamma, \alpha, l, k) \frac{\binom{L}{l}(-1)^{l+1}}{\left(1+\beta_2(\eta, \gamma, \alpha, l)\right)^k}, \tag{13}$$

$$P_{\text{cov,mf}}^{k,\text{l}}(K) = \sum_{l=1}^{L} \beta_1(1, \gamma, \alpha, l, k) \frac{\binom{L}{l}(-1)^{l+1}}{(1+\beta_2(1, \gamma, \alpha, l))^k}, \tag{14}$$

where

$$\beta_1(\eta,\gamma,\alpha,l,k)=\left[1-\frac{2(\eta\gamma l)^{\frac{2}{\alpha}}}{\alpha}B\left(\frac{2}{\alpha},1-\frac{2}{\alpha},\frac{1}{1+\eta\gamma l}\right)\right]^{k-1}, \tag{15}$$

$$\beta_2(\eta,\gamma,\alpha,l)=2\frac{(\eta\gamma l)^{\frac{2}{\alpha}}}{\alpha}B^{'}\left(\frac{2}{\alpha},1-\frac{2}{\alpha},\frac{1}{1+\eta\gamma l}\right), \tag{16}$$

where $\eta=(L!)^{-\frac{1}{L}}$, $B(x,y,z)\triangleq\int_0^z u^{x-1}(1-u)^{y-1}du$ is the incomplete Beta function and $B^{'}(x,y,z)\triangleq\int_z^1 u^{x-1}(1-u)^{y-1}du$ is the complementary incomplete Beta function.

Proof: See Appendix B. ■

In the special cases with $L=1$ (single antenna), the upper and lower bounds coincide and hence give the exact expression. In addition, if $\alpha=4$, this expression can be written as a closed form.

Corollary 1: The coverage probability of the typical user served by the k-th nearest SBS in the single-antenna network with $\alpha=4$ is given by:

$$P^k_{\text{cov,mf}}(K)=\frac{\left(1-\sqrt{\gamma}\text{arcsin}\frac{1}{\sqrt{1+\gamma}}\right)^{k-1}}{\left(1+\sqrt{\gamma}\text{arccos}\frac{1}{\sqrt{1+\gamma}}\right)^{k}}. \tag{17}$$

In the special case, when the user is served by its nearest SBS, e.g., $k=1$, (17) reduces to the results given in [14, Theorem 2].

B. Coverage Probability with ZF Beamforming

Lemma 3: The coverage probability of the typical user served by the k-th nearest SBS with ZF beamforming is given by:

$$P^k_{\text{cov,zf}}(K)=\mathbb{E}_{r_k,r_K}\left[\sum_{i=0}^{L-K}\frac{(-\gamma {r_k}^{\alpha})^i}{i!}\mathcal{L}^{(i)}_{I_r}(\gamma {r_k}^{\alpha})\right], \tag{18}$$

where $I_r=\sum_{j\in\Phi_b\backslash\mathcal{C}}g_{j,\text{zf}}\cdot r_j^{-\alpha}$ and its Laplace transform is given by:

$$\mathcal{L}_{I_r}(s)=\exp\left(-2\pi\lambda_b\int_{r_K}^{\infty}\frac{sr^{-\alpha}}{1+sr^{-\alpha}}rdr\right). \tag{19}$$

The proof of lemma 3 is similar to Appendix A, so we omit it here. Notice that the expectation in (18) is not only over r_k, but also over r_K. Thus, we need to know the joint probability density function (pdf) of r_k and r_K. The pdf of r_K is given in Lemma 2, we next focus on the pdf of r_k conditioned on r_K.

Lemma 4 ([15]): Consider K points randomly located in a 2-dimensional circle of radius r_K centered at the origin according to the uniform distribution, then the distance R_k from the origin to the k-th nearest point follows the distribution given by:

$$f_{R_k|R_K}(r_k|r_K)=\frac{2r_k}{r_K^2B(K-k,k)}\left(\frac{r_k^2}{r_K^2}\right)^{k-1}\left(1-\frac{r_k^2}{r_K^2}\right)^{K-k-1}, \tag{20}$$

where $B(x,y)$ denotes the Beta function given by $B(x,y)=\int_0^1 t^{x-1}(1-t)^{y-1}dt$.

Thus, the joint pdf of r_k and r_K can be obtained as:

$$\begin{aligned}f_{R_k,R_K}(r_k,r_K)&=f_{R_k|R_K}(r_k|r_K)f_{R_K}(r_K)\\&=\frac{4(\lambda_b\pi)^K}{\Gamma(K-k)\Gamma(k)}r_kr_K(r_k^2)^{k-1}\\&\quad\times(r_K^2-r_k^2)^{K-k-1}\exp\left(-\lambda_b\pi r_K^2\right).\end{aligned} \tag{21}$$

With Lemma 3 and Lemma 4, we can obtain a tractable expression for the coverage probability. More compact forms of the approximate coverage probability bounds are obtained in the following theorem.

Theorem 2: The coverage probability of the typical user served by the k-th nearest SBS with ZF beamforming can be approximately bounded as:

$$P^{k,\text{l}}_{\text{cov,zf}}(K)\lesssim P^{k}_{\text{cov,zf}}(K)\lesssim P^{k,\text{u}}_{\text{cov,zf}}(K), \tag{22}$$

with

$$P^{k,\text{u}}_{\text{cov,zf}}(K)=\sum_{l=1}^{L-K+1}\frac{\binom{L-K+1}{l}(-1)^{l+1}}{\left[1+(\kappa\gamma l)^{\frac{2}{\alpha}}\sqrt{\frac{k}{K}}\mathcal{A}\left(\frac{\sqrt{K}(\kappa\gamma l)^{-\frac{2}{\alpha}}}{\sqrt{k}}\right)\right]^k}, \tag{23}$$

$$P^{k,\text{l}}_{\text{cov,zf}}(K)=\sum_{l=1}^{L-K+1}\frac{\binom{L-K+1}{l}(-1)^{l+1}}{\left[1+(\gamma l)^{\frac{2}{\alpha}}\sqrt{\frac{k}{K}}\mathcal{A}\left(\frac{\sqrt{K}(\gamma l)^{-\frac{2}{\alpha}}}{\sqrt{k}}\right)\right]^k}, \tag{24}$$

where $\mathcal{A}(x)=\int_x^{\infty}\frac{1}{1+u^{\frac{\alpha}{2}}}du$ and $\kappa=(L-K+1!)^{-\frac{1}{L-K+1}}$.

The approximate upper and lower bounds coincide when $L=K$. Furthermore, when $\alpha=4$, they can be written in a closed form.

Corollary 2: The approximate coverage probability of the typical user served by the k-th nearest SBS using ZF beamforming with $L=K$ and $\alpha=4$ is given by:

$$P^k_{\text{cov,zf}}(K)\simeq\frac{1}{\left[1+\sqrt{\frac{k\gamma}{K}}\text{arccot}\left(\frac{K}{k\gamma}\right)\right]^k}. \tag{25}$$

Proof: See Appendix C. ■

When the user is served by the nearest SBS, e.g., $k=1$, (25) reduces to the results given in [9, Eqn (28)].

As we shall demonstrate numerically in Section V-A, the upper bounds (13) and (23) are very tight. Therefore, we shall use them as approximate coverage probabilities for caching optimization.

IV. Caching Optimization

' In this section, we find the optimal cache probabilities to maximize the successful transmission probability. The corre-

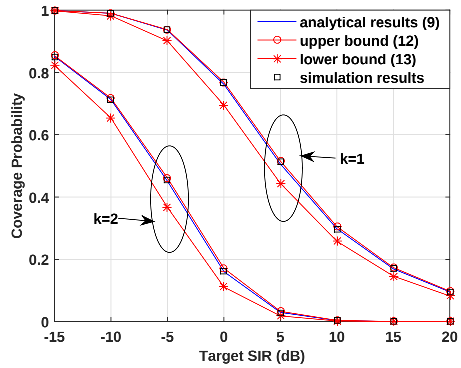

Fig. 1: Coverage probability with $K = 2$ and $L = 2$ in MF beamforming.

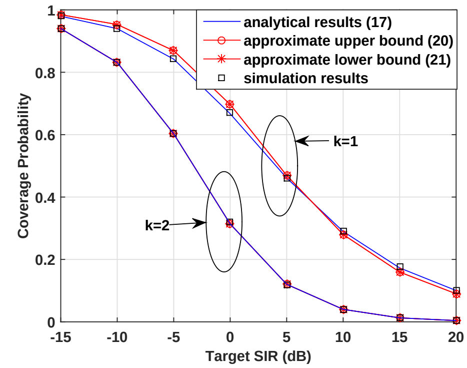

Fig. 2: Coverage probability with $K = 2$ and $L = 2$ in ZF beamforming.

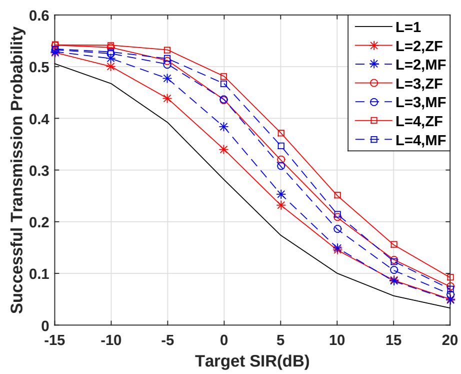

Fig. 3: MF vs ZF with different number of antennas with $K = 2$.

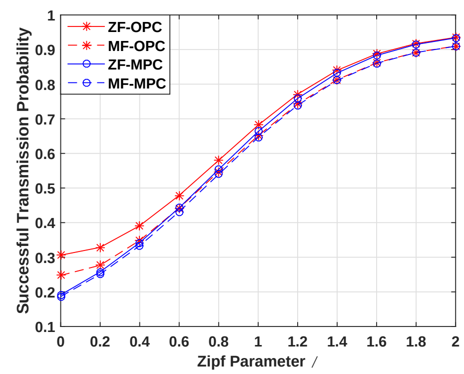

Fig. 4: STP for Different Zipf Parameter with $K = 2$, $L = 4$ and $M = 20$.

sponding problem can be formulated as:

$$\textbf{P1:} \quad \max_{b_n} \quad \sum_{n=1}^{N} p_n \sum_{k=1}^{K} b_n (1-b_n)^{k-1} P_{\text{cov}}(K), \tag{26a}$$

$$\text{s.t} \quad \sum_{n=1}^{N} b_n \leq M, \tag{26b}$$

$$0 \leq b_n \leq 1, \quad n = 1, 2, \ldots, N, \tag{26c}$$

where $P_{\text{cov}}^k(K)$ is given by (13) for MF beamforming and by (23) for ZF beamforming. Since caching more files increases the successful transmission probability, constraint (26b) can be rewritten as:

$$\sum_{n=1}^{N} b_n = M, \tag{27}$$

which will not change the optimal solution.

Lemma 5: The problem **P1** is convex for both MF and ZF beamforming, with the $P_{\text{cov}}^k(K)$ expressed in both the exact forms (9), (18) and the approximate forms (13), (23).

Proof: See Appendix D. ■

By using KKT condition, the optimal solution of **P1** satisfies the condition as follows.

Theorem 3: The optimal cache probabilities of **P1** satisfy

$$b_n(\mu^*) = \min\left(1, w_n(\mu^*)\right), \tag{28}$$

where $\mu^* \geq 0$ is the optimal dual variable satisfying the cache size constraint (27) and $w_n(\mu^*)$ is the real and non-negative root of the function:

$$p_n \sum_{k=1}^{K} [1 - w_n(\mu^*)]^{k-2} [1 - k w_n(\mu^*)] P_{\text{cov}}^k(K) - \mu^* = 0. \tag{29}$$

Proof: See Appendix E. ■

To obtain the optimal cache strategy, we should find the optimal dual variable μ^* by replacing (28) into the cache size constraint $\sum_{n=1}^{N} b_n(\mu^*) = M$. It is observed that $b_n(\mu)$ is a decreasing function respect to μ. Thus, the sum of $b_n(\mu)$ is also decreasing respect to μ. Therefore, we can use the bisection method to find the optimal μ^*.

V. Simulation Results and Discussion

In this section, we first validate the tightness of the approximate coverage probability in both MF and ZF cases by simulations. Next, we analyze the effects of number of antennas, and Zipf parameter on STP with optimal probabilistic caching (OPC) and compare the performance between ZF

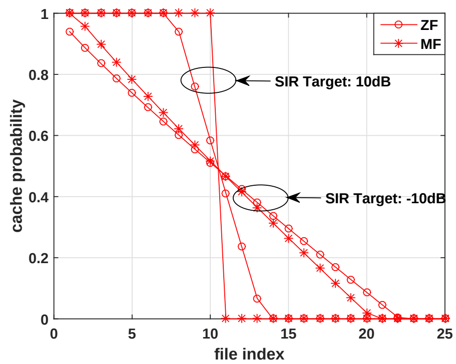

Fig. 5: Cache Probability with $K = 2$ and $L = 4$.

and MF. Besides, comparison with the most popular caching (MPC) is given in numerical results. Finally, the optimal cache probabilities are proposed in both MF and ZF cases.

Simulations are performed in a square area of 4000×4000 m^2. Unless otherwise stated, the simulation parameters are set as follows: SBS intensity $\lambda_b = 5 \times 10^{-5}/\mathrm{m}^2$, path loss exponent $\alpha = 4$, number of total files $N = 100$, cache size $M = 10$, SIR target $\gamma = 0$dB and Zipf parameter $\delta = 0.9$.

A. Validation of Analytical Results

From Fig. 1 and Fig. 2, we observe that the simulation and analytical results of the coverage probability match well for both MF and ZF beamforming. It is also seen that (13) and (23) are very tight. As such, we shall use them to approximate the coverage probabilities of the MF and ZF beamforming in problem $\mathbf{P1}$ to find the optimal cache probabilities.

B. Number of antennas

Fig. 3 shows the successful transmission probability with the optimized caching probabilities at different number of antennas. It is seen that multiple-antenna cache-enabled system has a significant gain compared to the single-antenna system. When the cluster size K is fixed, increasing the number of antennas increases the successful transmission probability for both MF and ZF beamforming but the gain diminishes as L grows. It is also seen that when $L > K$ (e.g. $L = 3, 4$), ZF beamforming always outperforms MF beamforming at all the considered SIR thresholds. Otherwise when $L = K$, MF beamforming performs better than ZF beamforming at most SIR threshold (e.g. $\gamma < 10$dB) and they are almost the same only at high SIR threshold (e.g. $\gamma > 10$dB). This is because when the number of antennas is the same as the cluster size, the effective channel gain of the desired signal with ZF beamforming is much smaller than that of MF beamforming although the former suffers less interference. However, when L is larger than K, the SBS has enough spatial dimensions to null out the intra-cluster interference and strengthen the effective channel gain of the desired signals simultaneously. Therefore, ZF outperforms MF.

C. Zipf Parameter

In Fig. 4, it is observed STP grows with Zipf parameter δ becomes larger. This is because larger δ means few popular files are requested with a larger probability. Besides, it results that the performance of MPC goes closer to that of OPC when δ grows. When $\delta > 1.8$, OPC performs almost identically to MPC. Moreover, it is observed that the gap between ZF and MF in OPC becomes smaller when δ increases. This is because that the coverage probability of the user not served by the nearest SBS in ZF is larger than that in MF. For small δ, the file popularity distribution is more like uniform distribution, which causes that SBSs prefer to cache more number of different files and users have a larger probability to connect to farther SBSs. In this case, ZF can make a better utilization of the multiple SBSs. While for large δ, OPC degenerates to MPC, and thus users have a high probability to connect with the nearest SBS and the performances of ZF and MF become closer.

D. Optimal Cache Strategy

Fig. 5 shows that at the high SIR target with $\gamma = 10$dB, the typical user most likely can only connect to the nearest SBS, and thus the most popular caching strategy is almost optimal. When the SIR target is small with $\gamma = -10$dB, the typical user can be served by farther SBSs. Therefore, caching more number of different files can increase the successful transmission probability. The optimal cache solutions in MF is closer to MPC compared to ZF since the users not served by the nearest SBS in MF case suffers strong interference and thus users are more willing to connect to the nearest SBS.

VI. Conclusions

In this work, we focus on the performance of ZF and MF beamformings in cache-enabled multi-antenna SCNs. Tractable and closed-form approximate expressions of the successful transmission probability in both MF and ZF cases are obtained by using stochastic geometry tools. We then formulate the optimal probabilistic caching problem for maximizing the successful transmission probability, which is proven to be a convex optimization. We solve this problem and obtain the closed-form optimal cache solutions. We numerically analyze the effects of the number of antennas and Zipf parameter on the successful transmission probability and make comparisons between ZF and MF beamformings. Numerical results also reveal that ZF outperforms MF when the number of antennas is larger than the cluster size and MF performs better when the number of antennas is the same as the cluster size.

Appendix A

The coverage probability can be written as:

$$
\begin{aligned}
P^k_{\mathrm{cov,mf}}(K) &= P\left[\frac{g_{k,\mathrm{mf}} \cdot r_k^{-\alpha}}{\sum_{j\in\Phi_b\setminus d_k} g_{j,\mathrm{mf}} \cdot r_j^{-\alpha}} \geq \gamma\right] \\
&= \mathbb{E}_{r_k,I_r}\left[P\left[g_{k,\mathrm{mf}} \geq \gamma r_k^{\alpha} I_r\right] | r_k, I_r\right] \\
&\overset{(a)}{=} \mathbb{E}_{r_k,I_r}\left[\sum_{i=0}^{L-1} \frac{(\gamma r_k^{\alpha} I_r)^i}{i!} e^{-\gamma r_k^{\alpha} I_r} | r_k, I_r\right] \\
&\overset{(b)}{=} \mathbb{E}_{r_k}\left[\sum_{i=0}^{L-1} \frac{(-\gamma {r_k}^{\alpha})^i}{i!} \mathcal{L}^{(i)}_{I_r}(\gamma {r_k}^{\alpha}) | r_k\right], \qquad (30)
\end{aligned}
$$

where (a) follows from the series expansion of the CCDF $\overline{F}(x;m,\theta)$ for Gamma distribution $\Gamma(m,\theta)$ when θ is a positive integer, i.e., $\overline{F}(x;m,\theta)=\sum_{i=0}^{m-1}\frac{1}{i!}(\frac{x}{\theta})^i e^{-\frac{x}{\theta}}$; and (b) follows from the derivative property of the Laplace transform: $\mathbb{E}[X^i e^{-sX}]=(-1)^i\mathcal{L}_X^{(i)}(s)$.

The interference I_r consists of two parts, the interference I_1 from the $k-1$ SBSs closer to u_0 than the serving SBS d_k and the interference I_2 from the SBSs farther than d_k. Therefore, the Laplace transform of interference $\mathcal{L}_{I_r}(s)$ can be given by the product of $\mathcal{L}_{I_1}(s)$ and $\mathcal{L}_{I_2}(s)$, which is under the condition that the distance between the k-th nearest SBS and the typical r_k is given.

The Laplace transform of the I_1 is given by:

$$\begin{aligned}\mathcal{L}_{I_1}(s)&=\mathbb{E}_{\Phi_b,g_{j,\mathrm{mf}}}\left[\prod_{j\in\Phi_b\bigcap\mathcal{B}(0,r_k)\setminus\{d_k\}}\exp\left(-sg_{j,\mathrm{mf}}\cdot r_j^{-\alpha}\right)\right]\\&\overset{(a)}{=}\mathbb{E}_{\Phi_b}\left[\prod_{j\in\Phi_b\bigcap\mathcal{B}(0,r_k)\setminus\{d_k\}}\frac{1}{1+sr_j^{-\alpha}}\right]\\&\overset{(b)}{=}\left(\int_0^{r_k}\frac{1}{1+sr^{-\alpha}}\frac{2r}{r_k^2}dr\right)^{k-1},\end{aligned}\tag{31}$$

where (a) follows from that $g_{j,\mathrm{mf}}$ are iid exponential random variables with parameter 1 and they are independent with the HPPP Φ_b. For the step (b), since the locations of $K-1$ BSs are iid and uniformly distributed in the circle $\mathcal{B}(0,r_k)$. Thus, the pdf of the distance between these BSs and the typical user is given by:

$$f_R(r)=\begin{cases}\frac{2r}{r_k^2}, & 0\le r\le r_k\\ 0, & r_k\le r\end{cases}.\tag{32}$$

For the interference I_2, the Laplace transform can be written as:

$$\begin{aligned}\mathcal{L}_{I_2}(s)&=\mathbb{E}_{\Phi_b}\left[\prod_{j\in\Phi_b\setminus\mathcal{B}(0,r_k)}\frac{1}{1+sr_j^{-\alpha}}\right]\\&=\exp\left(-2\pi\lambda_b\int_{r_k}^{\infty}\frac{sr^{-\alpha}}{1+sr^{-\alpha}}rdr\right),\end{aligned}\tag{33}$$

where the last step follows from the probability generating functional (PGFL) of the HPPP.

Appendix B

The proof relies on the following Lemma 5.

Lemma 6 (Alzer's Inequality [16], [17]): If Z is a random variable following the Gamma distribution with shape parameter T and scale parameter 1, denoted as $Z\sim\Gamma(T,1)$, the cumulative distribution function (CDF) $F_Z(z)=P[Z\le z]$ is bounded by:

$$(1-e^{-\kappa z})^T\le F_Z(z)\le(1-e^{-z})^T,\tag{34}$$

where $\kappa=(T!)^{-\frac{1}{T}}$.

From Lemma 5, it is observed that the upper bound can be obtained by setting $\kappa=1$ if the expression of the lower bound with parameter κ is known.

Thus, we only need to focus on the upper bound of the coverage probability. The upper bound for the CCDF of the Gamma random variable Z can be written as

$$\begin{aligned}\overline{F}_Z(z)=P[Z\ge z]&\ge 1-(1-e^{-\kappa z})^T\\&=\sum_{i=1}^{T}\binom{T}{i}(-1)^{i+1}e^{-\kappa zi}.\end{aligned}\tag{35}$$

Therefore, the upper bound of the coverage probability is given by:

$$\begin{aligned}P_{\mathrm{cov,mf}}^k(K)&=\mathbb{E}_{r_k,I_r}\left[P\left[g_{k,\mathrm{mf}}\ge\gamma r_k^\alpha I_r\right]|r_k,I_r\right]\\&\le\sum_{l=1}^{L}\binom{L}{l}(-1)^{l+1}\mathbb{E}_{r_k,I_r}\left[e^{-\eta\gamma r_k^\alpha I_r l}|r_k,I_r\right]\\&=\sum_{l=1}^{L}\binom{L}{l}(-1)^{l+1}\mathbb{E}_{r_k}\left[\mathcal{L}_{I_r}(\eta\gamma r_k^\alpha l)|r_k\right].\end{aligned}\tag{36}$$

where $\eta=(L!)^{-\frac{1}{L}}$.

Then, before simplifying the expression of the coverage probability upper bound, we make some mathematical transforms of the Laplace transform of the interference I_r.

$$\begin{aligned}\mathcal{L}_{I_r}(s)&=\left(\int_0^{r_k}\frac{1}{1+sr^{-\alpha}}\frac{2r}{r_k^2}dr\right)^{k-1}\\&\quad\times\exp\left(-2\pi\lambda_b\int_{r_k}^{\infty}\frac{sr^{-\alpha}}{1+sr^{-\alpha}}rdr\right)\\&=\left[1-\frac{2s^{2/\alpha}}{\alpha r_k^2}B\left(\frac{2}{\alpha},1-\frac{2}{\alpha},\frac{1}{1+sr_k^{-\alpha}}\right)\right]^{k-1}\\&\quad\times\exp\left[-2\pi\lambda_b\frac{s^{\frac{2}{\alpha}}}{\alpha}B^{'}\left(\frac{2}{\alpha},1-\frac{2}{\alpha},\frac{1}{1+sr_k^{-\alpha}}\right)\right],\end{aligned}\tag{37}$$

where the last step follows that first replacing $s^{-\frac{1}{\alpha}}r$ with u, then replacing $\frac{1}{1+u^{-\alpha}}$ with v.

Therefore, the Laplace transform in (36) can be written as:

$$\begin{aligned}\mathcal{L}_{I_r}(\eta\gamma r_k^\alpha l)&=\left[1-\frac{2(\eta\gamma l)^{2/\alpha}}{\alpha}B\left(\frac{2}{\alpha},1-\frac{2}{\alpha},\frac{1}{1+\eta\gamma l}\right)\right]^{k-1}\\&\quad\times\exp\left[-2\pi\lambda_b\frac{r_k^2(\eta\gamma l)^{\frac{2}{\alpha}}}{\alpha}B^{'}\left(\frac{2}{\alpha},1-\frac{2}{\alpha},\frac{1}{1+\eta\gamma l}\right)\right]\\&=\beta_1(\eta,\gamma,\alpha,l,k)\exp\left(-\pi\lambda_b r_k^2\beta_2\left(\eta,\gamma,\alpha,l\right)\right),\end{aligned}\tag{38}$$

where $\beta_1(\eta,\gamma,\alpha,l)$ and $\exp(-\pi\lambda_b r_k^2\beta_2(\eta,\gamma,\alpha,l))$ are defined for expression simplicity and they are the Laplace transform of intra-cluster and inter-cluster interference, respectively. $\beta_1(\eta,\gamma,\alpha,l)$ and $\beta_2(\eta,\gamma,\alpha,l)$ are expressed as:

$$\beta_1(\eta,\gamma,\alpha,l,k)=\left[1-\frac{2(\eta\gamma l)^{\frac{2}{\alpha}}}{\alpha}B\left(\frac{2}{\alpha},1-\frac{2}{\alpha},\frac{1}{1+\eta\gamma l}\right)\right]^{k-1},\tag{39}$$

$$\beta_2(\eta,\gamma,\alpha,l)=2\frac{(\eta\gamma l)^{\frac{2}{\alpha}}}{\alpha}B^{'}\left(\frac{2}{\alpha},1-\frac{2}{\alpha},\frac{1}{1+\eta\gamma l}\right),\tag{40}$$

Then we need to calculate the expectation of $\mathcal{L}_{I_r}(\eta\gamma r_k^\alpha l)$ over r_k. It is observed that only the inter-cluster interference is related to r_k. Therefore, evaluating the expectation of $\exp(-\pi\lambda_b r_k^2 \beta_2(\eta,\gamma,\alpha,l))$ is enough.

$$\begin{aligned}
&\mathbb{E}_{r_k}\left[\exp\left(-\pi\lambda_b r_k^2\beta_2(\eta,\gamma,\alpha,l)\right)\right] \\
&= \int_0^\infty \exp(-\pi\lambda_b r_k^2\beta_2(\eta,\gamma,\alpha,l))\frac{2(\lambda_b\pi r_k^2)^k}{r_k\Gamma(k)}\exp(-\lambda_b\pi r_k^2)dr_k \\
&\overset{(a)}{=} \int_0^\infty \left[\frac{z}{1+\beta_2(\eta,\gamma,\alpha,l)}\right]^{k-1} \times \frac{e^{-z}}{\Gamma(k)\left(1+\beta_2(\eta,\gamma,\alpha,l)\right)}dz \\
&\overset{(b)}{=} \left[\frac{1}{1+\beta_2(\eta,\gamma,\alpha,l)}\right]^k,
\end{aligned} \tag{41}$$

where step (a) follows from the random variable change

$$z = \pi\lambda_b r_k^2(1+\beta_2(\eta,\gamma,\alpha,l)), \tag{42}$$

and step (b) follows from the Gamma distribution property

$$\int_0^\infty t^k e^{-\lambda t}dt = \frac{k!}{\lambda^{k+1}}. \tag{43}$$

Therefore, we can obtain the upper bound of the coverage probability given by:

$$\begin{aligned}
P_{\text{cov,mf}}^{k,\text{u}}(K) &= \sum_{l=1}^{L}\binom{L}{l}(-1)^{l+1}\mathbb{E}_{r_k}[\mathcal{L}_{I_r}(\eta\gamma r_k^\alpha l)|r_k] \\
&= \sum_{l=1}^{L}\beta_1(\eta,\gamma,\alpha,l,k)\frac{\binom{L}{l}(-1)^{l+1}}{[1+\beta_2(\eta,\gamma,\alpha,l)]^k}.
\end{aligned} \tag{44}$$

By setting $\eta = 1$, the lower bound can be obtained directly, which completes the proof.

Appendix C

With the similar methods used in Appendix B, we can obtain the upper bound of the coverage probability in the ZF beamforming case.

$$\begin{aligned}
P_{\text{cov,zf}}^{k}(K) \le &\sum_{l=1}^{L-K+1}\binom{L-K+1}{l}(-1)^{l+1} \\
&\times \mathbb{E}_{r_k,r_K}[\mathcal{L}_{I_r}(\kappa\gamma r_k^\alpha l)|r_k,r_K],
\end{aligned} \tag{45}$$

where the expectation is over r_k and r_K because the distance of the k-th nearest and the K-th nearest BSs are both random variables. I_r is the inter-cluster interference and its Laplace transform is similar to the $\mathcal{L}_{I_2}(s)$ in the MF beamforming case. The only difference is to replace r_k in (33) with r_K.

$$\mathcal{L}_{I_r}(s) = \exp\left[-2\pi\lambda_b\int_{r_K}^\infty \frac{sr^{-\alpha}}{1+sr^{-\alpha}}rdr\right]. \tag{46}$$

Replace s with $\kappa\gamma r_k^\alpha l$ and introduce a geometric parameter $\delta_k = \frac{r_k}{r_K}$, the Laplace transform can be written as:

$$\begin{aligned}
\mathcal{L}_{I_r}(\kappa\gamma r_k^\alpha l) &= \exp\left(-2\pi\lambda_b\int_{r_K}^\infty \frac{r^{-\alpha}\kappa\gamma r_k^\alpha l}{1+r^{-\alpha}\kappa\gamma r_k^\alpha l}rdr\right) \\
&= \exp\left(-2\pi\lambda_b\int_{r_K}^\infty \frac{r}{1+(\frac{r}{r_K})^\alpha(\kappa\gamma\delta_k^\alpha l)^{-1}}dr\right) \\
&= \exp\left(-\pi\lambda_b r_K^2(\kappa\gamma\delta_k^\alpha l)^{\frac{2}{\alpha}}\int_{(\kappa\gamma\delta_k^\alpha l)^{-\frac{2}{\alpha}}}^\infty \frac{1}{1+v^{\frac{\alpha}{2}}}dv\right),
\end{aligned} \tag{47}$$

where the last step follows from the variable change

$$v = \left[\left(\frac{1}{\kappa\gamma\delta_k^\alpha l}\right)^{\frac{1}{\alpha}}\frac{r}{r_K}\right]^2. \tag{48}$$

For ease of illustration, we let

$$\beta_3(\kappa\gamma\delta_k^\alpha l,\alpha) = (\kappa\gamma\delta_k^\alpha l)^{\frac{2}{\alpha}}\int_{(\kappa\gamma\delta_k^\alpha l)^{-\frac{2}{\alpha}}}^\infty \frac{1}{1+v^{\frac{\alpha}{2}}}dv. \tag{49}$$

From (45) and (47), we find that we need to calculate the expectation over δ_k and r_K, rather than over r_k and r_K. Thus, we first calculate the expectation of (47) over r_K.

$$\begin{aligned}
&\mathbb{E}_{r_K}\left[\mathcal{L}_{I_r}\left(\kappa\gamma(\delta_k r_K)^\alpha l\right)|\delta_k,r_K\right] \\
&\quad= \int_0^\infty \exp(-\pi\lambda_b r_K^2\beta_3(\kappa\gamma\delta_k^\alpha l,\alpha)) \\
&\quad\quad\times \frac{2(\lambda_b\pi r_K^2)^K}{r_K\Gamma(K)}\exp(-\lambda_b\pi r_K^2)dr_K \\
&\quad= \frac{1}{[1+\beta_3(\kappa\gamma\delta_k^\alpha l,\alpha)]^K},
\end{aligned} \tag{50}$$

where the last step is similar to the (41).

Therefore, the upper bound and lower bound can be expressed as

$$P_{\text{cov,zf}}^{k,\text{u}}(K) = \mathbb{E}_{\delta_k}\left[\sum_{l=1}^{L-K+1}\frac{\binom{L-K+1}{l}(-1)^{l+1}}{[1+\beta_3(\kappa\gamma\delta_k^\alpha l,\alpha)]^K}\right], \tag{51}$$

$$P_{\text{cov,zf}}^{k,\text{l}}(K) = \mathbb{E}_{\delta_k}\left[\sum_{l=1}^{L-K+1}\frac{\binom{L-K+1}{l}(-1)^{l+1}}{[1+\beta_3(\gamma\delta_k^\alpha l,\alpha)]^K}\right]. \tag{52}$$

To obtain the expectation above over δ_k, we first need to know the pdf of δ_k. Utilizing the joint pdf of r_k and r_K given in (21), we can derive the CDF of $\delta_k = \frac{r_k}{r_K}$ as

$$\begin{aligned}
P[\delta_k \le x] &= P[r_k \le xr_K] \\
&= \int_0^\infty\int_0^{xr_K} f_{R_k,R_K}(r_k,r_K)dr_k dr_K \\
&= \int_0^\infty\int_0^{xr_K} \frac{4r_k r_k^{2(k-1)}r_K}{\Gamma(K-k)\Gamma(k)}(\lambda_b\pi)^K \\
&\quad\times (r_K^2-r_k^2)^{K-k-1}\exp(-\lambda_b\pi r_K^2)dr_k dr_K \\
&= 1-\sum_{i=0}^{k-1}\frac{(K-1)!x^{2(k-1-i)}(1-x^2)^{K-k+i}}{(K-k+i)!(k-1-i)!},
\end{aligned} \tag{53}$$

where $0 \le x \le 1$, and for the last step follows from the random variable change as $u = r_K^2$, $v = r_k^2$ and partial integration. Then the pdf of δ_k can be obtained by calculating the derivative of (53) respect to x.

$$\begin{aligned} f_{\delta_k}(x) &= \frac{dP[\delta_k \le x]}{dx} \\ &= \sum_{i=0}^{k-1} \frac{(K-1)!\left[(K-1)x^2-(k-i-1)\right]}{(K-k+i)!(k-1-i)!} \\ &\quad \times 2x^{2(k-1-i)-1}(1-x^2)^{K-k+i-1} \\ &= \frac{2(K-1)!}{(k-1)!(K-k-1)!} x^{2k-1}(1-x^2)^{K-k-1}. \end{aligned} \tag{54}$$

Recall (49), we approximate the integral in it as a constant value according to randomness of δ_k, which is given by:

$$\mathbb{E}\left[\int_{\delta_k^{-2}(\kappa\gamma l)^{-\frac{2}{\alpha}}}^{\infty} \frac{1}{1+v^{\frac{\alpha}{2}}} dv\right] \simeq \sqrt{\frac{k}{K}}\mathcal{A}\left(\frac{\sqrt{K}(\kappa\gamma l)^{-\frac{2}{\alpha}}}{\sqrt{k}}\right), \tag{55}$$

which follows from the expectation of δ_k^2 and is given by

$$\begin{aligned} \mathbb{E}(\delta_k^2) &= \int_0^1 x^2 f_{\delta_k}(x) dx \\ &= \int_0^1 \frac{2x^2(K-1)!}{(k-1)!(K-k-1)!} x^{2k-1}(1-x^2)^{K-k-1} dx \\ &= \frac{k}{K}. \end{aligned} \tag{56}$$

Thus,we can approximate $\beta_3(\kappa\gamma\delta_k^\alpha l, \alpha)$ as:

$$\beta_3(\kappa\gamma\delta_k^\alpha l, \alpha) \simeq \delta_k^2 (\kappa\gamma l)^{\frac{2}{\alpha}} \sqrt{\frac{k}{K}}\mathcal{A}\left(\frac{\sqrt{K}(\kappa\gamma l)^{-\frac{2}{\alpha}}}{\sqrt{k}}\right). \tag{57}$$

Then, the expectation of $\frac{1}{\left[1+\beta_3(\kappa\gamma\delta_k^\alpha l,\alpha)\right]^K}$ can be approximated as:

$$\begin{aligned} &\mathbb{E}_{\delta_k}\left[\left(\frac{1}{1+\beta_3(\kappa\gamma\delta_k^\alpha l, \alpha)}\right)^K\right] \\ &= \int_0^1 \left[\frac{1}{1+\beta_3(\kappa\gamma\delta_k^\alpha l, \alpha)}\right]^K f_{\delta_k}(x) dx \\ &\simeq \frac{2(K-1)!}{(k-1)!(K-k-1)!} \\ &\quad \times \int_0^1 \frac{x^{2k-1}(1-x^2)^{K-k-1}}{\left[1+(\kappa\gamma l)^{\frac{2}{\alpha}}\sqrt{\frac{k}{K}}\mathcal{A}\left(\frac{\sqrt{K}(\kappa\gamma l)^{-\frac{2}{\alpha}}}{\sqrt{k}}\right)x^2\right]^K} dx \\ &= \frac{1}{\left[1+(\kappa\gamma l)^{\frac{2}{\alpha}}\sqrt{\frac{k}{K}}\mathcal{A}\left(\frac{\sqrt{K}(\kappa\gamma l)^{-\frac{2}{\alpha}}}{\sqrt{k}}\right)\right]^k}, \end{aligned} \tag{58}$$

where the last step also follows from the random variable change as $u = x^2$ and partial integration, similar to (53). Then we plug this into (51) and (52), the proof is completed.

Appendix D

Before proving the optimization problem $\mathbf{P1}$ is convex, we first need to prove that the coverage probability $P_{\text{cov}}^k(K)$ is a non-increasing function respect to k. It is easy to understand this since the SIR for the user served by the farther SBS is lower, which causes the lower coverage probability. Then we prove this conclusion mathematically in both ZF and MF beamforming cases.

We first focus on the analytical results (9) and (18). In the ZF case, the Laplace transform of the interference I_r is given by:

$$\mathcal{L}_{I_r}(s) = \exp\left(-2\pi\lambda_b \int_{r_K}^{\infty} \frac{sr^{-\alpha}}{1+sr^{-\alpha}} r dr\right), \tag{59}$$

For ease of illustration, we let

$$X(s) = -2\pi\lambda_b \int_{r_K}^{\infty} \frac{sr^{-\alpha}}{1+sr^{-\alpha}} r dr. \tag{60}$$

Thus, the i-th order derivative of the Laplace transform can be written as

$$\mathcal{L}_{I_r}^{(i)}(s) = \sum_{m=0}^{i-1} \binom{i-1}{m} \mathcal{L}_{I_r}^{(m)}(s) X^{(i-m)}(s). \tag{61}$$

It is easy to find the even order derivative of $X(s)$ is positive while the odd order derivative is negative. Besides, it is observed that

$$\mathcal{L}_{I_r}^{(1)}(s) = \mathcal{L}_{I_r}(s) X^{(1)}(s) \le 0, \tag{62}$$

$$\mathcal{L}_{I_r}^{(2)}(s) = \mathcal{L}_{I_r}(s) X^{(2)}(s) + \mathcal{L}_{I_r}^{(1)}(s) X^{(1)}(s) \ge 0. \tag{63}$$

Therefore, we then we use mathematical induction to prove that the even order derivative of $\mathcal{L}_{I_r}(s)$ is positive and the odd order derivative is negative. First, we initialize $k = 1$. From (62) and (63), we have $\mathcal{L}_{I_r}^{(2k)}(s) \ge 0$ and $\mathcal{L}_{I_r}^{(2k-1)}(s) \le 0$. Then, we assume $\mathcal{L}_{I_r}^{(2k)}(s) \ge 0$ and $\mathcal{L}_{I_r}^{(2k-1)}(s) \le 0$ for any integer $k = 2, 3, \cdots n$. Next, we focus on $k = n+1$ and have:

$$\begin{aligned} \mathcal{L}_{I_r}^{(2k)}(s) &= \mathcal{L}_{I_r}^{(2n+2)}(s) \\ &= \sum_{m=0}^{2n+1} \binom{2n+1}{m} \mathcal{L}_{I_r}^{(m)}(s) X^{(2n+2-m)}(s) \ge 0 \end{aligned} \tag{64}$$

$$\begin{aligned} \mathcal{L}_{I_r}^{(2k-1)}(s) &= \mathcal{L}_{I_r}^{(2n+1)}(s) \\ &= \sum_{m=0}^{2n} \binom{2n}{m} \mathcal{L}_{I_r}^{(m)}(s) X^{(2n+1-m)}(s) \le 0 \end{aligned} \tag{65}$$

Therefore, we conclude that $\mathcal{L}_{I_r}^{(2k)}(s) \ge 0$ and $\mathcal{L}_{I_r}^{(2k-1)}(s) \le 0$ for any integer k. Then we define a function $G(s)$ as

$$G(s) = \sum_{i=0}^{L-K} \frac{(-s)^i}{i!} \mathcal{L}_{I_r}^{(i)}(s) \tag{66}$$

and we have

$$G'(s) = \sum_{i=0}^{L-K} \frac{(-1)^i}{i!} [is^{i-1}\mathcal{L}_{I_r}^{(i)}(s) + s^i \mathcal{L}_{I_r}^{(i+1)}(s)] \tag{67}$$

$$= (-1)^{L-K} s^{L-K} \mathcal{L}_{I_r}^{(L-K+1)}(s) \le 0 \tag{68}$$

which means $G(s)$ is non-increasing respect to s. Hence, $G(\gamma r_k^\alpha)$ is a non-increasing function respect to r_k. Since the coverage probability is given by:

$$\begin{aligned}P_{\text{cov,zf}}^k(K) &= \mathbb{E}_{r_k,r_K}[G(\gamma r_k^\alpha)] \\ &= \int_0^\infty \int_0^{r_K} G(\gamma r_k^\alpha) f_{R_k,R_K}(r_k,r_K) dr_k dr_K.\end{aligned} \tag{69}$$

Then, we define a function $A(k)$ represented the conditional pdf of r_k given in lemma 4 as: $A(k) = f_{R_k|R_K}(r_k|r_K)$. We assume that the distance of the k-th nearest SBS and the $k+1$ nearest SBS are the same ($r_k = r_{k+1}$). Thus, we have

$$\frac{A(k+1)}{A(k)} = \left(\frac{K-k-1}{k}\right)\left(\frac{r_k^2}{r_K^2 - r_k^2}\right) \tag{70}$$

which is a increasing function respect to r_k. When $r_k = 0$, the ratio equals to 0 and when $r_k = r_K$, we have $\frac{A(n+1)}{A(n)} = \infty$. Therefore, there must exist a threshold D. When $r_k = r_{k+1} = D$, we have $\frac{A(k+1)}{A(k)} = 1$. This means that the probability of $r_k = D$ and $r_{k+1} = D$ are the same. For the distance $D_1 < D$, we have $P[r_k = D_1] > P[r_{k+1} = D_1]$ since $\frac{A(k+1)}{A(k)}$ is a increasing function respect to r_k. For the distance $D_2 > D$, we have $P[r_k = D_2] < P[r_{k+1} = D_2]$. Combining this with $G(\gamma r_k^\alpha)$ is non-increasing respect to r_k, we can conclude that the coverage probability $P_{\text{cov,zf}}^k(K)$ is a non-increasing function respect to k. For the MF beamforming case, the coverage probability $P_{\text{cov,mf}}^k(K)$ is also a non-increasing function respect to k which can be proved easily.

While for the approximate coverage probability (13) and (23), we try to proof they are also non-increasing function respect to k.

In the ZF case, we define a non-negative random variable I_3 similar to the interference I_r and its Laplace transform is given by:

$$\mathcal{L}_{I_3}(s) = \exp\left(-2\pi\lambda_b \int_{r_k}^\infty \frac{sr^{-\alpha}}{1+sr^{-\alpha}} r dr\right). \tag{71}$$

Similar to (47), we have:

$$\begin{aligned}&\mathcal{L}_{I_3}\left(\kappa\gamma(\delta_k r_K)^\alpha l \left(\frac{k}{K}\right)^{\frac{\alpha}{4}}\right) \\ &= \exp\left(-\pi\lambda_b r_K^2 \delta_k^2 (\kappa\gamma l)^{\frac{2}{\alpha}} \sqrt{\frac{k}{K}} \mathcal{A}\left(\frac{\sqrt{K}(\kappa\gamma l)^{-\frac{2}{\alpha}}}{\sqrt{k}}\right)\right) \\ &= \exp\left(-\pi\lambda_b r_k^2 (\kappa\gamma l)^{\frac{2}{\alpha}} \sqrt{\frac{k}{K}} \mathcal{A}\left(\frac{\sqrt{K}(\kappa\gamma l)^{-\frac{2}{\alpha}}}{\sqrt{k}}\right)\right),\end{aligned} \tag{72}$$

Therefore, the approximate coverage probability in ZF case can be written as:

$$\begin{aligned}P_{\text{cov,zf}}^{k,\text{u}}(K) &= \sum_{l=1}^{L-K+1} \binom{L-K+1}{l} (-1)^{l+1} \\ &\quad \times \mathbb{E}_{r_k}\left[\mathcal{L}_{I_3}\left(\kappa\gamma r_k^\alpha l \left(\frac{k}{K}\right)^{\frac{\alpha}{4}}\right) \Big| r_k\right] \\ &= \mathbb{E}_{r_k,I_3}\Bigg[1 - \Bigg[1 - \exp\Bigg(-I_3\kappa\gamma r_k^\alpha \\ &\quad \times \left(\frac{k}{K}\right)^{\frac{\alpha}{4}}\Bigg)\Bigg]^{L-K+1} \Bigg| r_k, I_3\Bigg],\end{aligned} \tag{73}$$

which is a non-increasing function respect to k. Hence, we conclude that:

$$P_{\text{cov,zf}}^{k,\text{u}}(K) \geq \sum_{l=1}^{L-K+1} \frac{\binom{L-K+1}{l}(-1)^{l+1}}{\left[1 + (\kappa\gamma l)^{\frac{2}{\alpha}} \sqrt{\frac{k+1}{K}} \mathcal{A}\left(\frac{\sqrt{K}(\kappa\gamma l)^{-\frac{2}{\alpha}}}{\sqrt{k+1}}\right)\right]^k}. \tag{74}$$

Then, we define a non-negative random variable I_4 similarly and its Laplace transform is given by:

$$\begin{aligned}\mathcal{L}_{I_4}(s) &= \exp\left(-2\pi\lambda_b \int_{r_k}^\infty \frac{sr^{-\alpha}}{1+sr^{-\alpha}} r dr\right) \\ &\quad \times \int_{r_k}^\infty \frac{sr^{-\alpha}}{1+sr^{-\alpha}} \frac{2r}{r_k^2} dr \Big/ \left(1 + \int_{r_k}^\infty \frac{sr^{-\alpha}}{1+sr^{-\alpha}} \frac{2r}{r_k^2} dr\right).\end{aligned} \tag{75}$$

Hence, we have:

$$\begin{aligned}&\sum_{l=1}^{L-K+1} \frac{\binom{L-K+1}{l}(-1)^{l+1}}{\left[1 + (\kappa\gamma l)^{\frac{2}{\alpha}} \sqrt{\frac{k+1}{K}} \mathcal{A}\left(\frac{\sqrt{K}(\kappa\gamma l)^{-\frac{2}{\alpha}}}{\sqrt{k+1}}\right)\right]^k} - P_{\text{cov,zf}}^{k+1,\text{u}}(K) \\ &= \mathbb{E}_{r_k,I_4}\left[1 - \left(1 - e^{-I_4\kappa\gamma r_k^\alpha \left(\frac{k+1}{K}\right)^{\frac{\alpha}{4}}}\right)^{L-K+1}\right] \geq 0,\end{aligned} \tag{76}$$

Combining (74) and (76), we conclude that $P_{\text{cov,zf}}^{k,\text{u}}(K) \geq P_{\text{cov,zf}}^{k+1,\text{u}}(K)$, which means $P_{\text{cov,zf}}^{k,\text{u}}(K)$ is a non-increasing function respect to k.

For the MF case, we define a random variable I_5 similarly and its Laplace transform is given by:

$$\begin{aligned}\mathcal{L}_{I_5}(s) &= \left(\int_0^{r_k} \frac{1}{1+sr^{-\alpha}} \frac{2r}{r_k^2} dr\right)^{k-1} \exp\left(-2\pi\lambda_b \int_{r_k}^\infty \frac{sr^{-\alpha}}{1+sr^{-\alpha}} r dr\right) \\ &\quad \times \int_0^\infty \frac{sr^{-\alpha}}{1+sr^{-\alpha}} \frac{2r}{r_k^2} dr \Big/ \left(1 + \int_{r_k}^\infty \frac{sr^{-\alpha}}{1+sr^{-\alpha}} \frac{2r}{r_k^2} dr\right).\end{aligned} \tag{77}$$

Hence, we have

$$\begin{aligned}
&P^{k,\mathrm{u}}_{\mathrm{cov,mf}}(K) - P^{k+1,\mathrm{u}}_{\mathrm{cov,mf}}(K) \\
&= \sum_{l=1}^{L} \beta_1(\eta,\gamma,\alpha,l,k) \frac{\binom{L}{l}(-1)^{l+1}}{[1+\beta_2(\eta,\gamma,\alpha,l)]^k} \\
&\quad \times \frac{1-\beta_1(\eta,\gamma,\alpha,l,2)+\beta_2(\eta,\gamma,\alpha,l)}{1+\beta_2(\eta,\gamma,\alpha,l)} \\
&= \mathbb{E}_{r_k,I_5}\left[1-\left(1-e^{-I_5\eta\gamma r_k^\alpha}\right)^L\right] \geq 0, \qquad (78)
\end{aligned}$$

which means $P^{k,\mathrm{u}}_{\mathrm{cov,mf}}(K)$ is a non-increasing function respect to k.

Therefore, we conclude that the coverage probability $P^k_{cov}(K)$ is a non-increasing function respect to k for exact and approximate forms in both ZF and MF beamforming cases.

Utilizing the property of the coverage probability proved above, the second order derivative of the objective function respect to b_n can be expressed as

$$\begin{aligned}
&\frac{\partial^2 P_{\mathrm{suc}}(K)}{\partial b_n^2} \\
&= \sum_{n=1}^{N} p_n \sum_{k=1}^{K} (k-1)(1-b_n)^{k-3}(kb_n-2)P^k_{\mathrm{cov}}(K) \\
&= \sum_{n=1}^{N} p_n \Bigg[-2P^2_{\mathrm{cov}}(K) + 2(3b_n-2)P^3_{\mathrm{cov}}(K) \\
&\quad + \sum_{k=4}^{K} (k-1)(1-b_n)^{k-3}(kb_n-2)P^k_{\mathrm{cov}}(K) \Bigg] \\
&\leq \sum_{n=1}^{N} p_n \Bigg[6(b_n-1)P^3_{\mathrm{cov}}(K) + 3(1-b_n)(4b_n-2)P^4_{\mathrm{cov}}(K) \\
&\quad + \sum_{k=5}^{K} (k-1)(1-b_n)^{k-3}(kb_n-2)P^k_{\mathrm{cov}}(K) \Bigg] \\
&\cdots\cdots \\
&\leq K(K-1)(1-b_n)^{K-3}(b_n-1)P^K_{\mathrm{cov}}(K) \\
&\leq 0. \qquad (79)
\end{aligned}$$

where the first inequality follows from that $P^k_{\mathrm{cov}}(K)$ is non-increasing respect to k and the last step follows from that $0 \leq b_n \leq 1$. Thus, the objective function is a concave function and we want to maximize it. Besides, all constraints are linear. Therefore, the proof is completed.

Appendix E

The Lagrangian function can be written as:

$$\begin{aligned}
L(b_1, b_2, \cdots, b_N, \mu) &= \sum_{n=1}^{N} p_n \sum_{k=1}^{K} b_n(1-b_n)^{k-1} P^k_{\mathrm{cov}}(K) \\
&\quad + \mu\left(M - \sum_{n=1}^{N} b_n\right), \qquad (80)
\end{aligned}$$

where u is the Lagrangian multiplier associated with the constraint (27). The partial derivative of the Lagrangian function as:

$$\begin{aligned}
\mathcal{L} &= \frac{\partial L(b_1, b_2, \cdots, b_N, \mu)}{\partial b_n} \\
&= p_n \sum_{k=1}^{K} (1-b_n)^{k-2}(1-kb_n)P^k_{\mathrm{cov}}(K) - \mu. \qquad (81)
\end{aligned}$$

By letting $\mathcal{L} = 0$, we have

$$p_n \sum_{k=1}^{K} (1-b_n)^{k-2}(1-kb_n)P^k_{\mathrm{cov}}(K) = \mu \qquad (82)$$

It is easy to find that the left hand of (82) is a decreasing function respect to b_n since the objective function is concave. Notice that we have the constraint $0 \leq b_n \leq 1$. Thus, when $b_n = 1$, μ has the minimum value: $p_n\left[P^1_{\mathrm{cov}}(K) - P^2_{\mathrm{cov}}(K)\right]$. While for $b_n = 0$, it has the maximum value: $p_n \sum_{k=1}^{K} P^k_{\mathrm{cov}}(K)$. Therefore, for a given Lagrangian multiplier μ, combining this constraint and the monotonic decreasing property of the equation (82), we find that the solution of cache strategy $b_n(\mu)$ can be given by

$$b_n(\mu) = \begin{cases} 1, & \mu \leq p_n\left[P^1_{\mathrm{cov}}(K) - P^2_{\mathrm{cov}}(K)\right] \\ w_n(\mu), & \text{otherwise} \\ 0, & \mu \geq p_n \sum_{k=1}^{K} P^k_{\mathrm{cov}}(K) \end{cases}, \qquad (83)$$

which can be rewritten as:

$$b_n(\mu) = \min\left(1, w_n(\mu)\right), \qquad (84)$$

where $w_n(\mu)$ is the real and non-negative root of the function:

$$p_n \sum_{k=1}^{K} \left[1-w_n(\mu)\right]^{k-2}\left[1-kw_n(\mu)\right] P^k_{\mathrm{cov}}(K) = \mu. \qquad (85)$$

For the optimal dual variable μ^* which satisfies the cache size constraint (27), we obtain the optimal cache strategy is given by:

$$b_n(\mu^*) = \min\left(1, w_n(\mu^*)\right), \qquad (86)$$

which completes the proof.

References

[1] E. Bastug, M. Bennis, M. Kountouris, and M. Debbah, "Cache-enabled small cell networks: Modeling and tradeoffs," *EURASIP J. on Wireless Commun. and Netw.*, vol. 2015, no. 1, pp. 1–11, 2015.

[2] C. Yang, Y. Yao, Z. Chen, and B. Xia, "Analysis on cache-enabled wireless heterogeneous networks," *IEEE Trans. Wireless Commun.*, vol. 15, no. 1, pp. 131–145, 2016.

[3] B. Blaszczyszyn and A. Giovanidis, "Optimal geographic caching in cellular networks," in *Proc. IEEE ICC, 2015*.

[4] Y. Chen, M. Ding, J. Li, Z. Lin, G. Mao, and L. Hanzo, "Probabilistic small-cell caching: Performance analysis and optimization," *IEEE Trans. Veh. Technol.*, vol. PP, no. 99, 2016.

[5] S. H. Chae and W. Choi, "Caching placement in stochastic wireless caching helper networks: Channel selection diversity via caching," *IEEE Trans. Wireless Commun.*, vol. 15, no. 10, pp. 6626–6637, Oct 2016.

[6] K. Li, C. Yang, Z. Chen, and M. Tao, "Optimization and analysis of probabilistic caching in N-tier heterogeneous networks," *arXiv preprint arXiv:1612.04030*, 2016.

[7] 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. PP, no. 99, 2017.

[8] D. Liu and C. Yang, "Caching policy toward maximal success probability and area spectral efficiency of cache-enabled HetNets," *IEEE Trans. Commun.*, 2017.

[9] N. Lee, D. Morales-Jimenez, A. Lozano, and R. W. Heath, "Spectral efficiency of dynamic coordinated beamforming: A stochastic geometry approach," *IEEE Trans Wireless Commun.*, vol. 14, no. 1, pp. 230–241, 2015.

[10] N. Jindal, J. G. Andrews, and S. Weber, "Multi-antenna communication in ad hoc networks: Achieving MIMO gains with SIMO transmission," *IEEE Trans. Commun.*, vol. 59, no. 2, pp. 529–540, 2011.

[11] M. Di Renzo and W. Lu, "Stochastic geometry modeling and performance evaluation of MIMO cellular networks using the equivalent-in-distribution (EiD)-based approach," *IEEE Trans. Commun.*, vol. 63, no. 3, pp. 977–996, 2015.

[12] C. Li, J. Zhang, M. Haenggi, and K. B. Letaief, "User-centric intercell interference nulling for downlink small cell networks," *IEEE Trans. Commun.*, vol. 63, no. 4, pp. 1419–1431, 2015.

[13] M. Haenggi, "On distances in uniformly random networks," *IEEE Transactions on Information Theory*, vol. 51, no. 10, pp. 3584–3586, 2005.

[14] J. G. Andrews, F. Baccelli, and R. K. Ganti, "A tractable approach to coverage and rate in cellular networks," *IEEE Trans. Commun.*, vol. 59, no. 11, pp. 3122–3134, 2011.

[15] S. Srinivasa and M. Haenggi, "Distance distributions in finite uniformly random networks: Theory and applications," *IEEE Trans. Veh. Technol.*, vol. 59, no. 2, pp. 940–949, 2010.

[16] K. Huang, R. W. Heath, and J. G. Andrews, "Space division multiple access with a sum feedback rate constraint," *IEEE Transactions on Signal Processing*, vol. 55, no. 7, pp. 3879–3891, 2007.

[17] H. Alzer, "On some inequalities for the incomplete gamma function," *Mathematics of Computation of the American Mathematical Society*, vol. 66, no. 218, pp. 771–778, 1997.