Analysis and Optimization of Inter-Tier Interference Coordination in Downlink Multi-Antenna HetNets with Offloading

Yueping Wu, *Member, IEEE,* Ying Cui, *Member, IEEE,* and Bruno Clerckx, *Member, IEEE*

***Abstract*— Heterogeneous networks (HetNets) with offloading is considered as an effective way to meet the high data rate demand of future wireless service. However, offloaded users suffer from strong inter-tier interference, which reduces the benefits of offloading and is one of the main limiting factors of system performance. In this paper, we investigate an interference nulling (IN) scheme in improving system performance by carefully managing the inter-tier interference to the offloaded users in downlink two-tier HetNets with multi-antenna base stations. Utilizing tools from stochastic geometry, we first derive a tractable expression for the rate coverage probability of the IN scheme. Then, by studying its order, we obtain the optimal design parameter, i.e., the degree of freedom that can be used for IN, to maximize the rate coverage probability. Finally, we analyze the rate coverage probabilities of the simple offloading scheme without interference management and the multi-antenna version of the almost blank subframes (ABS) scheme in 3GPP LTE, and compare the performance of the IN scheme with these two schemes. Both analytical and numerical results show that the IN scheme can achieve good performance gains over both of these two schemes, especially in the large antenna regime.**

***Index Terms*— Heterogeneous networks, offloading, multiple antennas, inter-tier interference coordination, rate coverage probability, stochastic geometry, optimization.**

I. Introduction

The modern wireless networks have seen a significant increase in the number of users and the scope of high data rate applications. The growth of data rate demand is expected to continue for at least a few more years [1]. The conventional cellular solution, which comprises of high power base stations (BSs), each covering a large cellular area, will not be able to scale with the increasing data rate demand. A promising solution is the deployment of low power small cell nodes overlaid with high power macro-BSs, so called heterogeneous networks (HetNets). HetNets are capable of aggressively reusing existing spectrum assets to support high data rate applications. Due to the large power at macro-BSs, most of the users intend to connect with macro-BSs, which causes the problem of *load imbalancing* [2]. To address load imbalancing, some users are offloaded to the lightly loaded small cells via a bias factor [3]. The performance of HetNets with offloading has been investigated in various literature (see e.g., [2, 4]). However, in HetNets with offloading, the offloaded users (i.e., the users offloaded from the macro-cell tier to the small-cell tier) have degraded signal-to-interference ratio (SIR), which is one of the limiting factors of network performance. Interference management techniques are thus desirable in HetNets with offloading. One such technique is almost blank subframes (ABS) in 3GPP LTE [5]. In ABS, (time or frequency) resource is partitioned, whereby the offloaded users and the other users are served using different portions of the resource. The performance of ABS in HetNets with offloading was analyzed in [5] using tools from stochastic geometry. Another interference management technique was proposed for single-antenna HetNets in [6] to reduce the interference to each offloaded user by cooperation between its nearest macro-BS and nearest pico-BS. Under the scheme in [6], the scheduled offloaded user and the users of its nearest macro-BS cannot be served using the same resource. Note that [5, 6] considered single-antenna HetNets, and both schemes studied in [5, 6] may not fully utilize system resource.

Deploying multiple antennas in HetNets can further improve data rates for future wireless service. With multiple antennas, more effective interference management techniques (e.g., coordinated beamforming [7]) can be implemented. For example, references [8–11] investigated the performance of a HetNet with a single multi-antenna macro-BS and multiple small-BSs, where the multiple antennas at the macro-BS are used for serving its scheduled users as well as mitigating interference to the receivers in small cells using different interference coordination schemes. These schemes were analyzed and shown to have performance improvement. However, since only one macro-BS is considered, the analytical results obtained in [8–11] cannot reflect the macro-tier interference, and thus cannot offer accurate insights for practical HetNets. In [12], interference coordination among a *fixed* number of neighboring BSs was investigated in downlink large multi-antenna HetNets. However, this scheme may not thoroughly exploit the spatial properties of the interference in large HetNets, and thus cannot effectively improve system performance. Moreover, offloading

Manuscript received November 11, 2014; revised May 1, 2015; accepted July 1, 2015. The editor coordinating the review of this paper and approving it for publication was L. Song.

The work of Y. Wu and B. Clerckx was partially supported by the Seventh Framework Programme for Research of the European Commission under grant number HARP-318489. The work of Y. Cui was supported by the National Science Foundation of China grant 61401272. The material in this paper has been presented in part at the IEEE International Conference on Communications (ICC), London, United Kingdom, June 2015.

Y. Wu is with the Department of Electrical and Electronic Engineering, the University of Hong Kong, Hong Kong (e-mail: eewyp@hku.hk).

Y. Cui is with the Department of Electronic Engineering, Shanghai Jiao Tong University, China (e-mail: cuiying@sjtu.edu.cn).

B. Clerckx is with the Department of Electrical and Electronic Engineering, Imperial College London, United Kingdom, and is also with the School of Electrical Engineering, Korea University, Korea (e-mail: b.clerckx@imperial.ac.uk).

was not considered in [12]. So far, it is still not clear how the interference coordination schemes and the system parameters affect the performance of large multi-antenna HetNets with offloading.

In this paper, we consider offloading in a downlink two-tier large stochastic multi-antenna HetNet where a macro-cell tier is overlaid with a pico-cell tier, and investigate an interference nulling (IN) scheme in improving the performance of offloaded users. The IN scheme has a design parameter, which is the degree of freedom U that can be used at each macro-BS for avoiding its interference to some of its offloaded users. In particular, each macro-BS utilizes the low-complexity zero-forcing beamforming (ZFBF) precoder to suppress interference to at most U offloaded users as well as boost the signal to its scheduled user. Interference coordination using beamforming technique in large stochastic HetNets causes spatial dependence among macro-BSs and pico-BSs [11], and user dependence among offloaded users. Thus, it is more challenging to analyze than interference coordination in multi-antenna stochastic *single-tier* cellular networks [13–15]. In this paper, by adopting appropriate approximations and utilizing tools from stochastic geometry, we first present a tractable expression for the rate coverage probability of the IN scheme. To our best knowledge, this is the first work analyzing the interference coordination technique in large stochastic multi-antenna HetNets with offloading.

To further improve the rate coverage probability of the IN scheme, we consider the optimization of its design parameter. Note that optimization problems in large HetNets with single-antenna BSs were investigated in [16, 17]. The objective functions in [16, 17] are relatively simple, and bounds of the objective functions and the constraints are utilized to obtain near-optimal solutions. The optimization problem in large multi-antenna HetNets we consider is an integer programming problem with a very complicated objective function. Hence, it is quite challenging to obtain the optimal solution. First, for the asymptotic scenario where the rate threshold is small, by studying the order behavior of the rate coverage probability, we prove that the optimal design parameter converges to a fixed value, which equals to either the antenna number difference between each maco-BS and each pico-BS or the antenna number difference minus one. Next, for the general scenario, we show that besides the number of antennas, the optimal design parameter also depends on other system parameters.

Finally, we compare the IN scheme with the multi-antenna version of the existing simple offloading scheme without interference management [2] and the multi-antenna version of the existing ABS scheme in 3GPP LTE [5]. In particular, we first analyze the rate coverage probabilities of the simple offloading scheme and ABS. Then, we compare the IN scheme with the simple offloading scheme and ABS, respectively, in terms of the rate coverage probability of each user type and the overall rate coverage probability. Both the analytical and numerical results show that the IN scheme can achieve good rate coverage probability gains over both of these two schemes, especially in the large antenna regime.

The rest of the paper is organized as follows. Section II presents the system model. Section III presents the proposed inter-tier IN scheme. In Section IV, the rate coverage probability of the proposed IN scheme is analyzed. In Section V, the rate coverage probability of the proposed IN scheme is optimized. In Section VI, the rate coverage probability of the proposed IN scheme is compared with those of the multi-antenna version of the existing simple offloading scheme without interference management and the multi-antenna version of the existing ABS scheme in 3GPP LTE. Finally, Section VII concludes the paper.

The key notations used in the paper are listed in Table I.

TABLE I
KEY NOTATIONS

Notation	Description
Φ_j, Φ_u	PPP of BSs in the jth tier, PPP of users
λ_j, λ_u	Density of PPP Φ_j, density of PPP Φ_u
P_j	Transmit power at each BS in the jth tier
N_j	Number of transmit antennas at each BS in the jth tier
α_j	Path loss exponent in the jth tier
B	Bias factor of the pico-cell tier
$Z_{i,j}$	Distance between user i and its nearest BS in the jth tier
Y_j	Distance between the typical user and its serving BS in the jth tier
$\mathcal{U}_J$	Set of macro-users ($J=1$), pico-users ($J=2$), unoffloaded pico-users ($J=2\bar{O}$), offloaded users ($J=2O$), IN offloaded users ($J=2OC$), non-IN offloaded users ($J=2O\bar{C}$)
$\mathcal{A}_J$	Association probability of the typical user to $\mathcal{U}_J$
W	Available resource
L_0	Total number of associated users (i.e., load) of the typical user's serving BS
$\mathcal{R}$	Rate coverage probability
$\mathcal{S}$	SIR coverage probability
τ, β	Rate threshold, SIR threshold
U	Maximum DoF for IN in the IN scheme
η	Resource fraction used for serving offloaded users only in the ABS scheme

II. SYSTEM MODEL

A. Downlink Two-Tier Heterogeneous Networks

We consider a downlink two-tier HetNet where a macro-cell tier is overlaid with a pico-cell tier, as shown in Fig. 1(a). The locations of the macro-BSs and the pico-BSs are spatially distributed as two independent homogeneous Poisson point processes (PPPs) Φ_1 and Φ_2 with densities λ_1 and λ_2, respectively. The locations of the users are also distributed as an independent homogeneous PPP Φ_u with density λ_u. Without loss of generality (w.l.o.g.), denote the macro-cell tier as the 1st tier and the pico-cell tier as the 2nd tier. The frequency reuse is adopted at all the macro-BSs and the pico-BSs. We focus on the downlink scenario. Each macro-BS has N_1 antennas with total transmission power P_1, each pico-BS has N_2 antennas with total transmission power P_2, and

each user has a single antenna. We consider both large-scale fading and small-scale fading. Specifically, due to large-scale fading, transmitted signals from the jth tier with distance r are attenuated by a factor $\frac{1}{r^{\alpha_j}}$ $(j=1,2)$, where $\alpha_j > 2$ is the path loss exponent of the jth tier. For small-scale fading, we assume Rayleigh fading channels.[1]

B. User Association

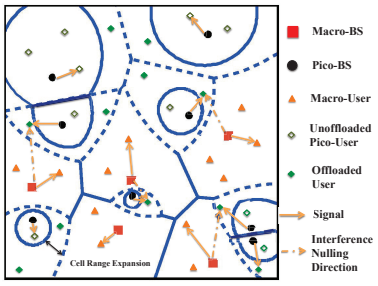
We assume open access [2]. As discussed in Section I, due to the larger power at the macro-BSs, the load imbalancing problem arises if the user association is only according to the long-term average received power (RP). To remit the load imbalancing problem, the bias factor B_j $(j=1,2)$ is introduced to tier j, where $B_2 > B_1$, to offload users from the heavily loaded macro-cell tier to the lightly loaded pico-cell tier. Specifically, user i (denoted as u_i) is associated with the BS which provides the maximum *long-term average* biased-received-power (BRP) (among all the macro-BSs and pico-BSs).[2] Here, the long-term average BRP is defined as the average RP multiplied by a bias factor. This associated BS is called the *serving BS* of u_i. Note that within each tier, the nearest BS to u_i provides the strongest long-term average BRP in this tier. Thus, u_i is associated with the nearest BS in the j_i^*th tier if[3] $j_i^* = \arg\max_{j\in\{1,2\}} P_j B_j Z_{i,j}^{-\alpha_j}$, where $Z_{i,j}$ is the distance between u_i and its nearest BS in the jth tier. We observe that, for given $\{P_j\}$, $\{Z_{i,j}\}$ and $\{\alpha_j\}$, user association is only affected by the ratio between B_1 and B_2. Thus, w.l.o.g., we assume $B_1 = 1$ and $B_2 = B > 1$. After user association, each BS schedules its associated users according to an orthogonal multiple access technique, e.g., TDMA and FDMA.

According to the above mentioned user association policy and the offloading strategy, all the users can be partitioned into the following three disjoint user sets: 1) *the set of macro-users* $\mathcal{U}_1 = \left\{u_i | P_1 Z_{i,1}^{-\alpha_1} \geq B P_2 Z_{i,2}^{-\alpha_2}\right\}$, 2) *the set of unoffloaded pico-users* $\mathcal{U}_{2\bar{O}} = \left\{u_i | P_2 Z_{i,2}^{-\alpha_2} > P_1 Z_{i,1}^{-\alpha_1}\right\}$, and 3) *the set of offloaded users* $\mathcal{U}_{2O} = \left\{u_i | P_2 Z_{i,2}^{-\alpha_2} \leq P_1 Z_{i,1}^{-\alpha_1} < B P_2 Z_{i,2}^{-\alpha_2}\right\}$, where the macro-users are associated with the maco-cell tier, the unoffloaded pico-users are associated with the pico-cell tier (even without bias), and the offloaded users are offloaded from the macro-cell tier to the pico-cell tier (due to bias $B > 1$), as illustrated in Fig. 1(b). Moreover, $\mathcal{U}_2 = \mathcal{U}_{2\bar{O}} \bigcup \mathcal{U}_{2O}$ represents *the set of pico-users*.

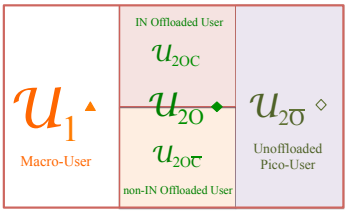
[1]Note that the Rayleigh fading channel is the most commonly used channel model in the existing literature (see e.g., [4–6]) to capture important features for practical channels to some extent and facilitate analysis for obtaining first-order insights. We will defer the investigation of the performance of stochastic large HetNets under more practical channel models to future work.

[2]The long-term average BRP user association model is commonly used in the existing literature (see e.g., [2, 4, 5]) and 3GPP standarization (see e.g., [3]) due to the following reasons. i) It is simple and practical, as each user only needs to find out the BS with the strongest (biased) RP among the nearby BSs. ii) It is tractable for analysis and helps provide first-order insights.

[3]In the user association procedure, the first antenna is normally used to transmit signal (using the total transmission power of each BS) for BRP determination according to LTE standards [18].



(a) System Model ($U = 1$)



(b) User Set Illustration

Fig. 1. System model and user set illustration.

C. Performance Metric

In this paper, we study the performance of the typical user denoted as[4] u_0, which is located at the origin and is scheduled [19]. Since HetNets are interference-limited, in this paper, we ignore the thermal noise in the analysis, as in [20]. Note that the analytical results with thermal noise can be calculated in a similar way. We investigate the *rate coverage probability* of the typical user, which is defined as the probability that the rate of the typical user is larger than a threshold [4, 5]. Specifically, let $R_0 = \frac{W}{L_0}\log_2(1+\mathrm{SIR}_0)$ denote the rate of the typical user, where W is the available resource (e.g., time or frequency), L_0 is the total number of associated users (i.e., *load*) of the typical user's serving BS, and SIR_0 is the SIR of the typical user. Then, the rate coverage probability can be mathematically written as

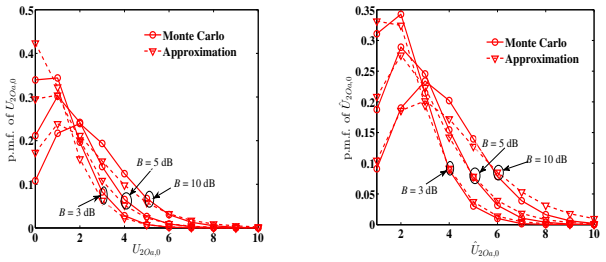
$$\mathcal{R}(\tau) \triangleq \Pr\left(R_0 > \tau\right) = \Pr\left(\frac{W}{L_0}\log_2\left(1+\mathrm{SIR}_0\right) > \tau\right) \quad (1)$$

where τ is the rate threshold. Note that R_0 is a random variable with randomness induced by SIR_0 and L_0. Thus, the rate coverage probability captures the effects of the distributions of both SIR_0 and L_0 [5]. Please note that the rate coverage probability is suitable for applications for which the rate can reflect the performance. For example, it is suitable for applications with strict rate requirements (e.g., video services) as well as applications which may not have strict rate requirements but still include rate as an important parameter reflecting the QoS requirement (e.g., data downloading).

III. Inter-tier Interference Nulling

In HetNets with offloading, offloaded users normally suffer from stronger interference than macro-users and unoffloaded

[4]The index of the typical user and its serving BS is 0.



(a) (b)

Fig. 2. P.m.f. of $U_{2O_a,0}$ and $\hat{U}_{2O_a,0}$ for different bias factors B, at $\frac{P_1}{P_2} = 20$ dB, $\alpha_1 = \alpha_2 = 4$, $\lambda_1 = 0.0001$ nodes/m^2, and $\lambda_2 = 0.0005$ nodes/m^2.

pico-users.[5] The dominant interference to each offloaded user, caused by its nearest macro-BS [4], is one of the limiting factors of system performance. In this section, we first elaborate on an inter-tier IN scheme to avoid the dominant interference to offloaded users, so as to improve system performance. Then, we obtain some results on the distributions of some related random variables of this scheme.

A. IN Scheme Description

There are roughly two types of interference management techniques, namely, interference cooperation techniques [21] and interference coordination techniques [7]. To implement interference cooperation techniques, both data and CSI need to be shared among different BSs; while to implement interference coordination techniques, only CSI needs to be shared among different BSs. Compared to interference cooperation techniques, interference coordination techniques have relatively low cost and complexity, although they may not perform as well as interference cooperation techniques. In this paper, we focus on low-complexity interference management designs and hence adopt interference coordination techniques.

We now describe an inter-tier IN scheme to avoid the dominant interference to offloaded users by making use of at most U ($U < N_1$) DoF at each macro-BS which has N_1 antennas. In particular, we use the low-complexity ZFBF precoder[6] at each macro-BS to perform inter-tier IN. Note that U is the design parameter of this scheme. When $U = 0$, the IN scheme reduces to the simple offloading scheme without interference management [2]. We first introduce several types of users related to this scheme. For each macro-BS, we refer to the users offloaded from it to their nearby pico-BSs as the *offloaded users* of this macro-BS. All these offloaded users may not be scheduled by their nearest pico-BSs simultaneously, as each BS schedules one user in each time slot. In each time slot, we refer to the offloaded users scheduled by their nearest pico-BSs as *active offloaded users* (of this slot). In the IN scheme, each macro-BS avoids its interference to some of its active offloaded users in a particular time slot, which are referred to as the *IN offloaded users* of this macro-BS. We refer to the remaining offloaded users as *non-IN offloaded users*. Hence, under the IN scheme, in a particular time slot, the offloaded users $\mathcal{U}_{2O}$ are further divided into two sets, i.e., $\mathcal{U}_{2O} = \mathcal{U}_{2OC} \bigcup \mathcal{U}_{2O\bar{C}}$, where $\mathcal{U}_{2OC}$ denotes the IN offloaded user set and $\mathcal{U}_{2O\bar{C}}$ denotes the non-IN offloaded user set. Note that under the IN scheme, the users can be partitioned into four disjoint user sets, namely, $\mathcal{U}_1$, $\mathcal{U}_{2\bar{O}}$, $\mathcal{U}_{2OC}$ and $\mathcal{U}_{2O\bar{C}}$, as illustrated in Fig. 1(b). Next, we discuss how to determine the IN offloaded users of each macro-BS. Specifically, let $U_{2O_a,\ell}$ denote the number of active offloaded users of macro-BS ℓ, each of which is scheduled by a different pico-BS. If $U_{2O_a,\ell} \leq U$, macro-BS ℓ can perform IN to all of its $U_{2O_a,\ell}$ active offloaded users using $U_{2O_a,\ell}$ DoF. However, if $U_{2O_a,\ell} > U$, macro-BS ℓ randomly selects U out of $U_{2O_a,\ell}$ active offloaded users according to the uniform distribution to perform IN using U DoF. Hence, macro-BS ℓ performs IN to $u_{2OC,\ell} \triangleq \min(U, U_{2O_a,\ell})$ out of $U_{2O_a,\ell}$ active offloaded users. Note that the DoF used for IN (referred to as IN DoF) at macro-BS ℓ is $u_{2OC,\ell}$. All the remaining $N_1 - u_{2OC,\ell}$ DoF at macro-BS ℓ are used for boosting the signal to its scheduled user.

Now, we introduce the precoding vectors at macro-BSs and pico-BSs in the IN scheme, respectively. First, each macro-BS utilizes the low-complexity ZFBF precoder to serve its scheduled user and simultaneously perform IN to its IN offloaded users. Specifically, denote $\mathbf{H}_{1,\ell} = \left[\mathbf{h}_{1,\ell}\, \mathbf{g}_{1,\ell 1}\, \ldots\, \mathbf{g}_{1,\ell u_{2OC,\ell}}\right]^\dagger$, where[7] $\mathbf{h}_{1,\ell} \overset{d}{\sim} \mathcal{CN}_{N_1,1}(\mathbf{0}_{N_1\times 1}, \mathbf{I}_{N_1})$ denotes the channel vector between macro-BS ℓ and its scheduled user, and $\mathbf{g}_{1,\ell i} \overset{d}{\sim} \mathcal{CN}_{N_1,1}(\mathbf{0}_{N_1\times 1}, \mathbf{I}_{N_1})$ denotes the channel vector between macro-BS ℓ and its IN offloaded user i ($i = 1, \ldots, u_{2OC,\ell}$). The ZFBF precoding matrix at macro-BS ℓ is designed to be $\mathbf{W}_{1,\ell} = \mathbf{H}_{1,\ell}^\dagger \left(\mathbf{H}_{1,\ell}\mathbf{H}_{1,\ell}^\dagger\right)^{-1}$ and the ZFBF vector at macro-BS ℓ is designed to be $\mathbf{f}_{1,\ell} = \frac{\mathbf{w}_{1,\ell}}{\|\mathbf{w}_{1,\ell}\|}$, where $\mathbf{w}_{1,\ell}$ is the first column of $\mathbf{W}_{1,\ell}$. To perform IN (via ZFBF precoder) under FDD, at most U active offloaded users need to estimate the CSI to their nearest macro-BS, and feedback the CSI to their associated pico-BSs through error-free feedback channels. Then, the associated pico-BSs share the CSI with these offloaded users' nearest macro-BS through backhaul. To perform IN under TDD, the CSI between each macro-BS and at most U of its active offloaded users can be obtained by exploiting reciprocity through standard channel sounding [22].

Next, each pico-BS utilizes the maximal ratio transmission precoder to serve its scheduled user. Specifically, the beamforming vector at pico-BS ℓ is $\mathbf{f}_{2,\ell} = \frac{\mathbf{h}_{2,\ell}}{\|\mathbf{h}_{2,\ell}\|}$, where $\mathbf{h}_{2,\ell} \overset{d}{\sim} \mathcal{CN}_{N_2,1}(\mathbf{0}_{N_2\times 1}, \mathbf{I}_{N_2})$ denotes the channel vector between pico-BS ℓ and its scheduled user.

We now discuss the received signal and the corresponding SIR of the typical user $u_0 \in \mathcal{U}_k$ ($k \in \mathcal{K} \triangleq \{1, 2\bar{O}, 2OC, 2O\bar{C}\}$).

[5]For each offloaded user, its nearest macro-BS, which provides the strongest long-term average RP, now becomes the dominant interferer of this offloaded user. However, for each macro-user or unoffloaded pico-user, the BS which provides the strongest long-term average RP is its serving BS. Therefore, offloaded users suffer the strongest interference.

[6]The ZFBF precoding scheme considered in this paper is one of the commonly used interference coordination techniques.

[7]The notation $X \overset{d}{\sim} Y$ means that X *is distributed as* Y.

1) Macro-User: The received signal and SIR of the typical user $u_0 \in \mathcal{U}_1$ are[8]

$$y_{1,0} = \frac{1}{Y_1^{\frac{\alpha_1}{2}}}\mathbf{h}_{1,00}^{\dagger}\mathbf{f}_{1,0}x_{1,0} + \sum_{\ell\in\Phi(\lambda_1)\setminus B_{1,0}} \frac{1}{|D_{1,\ell 0}|^{\frac{\alpha_1}{2}}}\mathbf{h}_{1,\ell 0}^{\dagger}\mathbf{f}_{1,\ell}x_{1,\ell} + \sum_{\ell\in\Phi(\lambda_2)} \frac{1}{|D_{2,\ell 0}|^{\frac{\alpha_2}{2}}}\mathbf{h}_{2,\ell 0}^{\dagger}\mathbf{f}_{2,\ell}x_{2,\ell}\ , \quad (2)$$

$$\mathrm{SIR}_{\mathrm{IN},1,0} = \frac{\frac{P_1}{Y_1^{\alpha_1}}\left|\mathbf{h}_{1,00}^{\dagger}\mathbf{f}_{1,0}\right|^2}{P_1\sum_{\ell\in\Phi(\lambda_1)\setminus B_{1,0}}\frac{\left|\mathbf{h}_{1,\ell 0}^{\dagger}\mathbf{f}_{1,\ell}\right|^2}{|D_{1,\ell 0}|^{\alpha_1}} + P_2\sum_{\ell\in\Phi(\lambda_2)}\frac{\left|\mathbf{h}_{2,\ell 0}^{\dagger}\mathbf{f}_{2,\ell}\right|^2}{|D_{2,\ell 0}|^{\alpha_2}}} \quad (3)$$

where $B_{1,0}$ is the serving macro-BS of u_0, Y_1 is the distance between u_0 and $B_{1,0}$, $|D_{j,\ell 0}|$ $(j=1,2)$ is the distance from BS ℓ in the jth tier to u_0, $x_{1,\ell}$ is the symbol sent from macro-BS ℓ to its scheduled user satisfying $\mathrm{E}\left[x_{1,\ell}x_{1,\ell}^*\right] = P_1$, and $x_{2,\ell}$ is the symbol sent from pico-BS ℓ to its scheduled user satisfying $\mathrm{E}\left[x_{2,\ell}x_{2,\ell}^*\right] = P_2$. Here, $\left|\mathbf{h}_{1,00}^{\dagger}\mathbf{f}_{1,0}\right|^2 \overset{d}{\sim} \mathrm{Gamma}\left(N_1 - u_{2OC,0}, 1\right)$, $\left|\mathbf{h}_{1,\ell 0}^{\dagger}\mathbf{f}_{1,\ell}\right|^2 \overset{d}{\sim} \mathrm{Gamma}(1,1)$, and $\left|\mathbf{h}_{2,\ell 0}^{\dagger}\mathbf{f}_{2,\ell}\right|^2 \overset{d}{\sim} \mathrm{Gamma}(1,1)$.

2) Unoffloaded Pico-User: The received signal and SIR of the typical user $u_0 \in \mathcal{U}_{2\bar{O}}$ are

$$y_{2\bar{O},0} = \frac{1}{Y_2^{\frac{\alpha_2}{2}}}\mathbf{h}_{2,00}^{\dagger}\mathbf{f}_{2,0}x_{2,0} + \sum_{\ell\in\Phi(\lambda_1)} \frac{1}{|D_{1,\ell 0}|^{\frac{\alpha_1}{2}}}\mathbf{h}_{1,\ell 0}^{\dagger}\mathbf{f}_{1,\ell}x_{1,\ell} + \sum_{\ell\in\Phi(\lambda_2)\setminus B_{2,0}} \frac{1}{|D_{2,\ell 0}|^{\frac{\alpha_2}{2}}}\mathbf{h}_{2,\ell 0}^{\dagger}\mathbf{f}_{2,\ell}x_{2,\ell}\ , \quad (4)$$

$$\mathrm{SIR}_{\mathrm{IN},2\bar{O},0} = \frac{\frac{P_2}{Y_2^{\alpha_2}}\left|\mathbf{h}_{2,00}^{\dagger}\mathbf{f}_{2,0}\right|^2}{P_1\sum_{\ell\in\Phi(\lambda_1)}\frac{\left|\mathbf{h}_{1,\ell 0}^{\dagger}\mathbf{f}_{1,\ell}\right|^2}{|D_{1,\ell 0}|^{\alpha_1}} + P_2\sum_{\ell\in\Phi(\lambda_2)\setminus B_{2,0}}\frac{\left|\mathbf{h}_{2,\ell 0}^{\dagger}\mathbf{f}_{2,\ell}\right|^2}{|D_{2,\ell 0}|^{\alpha_2}}} \quad (5)$$

where $B_{2,0}$ is the serving pico-BS of u_0, and Y_2 is the distance between u_0 and $B_{2,0}$. Here, $\left|\mathbf{h}_{2,00}^{\dagger}\mathbf{f}_{2,0}\right|^2 \overset{d}{\sim} \mathrm{Gamma}\left(N_2, 1\right)$.

3) IN Offloaded User: When $u_0 \in \mathcal{U}_{2OC}$, the typical user u_0 does not suffer interference from its nearest macro-BS. Thus, the received signal and SIR of $u_0 \in \mathcal{U}_{2OC}$ are

$$y_{2OC,0} = \frac{1}{Y_2^{\frac{\alpha_2}{2}}}\mathbf{h}_{2,00}^{\dagger}\mathbf{f}_{2,0}x_{2,0} + \sum_{\ell\in\Phi(\lambda_1)\setminus B_{1,0}} \frac{\mathbf{h}_{1,\ell 0}^{\dagger}\mathbf{f}_{1,\ell}}{|D_{1,\ell 0}|^{\frac{\alpha_1}{2}}}x_{1,\ell} + \sum_{\ell\in\Phi(\lambda_2)\setminus B_{2,0}} \frac{\mathbf{h}_{2,\ell 0}^{\dagger}\mathbf{f}_{2,\ell}}{|D_{2,\ell 0}|^{\frac{\alpha_2}{2}}}x_{2,\ell}\ , \quad (6)$$

$$\mathrm{SIR}_{\mathrm{IN},2OC,0} = \frac{\frac{P_2}{Y_2^{\alpha_2}}\left|\mathbf{h}_{2,00}^{\dagger}\mathbf{f}_{2,0}\right|^2}{P_1\sum_{\ell\in\Phi(\lambda_1)\setminus B_{1,0}}\frac{\left|\mathbf{h}_{1,\ell 0}^{\dagger}\mathbf{f}_{1,\ell}\right|^2}{|D_{1,\ell 0}|^{\alpha_1}} + P_2\sum_{\ell\in\Phi(\lambda_2)\setminus B_{2,0}}\frac{\left|\mathbf{h}_{2,\ell 0}^{\dagger}\mathbf{f}_{2,\ell}\right|^2}{|D_{2,\ell 0}|^{\alpha_2}}} \quad (7)$$

[8]In this paper, all macro-BSs and pico-BSs are assumed to be active. The same assumption can also be seen in the existing papers (see e.g., [5, 23]).

4) Non-IN Offloaded User: When $u_0 \in \mathcal{U}_{2O\bar{C}}$, the typical user u_0 is not selected for IN, and thus it still suffers interference from its nearest macro-BS. Hence, the received signal and SIR of $u_0 \in \mathcal{U}_{2O\bar{C}}$ are

$$y_{2O\bar{C},0} = \frac{\mathbf{h}_{2,00}^{\dagger}\mathbf{f}_{2,0}x_{2,0}}{Y_2^{\frac{\alpha_2}{2}}} + \frac{\mathbf{h}_{1,10}^{\dagger}\mathbf{f}_{1,1}x_{1,1}}{Y_1^{\frac{\alpha_1}{2}}} + \sum_{\ell\in\Phi(\lambda_1)\setminus B_{1,0}} \frac{\mathbf{h}_{1,\ell 0}^{\dagger}\mathbf{f}_{1,\ell}x_{1,\ell}}{|D_{1,\ell 0}|^{\frac{\alpha_1}{2}}} + \sum_{\ell\in\Phi(\lambda_2)\setminus B_{2,0}} \frac{\mathbf{h}_{2,\ell 0}^{\dagger}\mathbf{f}_{2,\ell}x_{2,\ell}}{|D_{2,\ell 0}|^{\frac{\alpha_2}{2}}} \quad (8)$$

$$\mathrm{SIR}_{\mathrm{IN},2O\bar{C},0} = \frac{\frac{P_2}{Y_2^{\alpha_2}}\left|\mathbf{h}_{2,00}^{\dagger}\mathbf{f}_{2,0}\right|^2}{\sum_{j=1}^{2}\sum_{\ell\in\Phi(\lambda_j)\setminus B_{j,0}}\frac{P_j\left|\mathbf{h}_{j,\ell 0}^{\dagger}\mathbf{f}_{,\ell}\right|^2}{|D_{j,\ell 0}|^{\alpha_j}} + \frac{P_1\left|\mathbf{h}_{1,10}^{\dagger}\mathbf{f}_{1,1}\right|^2}{Y_1^{\alpha_1}}}\ . \quad (9)$$

To facilitate the calculation of the rate coverage probability for $u_0 \in \mathcal{U}_{2O\bar{C}}$ in Section IV, different from (4) and (5), we separate the dominant interferer (i.e., the nearest macro-BS) and the other interferers (i.e., the other macro-BSs) in the macro-cell tier to $u_0 \in \mathcal{U}_{2O\bar{C}}$ in (8) and (9).

B. Preliminary Results

In this part, we provide some preliminary results, which are the basis of calculating the rate coverage probability in (1). Let $U_{2O_a,0}$ denote the number of active offloaded users of the typical user's serving macro-BS when $u_0 \in \mathcal{U}_1$. We first calculate the p.m.f. of $U_{2O_a,0}$. The p.m.f. of $U_{2O_a,0}$ depends on the distributions of the number of active offloaded users in a fixed area and the offloading area of the typical user's serving macro-BS, but its exact distribution is unknown. Similar to the approaches utilized in [15, 24], we approximate the locations of scheduled pico-users as a PPP. The resulting distribution of the number of active offloaded users in a fixed area is a Poisson distribution. Moreover, we approximate the distribution of the offloading area of a macro-BS using the first moment matching method [5]. Based on these approximations,[9] we calculate the p.m.f. of $U_{2O_a,0}$ as follows:

Lemma 1: When $u_0 \in \mathcal{U}_1$, the p.m.f. of $U_{2O_a,0}$ is approximated by

$$\Pr\left(U_{2O_a,0} = n\right) \approx \frac{3.5^{3.5}\Gamma\left(n+3.5\right)}{\Gamma(3.5)n!}\left(\frac{\lambda_2\mathcal{A}_{2O}}{\mathcal{A}_2\lambda_1}\right)^n \times \left(3.5 + \frac{\lambda_2\mathcal{A}_{2O}}{\mathcal{A}_2\lambda_1}\right)^{-(n+3.5)}\ ,\ n \geq 0 \quad (10)$$

where $\mathcal{A}_2 \triangleq \Pr\left(u_0 \in \mathcal{U}_2\right) = 2\pi\lambda_2\int_0^{\infty} z\exp\left(-\pi\lambda_2 z^2\right) \times \exp\left(-\pi\lambda_1\left(\frac{P_1 z^{\alpha_2}}{BP_2}\right)^{\frac{2}{\alpha_1}}\right)\mathrm{d}z$ and $\mathcal{A}_{2O} \triangleq \Pr\left(u_0 \in \mathcal{U}_{2O}\right) =$

[9]In this paper, we use these approximations to facilitate the analysis and get first-order insights. Later, in Fig. 2, we will show the good accuracy of the utilized approximations via numerical results. Please note that, due to the complexity of the considered system model, the analytical characterizations of the accuracy of these approximations are quite challenging and are still open problems. We will defer the analytical characterizations of the accuracy of these approximations to future work.

$$2\pi\lambda_2 \int_0^\infty z \exp\left(-\pi\lambda_2 z^2\right) \exp\left(-\pi\lambda_1 \left(\frac{P_1 z^{\alpha_2}}{BP_2}\right)^{\frac{2}{\alpha_1}}\right) \mathrm{d}z$$
$$-2\pi\lambda_2 \int_0^\infty z \exp\left(-\pi\lambda_2 z^2\right) \exp\left(-\pi\lambda_1 \left(\frac{P_1 z^{\alpha_2}}{P_2}\right)^{\frac{2}{\alpha_1}}\right) \mathrm{d}z.$$

Proof: See Appendix A. ∎

Fig. 2(a) illustrates the accuracy of the p.m.f. approximation of $U_{2O_a,0}$ in (10). We see that the p.m.f. approximation of $U_{2O_a,0}$ is reasonably accurate for different bias factors.

Next, let $\hat{U}_{2O_a,0}$ denote the number of active offloaded users that are offloaded from the typical user's nearest macro-BS when $u_0 \in \mathcal{U}_{2O}$. We now calculate the p.m.f. of $\hat{U}_{2O_a,0}$. Based on similar approximation approaches for deriving the p.m.f. of $U_{2O_a,0}$ in *Lemma 1*, we calculate the p.m.f. of $\hat{U}_{2O_a,0}$ as follows:

Lemma 2: When $u_0 \in \mathcal{U}_{2O}$, the p.m.f. of $\hat{U}_{2O_a,0}$ is approximated by

$$\Pr\left(\hat{U}_{2O_a,0}=n\right) \approx \frac{3.5^{3.5}\Gamma\left(n+3.5\right)}{\Gamma(n)\Gamma(3.5)} \left(\frac{\lambda_2\mathcal{A}_{2O}}{\mathcal{A}_2\lambda_1}\right)^{n-1} \times \left(3.5+\frac{\lambda_2\mathcal{A}_{2O}}{\mathcal{A}_2\lambda_1}\right)^{-(n+3.5)}, \; n \geq 1 \,. \quad (11)$$

Proof: Similar to the proof of (10). The difference is that, in this proof, the distribution of the offloading area (where the offloaded users including u_0 may reside) of u_0's *nearest* macro-BS is used, instead of the distribution of the offloading area (where the offloaded users excluding u_0 may reside) of u_0's *serving* macro-BS (used in the proof of (10)). ∎

Fig. 2(b) illustrates the accuracy of the p.m.f. approximation of $\hat{U}_{2O_a,0}$ in (11). We see that the p.m.f. approximation of $\hat{U}_{2O_a,0}$ is reasonably accurate for different bias factors.

IV. Rate Coverage Probability Analysis of Interference Nulling

In this section, we investigate the rate coverage probability of the IN scheme. First, we derive the SIR coverage probability of each user type. Next, based on the SIR coverage probabilities of all user types, we obtain the rate coverage probability and its mean load approximation (MLA).

A. SIR Coverage Probability of Each User Type

As discussed in Section III-A, under the IN scheme, the typical user u_0 can be in any user set $\mathcal{U}_k$, where $k \in \mathcal{K} \triangleq \{1, 2\bar{O}, 2OC, 2O\bar{C}\}$. Let[10] $\mathcal{S}_{\mathrm{IN},k}(\beta) \triangleq \Pr\left(\mathrm{SIR}_{\mathrm{IN},k,0} > \beta | u_0 \in \mathcal{U}_k\right)$ denote the SIR coverage probability of $u_0 \in \mathcal{U}_k$ ($k \in \mathcal{K}$) under the IN scheme, where $\mathrm{SIR}_{\mathrm{IN},k,0}$ denotes the SIR of $u_0 \in \mathcal{U}_k$ under the IN scheme and β is the SIR threshold. We first present the SIR coverage probability $\mathcal{S}_{\mathrm{IN},k}(\beta)$ of $u_0 \in \mathcal{U}_k$ ($k \in \mathcal{K}$) as follows:

Theorem 1: The SIR coverage probability of $u_0 \in \mathcal{U}_k$ under the IN scheme is

$$\mathcal{S}_{\mathrm{IN},1}(\beta) = \sum_{n=0}^{U} \left(\int_0^\infty \mathcal{S}_{\mathrm{IN},1,Y_1}(y,\beta) f_{Y_1}(y)\mathrm{d}y\right) \times \Pr\left(u_{2OC,0}=n\right) \,, \quad (12)$$

$$\mathcal{S}_{\mathrm{IN},2\bar{O}}(\beta) = \int_0^\infty \mathcal{S}_{\mathrm{IN},2\bar{O},Y_2}(y,\beta) f_{Y_2}(y)\mathrm{d}y \,, \quad (13)$$

$$\mathcal{S}_{\mathrm{IN},2OC}(\beta) = \int_0^\infty \int_{\left(\frac{P_2}{P_1}\right)^{\frac{1}{\alpha_2}} x^{\frac{\alpha_1}{\alpha_2}}}^{\left(\frac{BP_2}{P_1}\right)^{\frac{1}{\alpha_2}} x^{\frac{\alpha_1}{\alpha_2}}} \mathcal{S}_{\mathrm{IN},2OC,Y_1,Y_2}(x,y,\beta) \times f_{Y_1,Y_2}(x,y)\mathrm{d}y\mathrm{d}x \,, \quad (14)$$

$$\mathcal{S}_{\mathrm{IN},2O\bar{C}}(\beta) = \int_0^\infty \int_{\left(\frac{P_2}{P_1}\right)^{\frac{1}{\alpha_2}} x^{\frac{\alpha_1}{\alpha_2}}}^{\left(\frac{BP_2}{P_1}\right)^{\frac{1}{\alpha_2}} x^{\frac{\alpha_1}{\alpha_2}}} \mathcal{S}_{\mathrm{IN},2O\bar{C},Y_1,Y_2}(x,y,\beta) \times f_{Y_1,Y_2}(x,y)\mathrm{d}y\mathrm{d}x \,, \quad (15)$$

where $f_{Y_1}(y) = \frac{2\pi\lambda_1}{\mathcal{A}_1} y \exp\left(-\pi\left(\lambda_1 y^2 + \lambda_2\left(\frac{P_2 B}{P_1}\right)^{\frac{2}{\alpha_2}} y^{\frac{2\alpha_1}{\alpha_2}}\right)\right)$,
$f_{Y_2}(y) = \frac{2\pi\lambda_2}{\mathcal{A}_{2\bar{O}}} y \exp(-\pi\lambda_2 y^2) \exp\left(-\pi\lambda_1\left(\frac{P_1}{P_2}\right)^{\frac{2}{\alpha_1}} y^{\frac{2\alpha_2}{\alpha_1}}\right)$,
$f_{Y_1,Y_2}(x,y) = \frac{4\pi^2\lambda_1\lambda_2}{\mathcal{A}_{2O}} xy \exp\left(-\pi\left(\lambda_1 x^2 + \lambda_2 y^2\right)\right)$,
$\mathcal{S}_{\mathrm{IN},1,Y_1}(y,\beta)$, $\mathcal{S}_{\mathrm{IN},2\bar{O},Y_2}(y,\beta)$, $\mathcal{S}_{\mathrm{IN},2OC,Y_1,Y_2}(x,y,\beta)$ and $\mathcal{S}_{\mathrm{IN},2O\bar{C},Y_1,Y_2}(x,y,\beta)$ are given in *Lemma 7* in Appendix B, and

$$\Pr\left(u_{2OC,0}=n\right) = \begin{cases} \Pr\left(U_{2O_a,0}=n\right), & \text{for } 0 \leq n < U \\ \sum_{u=n}^{\infty} \Pr\left(U_{2O_a,0}=u\right), \text{for } n = U \end{cases} . \quad (16)$$

Here, $\mathcal{A}_{2O}$ is given in *Lemma 1*, $\mathcal{A}_1 = 2\pi\lambda_1 \int_0^\infty z \exp\left(-\pi\lambda_1 z^2 - \pi\left(\lambda_2\left(\frac{BP_2}{P_1}\right)^{\frac{2}{\alpha_2}} z^{\frac{2\alpha_1}{\alpha_2}}\right)\right)\mathrm{d}z$,
$\mathcal{A}_{2\bar{O}} = 2\pi\lambda_2 \int_0^\infty z \exp\left(-\pi\lambda_1\left(\frac{P_1 z^{\alpha_2}}{P_2}\right)^{\frac{2}{\alpha_1}} - \pi\lambda_2 z^2\right)\mathrm{d}z$, and $\Pr\left(U_{2O_a,0}=n\right)$ is given in (10).

Proof: See Appendix B. ∎

B. Rate Coverage Probability

Similar to (1), the rate coverage probability of $u_0 \in \mathcal{U}_k$ ($k \in \mathcal{K}$) under the IN scheme is defined as[11]

$$\begin{aligned}\mathcal{R}_{\mathrm{IN},k}(\tau) &\triangleq \Pr\left(R_{\mathrm{IN,k},0} > \tau | u_0 \in \mathcal{U}_k\right) \\ &= \Pr\left(\frac{W}{L_{0,j_k}} \log_2\left(1+\mathrm{SIR}_{\mathrm{IN},k,0}\right) > \tau | u_0 \in \mathcal{U}_k\right) \\ &= \mathrm{E}_{L_{0,j_k}}\left[\mathcal{S}_{\mathrm{IN},k}\left(f\left(\frac{L_{0,j_k}\tau}{W}\right)\right)\right] \end{aligned} \quad (17)$$

where $R_{\mathrm{IN,k},0}$ denotes the rate of $u_0 \in \mathcal{U}_k$ under the IN scheme, $f(x) = 2^x - 1$, and L_{0,j_k} is the load of the typical user's serving BS which is in the j_kth tier. Here, j_k is given in Table II. According to *Theorem 1* and total probability theorem, we have the rate coverage probability as follows:

Theorem 2: The rate coverage probability of the IN scheme

[10]Note that $\mathcal{S}_{\mathrm{IN},1}(\beta)$ is dependent of the design parameter U, while $\mathcal{S}_{\mathrm{IN},k}(\beta)$ is independent of U for all $k \in \{2\bar{O}, 2OC, 2O\bar{C}\}$. For notational simplicity, we do not make explicit the dependence of $\mathcal{S}_{\mathrm{IN},1}(\beta)$ on U.

[11]Note that $\mathcal{R}_{\mathrm{IN},1}(\tau)$ is dependent of the design parameter U, while $\mathcal{R}_{\mathrm{IN},k}(\tau)$ is independent of U for all $k \in \{2\bar{O}, 2OC, 2O\bar{C}\}$. For notational simplicity, we do not make explicit the dependence of $\mathcal{R}_{\mathrm{IN},1}(\tau)$ on U.

under U is

$$\mathcal{R}_{\mathrm{IN}}(U,\tau) = \mathcal{A}_1\mathcal{R}_{\mathrm{IN},1}(\tau) + \mathcal{A}_{2O}\Pr\left(\mathcal{E}_{2OC,0}(U)\right)\mathcal{R}_{\mathrm{IN},2OC}(\tau) + \mathcal{A}_{2\bar{O}}\mathcal{R}_{\mathrm{IN},2\bar{O}}(\tau) + \mathcal{A}_{2O}\left(1-\Pr\left(\mathcal{E}_{2OC,0}(U)\right)\right)\mathcal{R}_{\mathrm{IN},2O\bar{C}}(\tau) \tag{18}$$

where $\Pr\left(\mathcal{E}_{2OC,0}(U)\right) = U\left(\frac{\lambda_1\mathcal{A}_2}{\lambda_2\mathcal{A}_{2O}}\left(1-\left(1+\frac{\lambda_2\mathcal{A}_{2O}}{3.5\lambda_1\mathcal{A}_2}\right)^{-3.5}\right) - \sum_{n=1}^{U}\frac{1}{n}\Pr\left(\hat{U}_{2O_a,0}=n\right)\right) + \sum_{n=1}^{U}\Pr\left(\hat{U}_{2O_a,0}=n\right)$ with $\Pr\left(\hat{U}_{2O_a,0}=n\right)$ given in (11), $\mathcal{A}_1$, $\mathcal{A}_{2\bar{O}}$ and $\mathcal{A}_{2O}$ given in *Theorem 1*, and

$$\mathcal{R}_{\mathrm{IN},1}(\tau) = \sum_{n\geq 1}\Pr\left(L_{0,1}=n\right)\mathcal{S}_{\mathrm{IN},1}\left(f\left(\frac{n\tau}{W}\right)\right), \tag{19}$$

$$\mathcal{R}_{\mathrm{IN},2\bar{O}}(\tau) = \sum_{n\geq 1}\Pr\left(L_{0,2}=n\right)\mathcal{S}_{\mathrm{IN},2\bar{O}}\left(f\left(\frac{n\tau}{W}\right)\right), \tag{20}$$

$$\mathcal{R}_{\mathrm{IN},2OC}(\tau) = \sum_{n\geq 1}\Pr(L_{0,2}=n)\mathcal{S}_{\mathrm{IN},2OC}\left(f\left(\frac{n\tau}{W}\right)\right), \tag{21}$$

$$\mathcal{R}_{\mathrm{IN},2O\bar{C}}(\tau) = \sum_{n\geq 1}\Pr(L_{0,2}=n)\mathcal{S}_{\mathrm{IN},2O\bar{C}}\left(f\left(\frac{n\tau}{W}\right)\right). \tag{22}$$

Here, $\mathcal{S}_{\mathrm{IN},k}(\cdot)$ is given by (12)–(15), and $\Pr\left(L_{0,j}=n\right) = \frac{3.5^{3.5}\Gamma(n+3.5)\left(\frac{\lambda_u\mathcal{A}_j}{\lambda_j}\right)^{n-1}}{\Gamma(3.5)(n-1)!\left(\frac{\lambda_u\mathcal{A}_j}{\lambda_j}+3.5\right)^{n+3.5}}$ $(j=1,2)$.

Proof: Follows by conditioning on the load (i.e., $L_{1,0}$ or $L_{2,0}$), calculating the conditional rate coverage probability according to *Lemma 7*, and removing the conditions on the load (i.e., $L_{0,1}$ or $L_{0,2}$). The proof of $\Pr\left(\mathcal{E}_{2OC,0}(U)\right)$ follows the total probability theorem and [25, *Proposition 2*]. Note that the p.m.f. of $L_{0,1}$ is given in [5, *Lemma 3*], and the p.m.f. of $L_{0,2}$ can be calculated using a similar approach. ■

Note that the expression of $\mathcal{R}_{\mathrm{IN}}(U,\tau)$ in (18) of *Theorem 2* is difficult to compute and analyze due to the infinite summations over n in (19)–(22). To simplify the expression of $\mathcal{R}_{\mathrm{IN}}(U,\tau)$ in (18), we use the mean of the random load (i.e., $\mathrm{E}\left[L_{0,j}\right]$) to approximate the random load (i.e., $L_{0,j}$), where $j=1,2$ [4, 5]. The simplification is achieved due to the elimination of the infinite summation over n. In other words, by replacing $L_{0,j}$ with $\mathrm{E}\left[L_{0,j}\right]$ in (17), we can obtain the rate coverage probability with MLA of the IN scheme under U, denoted as $\bar{\mathcal{R}}_{\mathrm{IN}}(U,\tau)$, as follows:

Corollary 1: The rate coverage probability with MLA of the IN scheme under U is

$$\bar{\mathcal{R}}_{\mathrm{IN}}(U,\tau) = \mathcal{A}_1\bar{\mathcal{R}}_{\mathrm{IN},1}(\tau) + \mathcal{A}_{2O}\Pr\left(\mathcal{E}_{2OC,0}(U)\right)\mathcal{R}_{\mathrm{IN},2OC}(\tau) + \mathcal{A}_{2\bar{O}}\bar{\mathcal{R}}_{\mathrm{IN},2\bar{O}}(\tau) + \mathcal{A}_{2O}\left(1-\Pr\left(\mathcal{E}_{2OC,0}(U)\right)\right)\bar{\mathcal{R}}_{\mathrm{IN},2O\bar{C}}(\tau) \tag{23}$$

where $\bar{\mathcal{R}}_{\mathrm{IN},1}(\tau) = \mathcal{S}_{\mathrm{IN},1}\left(f\left(\frac{\mathrm{E}[L_{0,1}]\tau}{W}\right)\right)$, $\bar{\mathcal{R}}_{\mathrm{IN},2\bar{O}}(\tau) = \mathcal{S}_{\mathrm{IN},2\bar{O}}\left(f\left(\frac{\mathrm{E}[L_{0,2}]\tau}{W}\right)\right)$, $\bar{\mathcal{R}}_{\mathrm{IN},2OC}(\tau) = \mathcal{S}_{\mathrm{IN},2OC}\left(f\left(\frac{\mathrm{E}[L_{0,2}]\tau}{W}\right)\right)$, $\bar{\mathcal{R}}_{\mathrm{IN},2O\bar{C}}(\tau) = \mathcal{S}_{\mathrm{IN},2O\bar{C}}\left(f\left(\frac{\mathrm{E}[L_{0,2}]\tau}{W}\right)\right)$ with $\mathcal{S}_{\mathrm{IN},k}(\cdot)$ given by (12)–(15), $\mathrm{E}\left[L_{0,1}\right] = 1+1.28\frac{\lambda_u\mathcal{A}_1}{\lambda_1}$, and $\mathrm{E}\left[L_{0,2}\right] = 1+1.28\frac{\lambda_u\mathcal{A}_2}{\lambda_2}$. Here, $\mathcal{A}_1$, $\mathcal{A}_{2\bar{O}}$ and $\mathcal{A}_{2O}$ are given in *Theorem 1*, and $\mathcal{A}_2$ is given in *Lemma 1*.

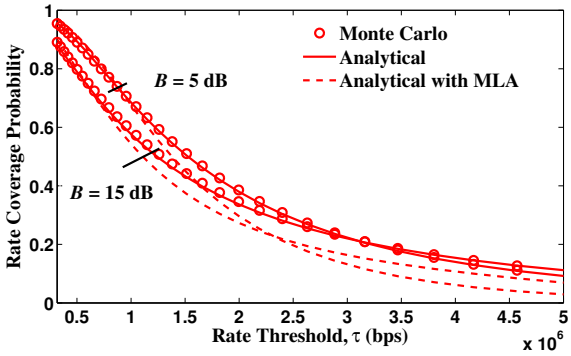


Fig. 3. Rate coverage probability vs. rate threshold τ for different bias factors B, at $\alpha_1=\alpha_2=4$, $\frac{P_1}{P_2}=10$ dB, $N_1=8$, $N_2=4$, $U=4$, $W=10$ MHz, $\lambda_1=0.0001$ nodes/m^2, and $\lambda_2=0.0005$ nodes/m^2.

Proof: Follows by replacing $L_{0,j}$ in (18) with $\mathrm{E}\left[L_{0,j}\right]$, where $j=1,2$. Note that $\mathrm{E}\left[L_{0,1}\right]$ is given in [5], and $\mathrm{E}\left[L_{0,2}\right]$ can be calculated using a similar approach. ■

Fig. 3 plots the rate coverage probability of the IN scheme vs. rate threshold τ for different bias factors B. We see from Fig. 3 that the 'Analytical' curves (i.e., $\mathcal{R}_{\mathrm{IN}}(U,\tau)$ in *Theorem 2*) closely match with the 'Monte Carlo' curves, although $\mathcal{R}_{\mathrm{IN}}(U,\tau)$ is derived based on some approximations, as illustrated in Section III-B. Moreover, we observe that the 'Analytical with MLA' curves (i.e., $\bar{\mathcal{R}}_{\mathrm{IN}}(U,\tau)$ in *Corollary 1*) are close to the 'Analytical' curves (i.e., $\mathcal{R}_{\mathrm{IN}}(U,\tau)$ in *Theorem 2*), especially when τ is not very large. Hence, for analytical tractability, we will investigate the rate coverage probability with MLA $\bar{\mathcal{R}}_{\mathrm{IN}}(U,\tau)$ in the remaining part of this paper.

V. Rate Coverage Probability Optimization of Interference Nulling

In this section, we consider the rate coverage probability optimization of the IN scheme. For a fixed bias factor B, the optimal design parameter $U^*(\tau)$, which maximizes the (overall) rate coverage probability $\bar{\mathcal{R}}_{\mathrm{IN}}(U,\tau)$, is defined as follows:[12]

$$U^*(\tau) \triangleq \arg\max_{U\in\{0,1,\ldots,N_1-1\}}\bar{\mathcal{R}}_{\mathrm{IN}}(U,\tau). \tag{24}$$

Note that (24) is an integer programming problem with a very complicated objective function $\bar{\mathcal{R}}_{\mathrm{IN}}(U,\tau)$. It is thus difficult to obtain the closed-form optimal solution $U^*(\tau)$ to the problem in (24). To address this challenge, in the following, we first characterize the rate coverage probability change when the design parameter is changed from $U-1$ to U. Then, based on it, we study some properties of $U^*(\tau)$ for the small and general rate threshold regimes, respectively.

[12]Please note that some constraints can be imposed to the optimization problem in (24), e.g., the rate coverage probability constraints of different sets of users. Our analysis of the rate coverage probability is the basis for formulating the new optimization problems, and our analysis of the optimality properties may be extended to analyze the new optimization problems. However, we will defer the detailed investigation of the new optimization problems to future work.

A. Rate Coverage Probability Change

First, we define $\Delta\bar{\mathcal{R}}_{\mathrm{IN}}(U,\tau) \triangleq \bar{\mathcal{R}}_{\mathrm{IN}}(U,\tau) - \bar{\mathcal{R}}_{\mathrm{IN}}(U-1,\tau)$ as the change of $\bar{\mathcal{R}}_{\mathrm{IN}}(U,\tau)$ when the design parameter is changed from $U-1$ to U, where $U \in \{1,\ldots,N_1-1\}$. By (23), $\Delta\bar{\mathcal{R}}_{\mathrm{IN}}(U,\tau)$ can be decomposed into three parts as follows:

$$\begin{aligned}\Delta\bar{\mathcal{R}}_{\mathrm{IN}}(U,\tau) =& \mathcal{A}_1\Delta\bar{\mathcal{R}}_{\mathrm{IN},1}(U,\tau) + \mathcal{A}_{2\bar{O}}\Delta\bar{\mathcal{R}}_{\mathrm{IN},2\bar{O}}(\tau) \\ &+ \mathcal{A}_{2O}\Delta\bar{\mathcal{R}}_{\mathrm{IN},2O}(U,\tau) \end{aligned} \quad (25)$$

where[13] $\Delta\bar{\mathcal{R}}_{\mathrm{IN},1}(U,\tau) \triangleq \bar{\mathcal{R}}_{\mathrm{IN},1}(U,\tau) - \bar{\mathcal{R}}_{\mathrm{IN},1}(U-1,\tau)$ denotes the rate coverage probability change of a macro-user, $\Delta\bar{\mathcal{R}}_{\mathrm{IN},2\bar{O}}(\tau) \triangleq \bar{\mathcal{R}}_{\mathrm{IN},2\bar{O}}(\tau) - \bar{\mathcal{R}}_{\mathrm{IN},2\bar{O}}(\tau)$ denotes the rate coverage probability change of an unoffloaded pico-user, and $\Delta\bar{\mathcal{R}}_{\mathrm{IN},2O}(U,\tau) \triangleq \bar{\mathcal{R}}_{\mathrm{IN},2O}(U,\tau) - \bar{\mathcal{R}}_{\mathrm{IN},2O}(U-1,\tau) = \left(\Pr\left(\mathcal{E}_{2OC,0}(U)\right) - \Pr\left(\mathcal{E}_{2OC,0}(U-1)\right)\right)\left(\bar{\mathcal{R}}_{\mathrm{IN},2OC}(\tau) - \bar{\mathcal{R}}_{\mathrm{IN},2O\bar{C}}(\tau)\right)$ denotes the rate coverage probability change of an offloaded user. Here, $\bar{\mathcal{R}}_{\mathrm{IN},2O}(U,\tau) \triangleq \Pr\left(\mathcal{E}_{2OC,0}(U)\right)\bar{\mathcal{R}}_{\mathrm{IN},2OC}(\tau) + \left(1-\Pr\left(\mathcal{E}_{2OC,0}(U)\right)\right)\bar{\mathcal{R}}_{\mathrm{IN},2O\bar{C}}(\tau)$ denotes the rate coverage probability of an offloaded user.

Next, we analyze $\Delta\bar{\mathcal{R}}_{\mathrm{IN},1}(U,\tau)$, $\Delta\bar{\mathcal{R}}_{\mathrm{IN},2\bar{O}}(\tau)$ and $\Delta\bar{\mathcal{R}}_{\mathrm{IN},2O}(U,\tau)$ in the following lemma:

Lemma 3: i) $\Delta\bar{\mathcal{R}}_{\mathrm{IN},1}(U,\tau) < 0$, ii) $\Delta\bar{\mathcal{R}}_{\mathrm{IN},2\bar{O}}(\tau) = 0$, and iii) $\Delta\bar{\mathcal{R}}_{\mathrm{IN},2O}(U,\tau) > 0$.

Proof: See Appendix C. ∎

Based on *Lemma 3*, $\Delta\bar{\mathcal{R}}_{\mathrm{IN}}(U,\tau)$ can be simplified as follows:

$$\Delta\bar{\mathcal{R}}_{\mathrm{IN}}(U,\tau) = \mathcal{A}_{2O}\Delta\bar{\mathcal{R}}_{\mathrm{IN},2O}(U,\tau) - \mathcal{A}_1\left|\Delta\bar{\mathcal{R}}_{\mathrm{IN},1}(U,\tau)\right| . \quad (26)$$

where $\mathcal{A}_{2O}\Delta\bar{\mathcal{R}}_{\mathrm{IN},2O}(U,\tau)$ and $\mathcal{A}_1\left|\Delta\bar{\mathcal{R}}_{\mathrm{IN},1}(U,\tau)\right|$ are referred to as the "gain" and the "penalty" of the IN scheme, respectively. Whether $\Delta\bar{\mathcal{R}}_{\mathrm{IN}}(U,\tau)$ is positive or not depends on whether the "gain" dominates the "penalty" or not. Therefore, to maximize $\bar{\mathcal{R}}_{\mathrm{IN}}(U,\tau)$, we can study the properties of $\Delta\bar{\mathcal{R}}_{\mathrm{IN}}(U,\tau)$ in (26) w.r.t. U by comparing $\mathcal{A}_{2O}\Delta\bar{\mathcal{R}}_{\mathrm{IN},2O}(U,\tau)$ and $\mathcal{A}_1\left|\Delta\bar{\mathcal{R}}_{\mathrm{IN},1}(U,\tau)\right|$.

B. Rate Coverage Probability Optimization for Small τ

In this part, we obtain $U^*(\tau)$ when $\tau \to 0$ by comparing $\mathcal{A}_{2O}\Delta\bar{\mathcal{R}}_{\mathrm{IN},2O}(U,\tau)$ and $\mathcal{A}_1\left|\Delta\bar{\mathcal{R}}_{\mathrm{IN},1}(U,\tau)\right|$. First, we characterize $\left|\Delta\bar{\mathcal{R}}_{\mathrm{IN},1}(U,\tau)\right|$ and $\Delta\bar{\mathcal{R}}_{\mathrm{IN},2O}(U,\tau)$.

We first obtain the order of the rate coverage probability loss of a macro-user, i.e., $\left|\Delta\bar{\mathcal{R}}_{\mathrm{IN},1}(U,\tau)\right|$, which is shown in the following proposition:

Proposition 1: When $\tau \to 0$, we have[14] $\left|\Delta\bar{\mathcal{R}}_{\mathrm{IN},1}(U,\tau)\right| = \Theta\left(\tau^{N_1-U}\right)$.

Proof: See Appendix D. ∎

Proposition 1 shows that when $\tau \to 0$, the rate coverage probability loss of a macro-user, i.e., $\left|\Delta\bar{\mathcal{R}}_{\mathrm{IN},1}(U,\tau)\right|$ in (26), decreases as $N_1 - U$ increases, and the decrease is in the order of τ^{N_1-U}. Furthermore, for a fixed N_1, $\left|\Delta\bar{\mathcal{R}}_{\mathrm{IN},1}(U,\tau)\right|$ increases as U increases.

[13]From now on, we make explicit the dependence of $\bar{\mathcal{R}}_{\mathrm{IN},1}(\tau)$ on U.

[14]$f(x) = \Theta\left(g(x)\right)$ means that $\lim_{x\to 0}\frac{f(x)}{g(x)} = c$ where $0 < c < \infty$.

Next, we characterize the rate coverage probability gain achieved by an offloaded user, i.e., $\Delta\bar{\mathcal{R}}_{\mathrm{IN},2O}(U,\tau)$. Using a mean interference-to-signal ratio based approach proposed in [26], we obtain the order of $\Delta\bar{\mathcal{R}}_{\mathrm{IN},2O}(U,\tau)$, which is shown as follows:

Proposition 2: When $\tau \to 0$, we have $\Delta\bar{\mathcal{R}}_{\mathrm{IN},2O}(U,\tau) = \Theta\left(\tau^{N_2}\right)$.

Proof: See Appendix E. ∎

From *Proposition 2*, we see that when $\tau \to 0$, the rate coverage probability gain of an offloaded user, i.e., $\Delta\bar{\mathcal{R}}_{\mathrm{IN},2O}(U,\tau)$ in (26), decreases as the number of antennas at each pico-BS N_2 increases, and the decrease is in the order of τ^{N_2}.

According to (26), *Proposition 1* and *Proposition 2*, and noting that $\mathcal{A}_{2O}$ and $\mathcal{A}_1$ are independent of τ, we have the following theorem:

Theorem 3: When $\tau \to 0$, the optimal design parameter $U^*(\tau) \to U_0^*$, where $U_0^* \in \{N_1 - N_2 - 1, N_1 - N_2\}$.

Proof: See Appendix F. ∎

Theorem 3 shows that when $\tau \to 0$, the optimal design parameter $U^*(\tau)$ converges to a fixed value in the set $\{N_1 - N_2 - 1, N_1 - N_2\}$, which is only related to the number of antennas at each macro-BS and each pico-BS. The optimality structure of $U^*(\tau)$ in *Theorem 3* is helpful for us to understand the insights. In addition, the complexity of obtaining $U^*(\tau)$ based on the optimality structure in *Theorem 3* (i.e., $U^*(\tau) \in \{N_1-N_2, N_1-N_2-1\}$) is much lower than that of exhaustive search (i.e., $U^*(\tau) \in \{0,\ldots,N_1-1\}$).

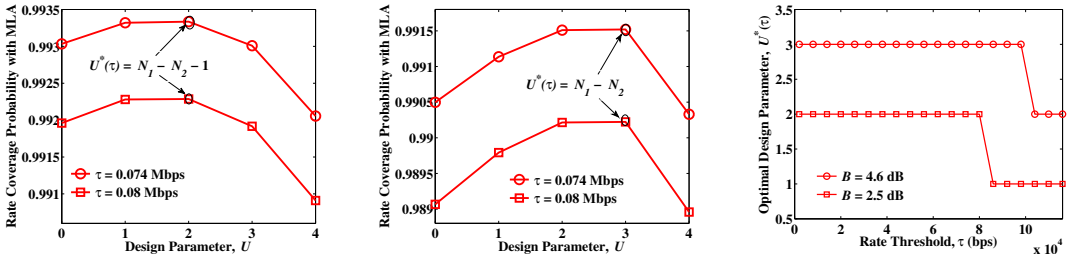
Figs. 4(a) and 4(b) plot $\bar{\mathcal{R}}_{\mathrm{IN}}(U,\tau)$ vs. the design parameter U for different bias factors B. We see that when $B = 2.5$ dB, $U^*(\tau) = N_1 - N_2 - 1 = 2$; when $B = 4.6$ dB, $U^*(\tau) = N_1 - N_2 = 3$ (note that $U^*(\tau)$ increases with B). Moreover, Fig. 4(c) plots the optimal design parameter $U^*(\tau)$ vs. rate threshold τ for different bias factors B, from which we see that $U^*(\tau)$ converges to a fixed value $U_0^* \in \{N_1-N_2-1, N_1-N_2\}$ when τ is sufficiently small (e.g., $\tau < 0.1$ Mbps for $B = 4.6$ dB). These observations verify *Theorem 3*.

C. Rate Coverage Probability Optimization for General τ

In this part, we discuss the optimality property of $U^*(\tau)$ for general τ. Different from the case for small τ, for general τ, $U^*(\tau)$ also depends on other system parameters besides N_1 and N_2. Fig. 5 plots the rate coverage probability with MLA vs. U for different bias factors B. We can see that besides $N_1 - N_2 - 1$ and $N_1 - N_2$, $U^*(\tau)$ can also take other values in set $\{0, 1, \ldots, N_1 - 1\}$. In particular, we see that $U^*(\tau)$ can be 0 (at $B = 2$ dB), 2 (at $B = 5$ dB), and 4 (at $B = 10$ dB). Interestingly, similar to the case for small τ in Fig. 4, from Fig. 5, we can also see that for general τ, $U^*(\tau)$ increases with the bias factor B.

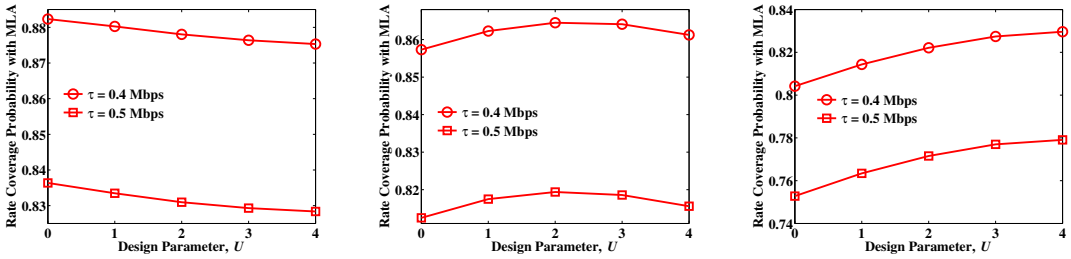
VI. Rate Coverage Probability Comparison

In this section, we first analyze the rate coverage probabilities of the multi-antenna version of the existing simple offloading scheme without interference management (i.e., $U = 0$) [2] and the multi-antenna version of the existing ABS scheme in 3GPP-LTE [5]. Then, we compare the rate coverage probability of each user type and the overall rate



(a) $B = 2.5$ dB (b) $B = 4.6$ dB (c) $U^*(\tau)$ vs. τ

Fig. 4. Optimal design parameter $U^*(\tau)$, at $\alpha_1 = \alpha_2 = 3$, $\frac{P_1}{P_2} = 10$ dB, $W = 10 \times 10^6$ Hz, $N_1 = 5$, $N_2 = 2$, $\lambda_u = 0.01$ nodes/m^2, $\lambda_1 = 0.0001$ nodes/m^2, and $\lambda_2 = 0.0015$ nodes/m^2.



(a) $B = 2$ dB (b) $B = 5$ dB (c) $B = 10$ dB

Fig. 5. Rate coverage probability with MLA vs. U for different bias factors B and general τ, at $\alpha_1 = \alpha_2 = 3$, $\frac{P_1}{P_2} = 10$ dB, $N_1 = 5$, $N_2 = 2$, $W = 10 \times 10^6$ Hz, $\lambda_1 = 0.0001$ nodes/m^2, $\lambda_2 = 0.0015$ nodes/m^2, and $\lambda_u = 0.01$ nodes/m^2.

coverage probability of the IN scheme with those of the simple offloading scheme and ABS.

A. Rate Coverage Probability Analysis for Simple Offloading Scheme and ABS

1) Analysis for Simple Offloading Scheme (i.e., $U = 0$) [2]: Note that $U = 0$ is a special case of the IN scheme. As such, by letting $U = 0$ in *Theorem 2* and *Corollary 1*, we can obtain the rate coverage probability and its MLA of the simple offloading scheme, respectively. In addition, from the resultant expressions, we can know the rate coverage probabilities of the macro-users $\bar{\mathcal{R}}_{U=0,1}(\tau)$, the unoffloaded pico-users $\bar{\mathcal{R}}_{U=0,2\bar{O}}(\tau)$ and the offloaded users $\bar{\mathcal{R}}_{U=0,2O}(\tau)$, where $\bar{\mathcal{R}}_{U=0,k}(\tau) = \bar{\mathcal{R}}_{\mathrm{IN},k}(0,\tau)$ $(k \in \{1, 2O\})$ and $\bar{\mathcal{R}}_{U=0,2\bar{O}}(\tau) = \mathcal{R}_{\mathrm{IN},2\bar{O}}(\tau)$. Due to page limit, we omit the expressions of the rate coverage probability and its MLA of the simple offloading scheme. Note that [27] also derived the rate coverage probability and its MLA of the macro-users and the pico-users under the simple offloading scheme in large multi-antenna HetNets. However, they did not further obtain the results for unoffloaded pico-users and offloaded users.

2) Analysis for ABS [5]: We consider ABS with a given design parameter $\eta \in (0, 1)$. Specifically, in ABS, η fraction of the resource W is utilized by pico-BSs to serve offloaded users only, while the remaining $1 - \eta$ fraction of the resource W is utilized simultaneously by macro-BSs and pico-BSs to serve macro-users and unoffloaded pico-users, respectively [5]. In other words, to avoid interference to offloaded users from all the macro-BSs, the resource used at each BS to serve its associated users in ABS is reduced due to the resource partition (parameterized by η). Note that different from ABS, in the IN scheme and the simple offloading scheme, each BS utilizes all the resource W to serve its associated users. Similar to (17), the rate coverage probability of $u_0 \in \mathcal{U}_k$ $\left(k \in \{1, 2\bar{O}, 2O\}\right)$ under ABS is defined as $\mathcal{R}_{\mathrm{ABS},k}(\eta,\tau) \triangleq \Pr\left(R_{\mathrm{ABS},k,0} > \tau | u_0 \in \mathcal{U}_k\right) = \Pr\left(\frac{\eta_k W}{L_{0,k}} \log_2\left(1 + \mathrm{SIR}_{\mathrm{ABS},k,0}\right) > \tau | u_0 \in \mathcal{U}_k\right)$, where $R_{\mathrm{ABS},k,0}$ and $\mathrm{SIR}_{\mathrm{ABS},k,0}$ denote the rate and SIR of $u_0 \in \mathcal{U}_k$ in ABS, respectively, $\eta_1 = \eta_{2\bar{O}} = 1 - \eta$, and $\eta_{2O} = \eta$. The rate coverage probability of ABS under the design parameter η can be written as: $\mathcal{R}_{\mathrm{ABS}}(\eta,\tau) \triangleq \Pr\left(R_{\mathrm{ABS},0} > \tau\right) = \sum_{k\in\{1,2\bar{O},2O\}} \mathcal{A}_k \mathcal{R}_{\mathrm{ABS},k}(\eta,\tau)$, where $R_{\mathrm{ABS},0}$ is the rate of u_0 (which can be in any user set) in ABS. Applying similar methods in calculating the rate coverage probability and its MLA of the IN scheme in *Theorem 2* and *Corollary 1*, we can obtain the rate coverage probability and its MLA of ABS, respectively. Due to page limit, we omit the rate coverage probability expression of ABS, and only focus on its MLA. Note that as shown in Section IV-B, the rate coverage probability with MLA, which has a simpler expression, is sufficiently accurate. The rate coverage probability with MLA for ABS $\bar{\mathcal{R}}_{\mathrm{ABS}}(\eta,\tau)$ under $\eta \in (0, 1)$ is given as follows:

Proposition 3: The rate coverage probability with MLA of ABS under $\eta \in (0, 1)$ is

$$\begin{aligned}\bar{\mathcal{R}}_{\mathrm{ABS}}(\eta,\tau) =& \mathcal{A}_1 \bar{\mathcal{R}}_{\mathrm{ABS},1}(\eta,\tau) + \mathcal{A}_{2\bar{O}} \bar{\mathcal{R}}_{\mathrm{ABS},2\bar{O}}(\eta,\tau) \\ &+ \mathcal{A}_{2O} \bar{\mathcal{R}}_{\mathrm{ABS},2O}(\eta,\tau) \end{aligned} \quad (27)$$

where $\mathcal{A}_1$, $\mathcal{A}_{2\bar{O}}$ and $\mathcal{A}_{2O}$ are given in *Theorem 1*, and

$$\bar{\mathcal{R}}_{\mathrm{ABS},1}(\eta,\tau)=\int_0^{\infty}\sum_{n=0}^{N_1-1}\frac{1}{n!}\sum_{n_1=0}^{n}\binom{n}{n_1}\tilde{\mathcal{L}}_{I_1}^{(n_1)}(s,y)\Big|_{s=\tilde{\beta}_1(\eta)y^{\alpha_1}}\times\tilde{\mathcal{L}}_{I_2}^{(n-n_1)}\left(s,(P_2B/P_1)^{\frac{1}{\alpha_2}}y^{\frac{\alpha_1}{\alpha_2}}\right)\Big|_{s=\tilde{\beta}_1(\eta)y^{\alpha_1}\frac{P_2}{P_1}}f_{Y_1}(y)\mathrm{d}y\,, \tag{28}$$

$$\bar{\mathcal{R}}_{\mathrm{ABS},2\bar{O}}(\eta,\tau)=\int_0^{\infty}\mathcal{S}_{\mathrm{IN},2\bar{O},Y_2}\left(y,\tilde{\beta}_{2\bar{O}}(\eta)\right)f_{Y_2}(y)\mathrm{d}y\,, \tag{29}$$

$$\bar{\mathcal{R}}_{\mathrm{ABS},2O}(\eta,\tau)=\int_0^{\infty}\sum_{n=0}^{N_2-1}\frac{1}{n!}\tilde{\mathcal{L}}_{I_2}^{(n)}(s,y)\Big|_{s=\tilde{\beta}_{2O}(\eta)y^{\alpha_2}}\times f_{Y_{2O}}(y)\mathrm{d}y\,. \tag{30}$$

Here, $\tilde{\beta}_1(\eta)=2^{\frac{\tau\mathrm{E}[L_{0,1}]}{W(1-\eta)}}-1$, $\tilde{\beta}_{2\bar{O}}(\eta)=2^{\frac{\tau\mathrm{E}\left[L_{0,2\bar{O}}\right]}{W(1-\eta)}}-1$, $\tilde{\beta}_{2O}(\eta)=2^{\frac{\tau\mathrm{E}\left[L_{0,2O}\right]}{W\eta}}-1$, $f_{Y_{2O}}(y)=\frac{2\pi\lambda_2}{\mathcal{A}_{2O}}\Big(\exp\left(-\pi\lambda_1\left(P_1/(P_2B)\right)^{\frac{2}{\alpha_1}}y^{\frac{2\alpha_2}{\alpha_1}}\right)-\exp\left(-\pi\lambda_1\left(P_1/P_2\right)^{\frac{2}{\alpha_1}}y^{\frac{2\alpha_2}{\alpha_1}}\right)\Big)y\exp\left(-\pi\lambda_2y^2\right)$ [5], $\mathrm{E}\left[L_{0,2\bar{O}}\right]=1+1.28\frac{\lambda_u\mathcal{A}_{2\bar{O}}}{\lambda_2}$, and $\mathrm{E}\left[L_{0,2O}\right]=1+1.28\frac{\lambda_u\mathcal{A}_{2O}}{\lambda_2}$.

Proof: Similar to the proof of *Corollary 1*. Omitted due to page limit. ■

Note that the rate coverage probability with MLA of multi-antenna ABS shown in *Proposition 3* is derived using higher order derivatives of the Laplace transform of the aggregate interference, and can be treated as an extension of the single-antenna result derived using the Laplace transform of the aggregate interference in [5].

B. Rate Coverage Probability Comparison for Each User Type

In this part, we compare the rate coverage probability of $u_0\in\mathcal{U}_k$ $\left(k\in\left\{1,2\bar{O},2O\right\}\right)$ in the IN scheme (under a given $U\in\{0,1,\ldots,N_1-1\}$) with those in the simple offloading scheme (i.e., $U=0$) and ABS (under a given $\eta\in(0,1)$), respectively, for a fixed bias factor B.

1) Comparison with simple offloading scheme [2]: First, we compare the rate coverage probability of $u_0\in\mathcal{U}_k$ $\left(k\in\left\{1,2\bar{O},2O\right\}\right)$ in the IN scheme with that in the simple offloading scheme. We can easily show the following lemma:

Lemma 4: For all $U\in\{0,1,\ldots,N_1-1\}$, we have: i) $\bar{\mathcal{R}}_{\mathrm{IN},1}(U,\tau)\le\bar{\mathcal{R}}_{U=0,1}(\tau)$, ii) $\bar{\mathcal{R}}_{\mathrm{IN},2\bar{O}}(\tau)=\bar{\mathcal{R}}_{U=0,2\bar{O}}(\tau)$, and iii) $\bar{\mathcal{R}}_{\mathrm{IN},2O}(U,\tau)\ge\bar{\mathcal{R}}_{U=0,2O}(\tau)$. The equalities in i) and ii) hold i.f.f. $U=0$.

Now we compare the IN scheme under $U>0$ with the simple offloading scheme (i.e., $U=0$). *Lemma 4* can be interpreted below: i) the IN scheme achieves a smaller rate coverage probability for $u_0\in\mathcal{U}_1$, since the DoF used to serve u_0 are reduced by $\min\left(U,u_{2OC,0}\right)$; ii) the IN scheme achieves the same rate coverage probability of $u_0\in\mathcal{U}_{2\bar{O}}$ as the simple offloading scheme, since $\bar{\mathcal{R}}_{\mathrm{IN},2\bar{O}}(\tau)$ is independent of U; iii) the IN scheme achieves a larger rate coverage probability for $u_0\in\mathcal{U}_{2O}$, since $\min\left(U,u_{2OC,0}\right)$ DoF at the nearest macro-BS of u_0 is used to avoid dominant macro-interference to its $\min\left(U,u_{2OC,0}\right)$ IN offloaded users.

2) Comparison with ABS [5]: Now, we compare the rate coverage probability of $u_0\in\mathcal{U}_k$ $(k\in\{1,2\bar{O},2O\})$ in the IN scheme with that in ABS, which is summarized in the following:

Lemma 5: i) A sufficient condition for $\bar{\mathcal{R}}_{\mathrm{IN},1}(U,\tau)>\bar{\mathcal{R}}_{\mathrm{ABS},1}(\eta,\tau)$ when $N_1,U\to\infty$ with $\frac{U}{N_1}\to\kappa\in(0,1)$ and $\tau\to0$ is $\kappa<\eta$; ii) the necessary and sufficient condition for $\bar{\mathcal{R}}_{\mathrm{IN},2\bar{O}}(\tau)>\bar{\mathcal{R}}_{\mathrm{ABS},2\bar{O}}(\eta,\tau)$ is $\frac{1}{\mathrm{E}[L_{0,2}]}>\frac{1-\eta}{\mathrm{E}\left[L_{0,2\bar{O}}\right]}$; iii) a necessary condition for $\bar{\mathcal{R}}_{\mathrm{IN},2O}(U,\tau)>\bar{\mathcal{R}}_{\mathrm{ABS},2O}(\eta,\tau)$ is $\frac{1}{\mathrm{E}[L_{0,2}]}>\frac{\eta}{\mathrm{E}[L_{0,2O}]}$.

Proof: See Appendix G. ■

Note that the rate coverage probability of $u_0\in\mathcal{U}_k$ $\left(k\in\left\{1,2\bar{O},2O\right\}\right)$ depends on both the SIR of u_0 and the average resource used to serve u_0. Thus, *Lemma 5* can be understood below: i) the IN scheme (with DoF fraction $1-\kappa$ and resource fraction 1 for scheduled u_0) achieves a larger rate coverage probability for $u_0\in\mathcal{U}_1$ than ABS (with DoF fraction 1 and resource fraction $1-\eta$ for scheduled u_0) if $\kappa<\eta$; ii) The IN scheme achieves a larger rate coverage probability for $u_0\in\mathcal{U}_{2\bar{O}}$ i.f.f. the average resource (i.e., $\frac{1}{\mathrm{E}[L_{0,2}]}$ under MLA) used to serve u_0 in the IN scheme is larger than that (i.e., $\frac{1-\eta}{\mathrm{E}\left[L_{0,2\bar{O}}\right]}$ under MLA) in ABS, as the SIRs of $u_0\in\mathcal{U}_{2\bar{O}}$ are the same in both schemes; iii) Note that the SIR of $u_0\in\mathcal{U}_{2O}$ in the IN scheme is worse than that in ABS, as $u_0\in\mathcal{U}_{2O}$ does not experience any macro-interference in ABS, while $u_0\in\mathcal{U}_{2O}$ still experiences macro-interference (except the dominant one) in the IN scheme. Hence, it is possible for the IN scheme to achieve a larger rate coverage probability for $u_0\in\mathcal{U}_{2O}$ only when the average resource (i.e., $\frac{1}{\mathrm{E}[L_{0,2}]}$ under MLA) used to serve $u_0\in\mathcal{U}_{2O}$ in the IN scheme is larger than that (i.e., $\frac{\eta}{\mathrm{E}[L_{0,2O}]}$ under MLA) in ABS.

Fig. 6 plots the rate coverage probability with MLA of the IN scheme at $U=7$, and the rate coverage probability with MLA of ABS vs. η. Note that under the parameters in Fig. 6, we have: i) $\kappa=\frac{U}{N_1}=0.7$, ii) $1-\frac{\mathrm{E}\left[L_{0,2\bar{O}}\right]}{\mathrm{E}[L_{0,2}]}\approx0.09$, and iii) $\frac{\mathrm{E}[L_{0,2O}]}{\mathrm{E}[L_{0,2}]}\approx0.12$, with $\mathrm{E}\left[L_{0,2\bar{O}}\right]\approx28.57$, $\mathrm{E}\left[L_{0,2O}\right]\approx3.86$ and $\mathrm{E}\left[L_{0,2}\right]\approx31.43$ calculated according to *Proposition 3* and *Corollary 1*. From Fig. 6, we observe that i) $\eta>0.7$ is sufficient to achieve $\bar{\mathcal{R}}_{\mathrm{IN},1}(7,\tau)>\bar{\mathcal{R}}_{\mathrm{ABS},1}(\eta,\tau)$; ii) $\bar{\mathcal{R}}_{\mathrm{IN},2\bar{O}}(\tau)>\bar{\mathcal{R}}_{\mathrm{ABS},2\bar{O}}(\eta,\tau)$ i.f.f. $\eta>0.09\approx1-\frac{\mathrm{E}\left[L_{0,2\bar{O}}\right]}{\mathrm{E}[L_{0,2}]}$; iii) $\bar{\mathcal{R}}_{\mathrm{IN},2O}(7,\tau)>\bar{\mathcal{R}}_{\mathrm{ABS},2O}(\eta,\tau)$ when $\eta<0.1<0.12\approx\frac{\mathrm{E}[L_{0,2O}]}{\mathrm{E}[L_{0,2}]}$. These observations verify *Lemma 5*.

C. Overall Rate Coverage Probability Comparison

In this part, we compare the overall rate coverage probability of the *IN scheme under its optimal design parameter* $U^*(\tau)$ with those of the *simple offloading scheme without interference management (i.e., $U=0$)* [2] and the *multi-antenna version of ABS under its optimal design parameter* $\eta^*(\tau)\triangleq\arg\max_{\eta\in(0,1)}\mathcal{R}_{\mathrm{ABS}}(\eta,\tau)$ [5].

First, we compare the rate coverage probability of the IN scheme with that of the simple offloading scheme [2]. Based on the discussions of *Lemma 4*, we know that the IN scheme has the benefit of avoiding the dominant macro-interference to offloaded users. When B is sufficiently large (implying that

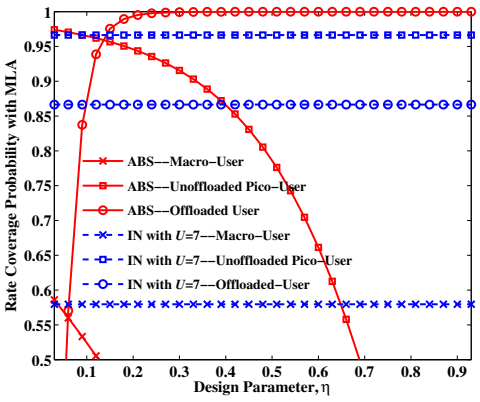
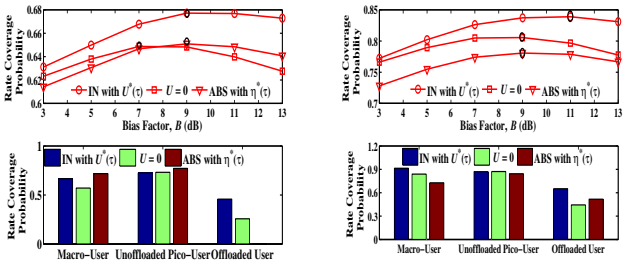


Fig. 6. Rate coverage probability vs. resource fraction η for ABS and IN, at $\frac{P_1}{P_2} = 13$ dB, $W = 10$ MHz, $\tau = 5 \times 10^5$ bps, $N_1 = 10$, $N_2 = 8$, $\lambda_1 = 0.00008$ nodes/m^2, $\lambda_2 = 0.001$ nodes/m^2, $\lambda_u = 0.03$ nodes/m^2, $\alpha_1 = 4.5$, $\alpha_2 = 4.7$, $B = 4$ dB, $\mathcal{A}_1 \approx 0.21$, $\mathcal{A}_{2\bar{O}} \approx 0.72$, and $\mathcal{A}_{2O} \approx 0.07$.



(a) $N_1 = 8$, $N_2 = 6$, $\eta^*(\tau) = 0.01$ at B^*_{ABS} (b) $N_1 = 18$, $N_2 = 16$, $\eta^*(\tau) = 0.19$ at B^*_{ABS}

Fig. 7. Rate coverage probability vs. bias factors B, at $\alpha_1 = 4.5$, $\alpha_2 = 4.7$, $\frac{P_1}{P_2} = 13$ dB, $W = 10 \times 10^6$ Hz, $\tau = 5 \times 10^5$ bps, $\lambda_1 = 0.00008$ nodes/m^2, $\lambda_2 = 0.001$ nodes/m^2, and $\lambda_u = 0.05$ nodes/m^2. In the figures on the top, the points at B^*_{IN}, $B^*_{U=0}$, and B^*_{ABS} are highlighted using black ellipse. In the figures at the bottom, the rate coverage probability of each user type in different schemes are plotted at B^*_{IN}, $B^*_{U=0}$, and B^*_{ABS}, respectively. Note that $\eta^*(\tau)$ of ABS is obtained by bisection method with N_1 iterations, while $U^*(\tau)$ of the IN scheme is obtained by exhaustive search over $\{0, 1, \ldots, N_1 - 1\}$.

$\mathcal{A}_{2O}$ is sufficiently large), sufficient offloaded users can benefit from the avoidance of the dominant macro-interference (i.e., the benefit is large). On the other hand, we also know that the loss of the IN scheme compared to the simple offloading scheme is caused by the reduction of the DoF used to serve the macro-users (i.e., at most $\frac{U}{N_1}$ reduction of the DoF fraction at each macro-BS in the IN scheme). Thus, when N_1 is relatively large (implying that the DoF fraction reduction $\frac{U}{N_1}$ is minor), the loss due to the DoF reduction is small. Therefore, when B and N_1 are relatively large (e.g., $B = 9$ dB and $N_1 = 8$ in Fig. 7(a)), the IN scheme can achieve a larger rate coverage probability than the simple offloading scheme.

Next, we compare the rate coverage probability of the IN scheme with that of ABS [5]. Based on the discussions of *Lemma 5*, we know that the benefit of the IN scheme compared to ABS is that it does not have (time or frequency) resource sacrifice. On the other hand, we also know that one loss of the IN scheme compared to ABS is due to the $\frac{U}{N_1}$ DoF fraction reduction (as discussed above). Thus, when N_1 is relatively large (implying that the DoF fraction reduction $\frac{U}{N_1}$ is minor), the loss due to the DoF reduction is small. Another loss of the IN scheme compared to ABS is caused by the macro-interference (except the dominant one), as the IN scheme only avoids the dominant macro-interference to offloaded users, while ABS avoids all the macro-interference to offloaded users. When α_1 is relatively large (implying that the dominant macro-interference is sufficiently strong compared to the remaining macro-interference), the loss due to the remaining macro-interference is small. Therefore, when N_1 and α_1 are relatively large (e.g., $N_1 = 8$ and $\alpha_1 = 4.5$ in Fig. 7(a)), the IN scheme can achieve a larger rate coverage probability than ABS.

The two figures on the top in Fig. 7 plot the rate coverage probability vs. the bias factor B for the IN scheme under $U^*(\tau)$, the simple offloading scheme, and ABS under $\eta^*(\tau)$. We see that the IN scheme achieves a larger rate coverage probability than both the simple offloading scheme and ABS when the bias factor B is relatively large.[15] In addition, we consider rate coverage probability maximization over B for these three schemes. We observe that the IN scheme achieves a larger rate coverage probability than both the simple offloading scheme and ABS at their optimal bias factors. Denote the optimal bias factors of the IN scheme, simple offloading scheme and ABS as B^*_{IN}, $B^*_{U=0}$ and B^*_{ABS}, respectively. We have the following observations for B^*_{IN}, $B^*_{U=0}$ and B^*_{ABS}. Firstly, B^*_{IN}, $B^*_{U=0}$ and B^*_{ABS} are all positive. This implies that the rate coverage probability can be improved by offloading users from the heavily loaded macro-cell tier to the lightly loaded pico-cell tier. Secondly, both B^*_{IN} and B^*_{ABS} can be larger than $B^*_{U=0}$. This implies that the IN scheme and ABS allow more users to be offloaded to the lightly loaded pico-cell tier than the simple offloading scheme, as the IN scheme and ABS can effectively improve the performance of the offloaded users.

We now further investigate the rate coverage probability of offloaded users, which is one of the main limiting factors for the performance of HetNets with offloading. In the IN scheme, the offloaded users do not have (time or frequency) resource sacrifice and dominant macro-interference. However, the offloaded users in ABS suffer from resource limitations, and the offloaded users in the simple offloading scheme suffer from strong interference caused by their dominant macro-interfererence. Hence, the offloaded users in the IN scheme can achieve a larger rate coverage probability than those in both the simple offloading scheme and ABS (e.g., when $\alpha_1 = 4.5$ and $\eta(\tau) = 0.01$ in Fig. 7(a)). The two figures at the bottom in Fig. 7 plot the rate coverage probability of three user types at B^*_{IN}, $B^*_{U=0}$, and B^*_{ABS}, respectively. We can clearly see that the offloaded user in the IN scheme achieves the largest rate coverage probability.

VII. Conclusions

In this paper, we investigated the IN scheme in downlink two-tier multi-antenna HetNets with offloading. Utilizing tools

[15]Note that the IN scheme may not provide gains in all scenarios, as suggested in Fig. 5.

from stochastic geometry, we first derived a tractable expression for the rate coverage probability of the IN scheme. Then, we considered the rate coverage probability optimization of the IN scheme by solving the optimal design parameter. Finally, we analyzed the performance of the multi-antenna version of the existing simple offloading scheme without interference management and the multi-antenna version of the existing ABS scheme, and compared the performance of the IN scheme with both of the two schemes in terms of the rate coverage probability of each user type and the overall rate coverage probability. Both the analytical and numerical results showed that the IN scheme can achieve good performance gains over both of the two schemes, especially in the large antenna regime.

APPENDIX

A. Proof of Lemma 1

We first note that i) the total number of scheduled pico-users are the same with the total number of pico-BSs, ii) the association area of pico-BSs is $\mathcal{A}_2$ fraction of the total area, and iii) the scheduled pico-users are only in the association area of pico-BSs. Hence, the effective density of the scheduled pico-users is $\frac{\lambda_2}{\mathcal{A}_2}$. Next, we approximate the scheduled pico-users as a homogeneous PPP, so that the number of scheduled pico-users in a fixed area is Poisson distributed with density $\frac{\lambda_2}{\mathcal{A}_2}$. Note that similar approximation approaches are utilized in [15, 24]. Obviously, the number of active offloaded users in a fixed area is also Poisson distributed with density $\frac{\lambda_2}{\mathcal{A}_2}$. Further, using the approach in [5], we can calculate the mean of the offloading area (where the offloaded users may reside) of a randomly selected macro-BS, which is $\frac{\mathcal{A}_{2O}}{\lambda_1}$. Finally, we obtain (10) by following similar steps in calculating the load p.m.f. in [5, 25]. Note that $\mathcal{A}_2$ and $\mathcal{A}_{2O}$ are given in [2] and [5], respectively.

B. Proof of Theorem 1

To prove *Theorem 1*, we first prove the following lemma related to the p.m.f. of the IN DoF.

Lemma 6: When $u_0 \in \mathcal{U}_1$, the p.m.f. of the IN DoF at the typical user's serving macro-BS is

$$\Pr\left(u_{2OC,0}=n\right)=\begin{cases}\Pr\left(U_{2O_a,0}=n\right), & \text{for } 0\le n<U\\ \sum_{u=n}^{\infty}\Pr\left(U_{2O_a,0}=u\right), & \text{for } n=U\end{cases}. \tag{31}$$

Proof: Based on *Lemma 1*, we can easily compute the p.m.f. of $u_{2OC,0}=\min\left(U,U_{2O_a,0}\right)$. ∎

Next, let R_{jk} denote the minimum possible distance between $u_0 \in \mathcal{U}_k$ ($k \in \mathcal{K}$) and its nearest interferer in the jth tier ($j = 1, 2$). Note that $\mathcal{S}_{\mathrm{IN},k}(\beta) = \mathrm{E}_{R_{1k},R_{2k}}\left[\mathcal{S}_{\mathrm{IN},k,R_{1k},R_{2k}}(r_{1k},r_{2k},\beta)\right]$, where $\mathcal{S}_{\mathrm{IN},k,R_{1k},R_{2k}}(r_{1k},r_{2k},\beta) \triangleq \Pr\Big(\mathrm{SIR}_{\mathrm{IN},k,0} > \beta | u_0 \in \mathcal{U}_k, R_{1k} = r_{1k}, R_{2k} = r_{2k}\Big)$ denotes the conditional SIR coverage probability[16]. To calculate $\mathcal{S}_{\mathrm{IN},k}(\beta)$, we first need to calculate $\mathcal{S}_{\mathrm{IN},k,R_{1k},R_{2k}}(r_{1k},r_{2k},\beta)$, which is provided as follows:

[16]When $u_0 \in \mathcal{U}_1$, we also condition on $u_{2OC,0}$. For notational simplicity, we do not make this dependence explicit.

TABLE II
PARAMETER VALUES UNDER THE IN SCHEME WHEN $u_0 \in \mathcal{U}_k$ WITH $k \in \mathcal{K}$

k	j_k	r_{1k}	r_{2k}	M_k
1	1	Y_1	$\left(\frac{P_2 B}{P_1}\right)^{\frac{1}{\alpha_2}} Y_1^{\frac{\alpha_1}{\alpha_2}}$	$N_1 - u_{2O,0}$
$2\bar{O}$	2	$\left(\frac{P_1}{P_2}\right)^{\frac{1}{\alpha_1}} Y_2^{\frac{\alpha_2}{\alpha_1}}$	Y_2	N_2
$2OC$	2	Y_1	Y_2	N_2
$2O\bar{C}$	2	Y_1	Y_2	N_2

Lemma 7: The conditional SIR coverage probability of $u_0 \in \mathcal{U}_k$ under the IN scheme is

$$\mathcal{S}_{\mathrm{IN},k,R_{1k},R_{2k}}(r_{1k},r_{2k},\beta)=\sum_{n=0}^{M_k-1}\mathcal{T}_{k,R_{1k},R_{2k}}\left(n,r_{1k},r_{2k},\beta\right) \tag{32}$$

where $k \in \mathcal{K}$ and

$$\mathcal{T}_{k,R_{1k},R_{2k}}\left(n,r_{1k},r_{2k},\beta\right)= \begin{cases} \frac{1}{n!}\sum_{n_1=0}^{n}\binom{n}{n_1}\tilde{\mathcal{L}}_{I_1}^{(n_1)}\left(s,r_{1k}\right)\Big|_{s=\beta Y_{j_k}^{\alpha_{j_k}}\frac{P_1}{P_{j_k}}} \\ \quad\times\tilde{\mathcal{L}}_{I_2}^{(n-n_1)}\left(s,r_{2k}\right)\Big|_{s=\beta Y_{j_k}^{\alpha_{j_k}}\frac{P_2}{P_{j_k}}}, \quad \text{if } k\in\{1,2\bar{O},2OC\} \\ \frac{1}{n!}\sum_{(q_a)_{a=1}^3\in\mathcal{Q}_3}\binom{n}{q_1,q_2,q_3}\tilde{\mathcal{L}}_{I_1}^{(q_1)}\left(s,r_{1k}\right)\Big|_{s=\beta Y_{j_k}^{\alpha_{j_k}}\frac{P_1}{P_{j_k}}}\left(\beta\frac{P_1 Y_{j_k}^{\alpha_{j_k}}}{P_{j_k} r_{1k}^{\alpha_1}}\right)^{q_3} \\ \quad\times\tilde{\mathcal{L}}_{I_2}^{(q_2)}\left(s,r_{2k}\right)\Big|_{s=\beta Y_{j_k}^{\alpha_{j_k}}\frac{P_2}{P_{j_k}}}\Gamma\left(q_3+1\right)\left(1+\beta\frac{P_1 Y_{j_k}^{\alpha_{j_k}}}{P_{j_k} r_{1k}^{\alpha_1}}\right)^{-(q_3+1)}, \\ \qquad\qquad \text{if } k=2O\bar{C} \end{cases} \tag{33}$$

Here, j_k, r_{1k}, r_{2k}, and M_k are given in Table II, $\mathcal{Q}_3 \triangleq \{(q_a)_{a=1}^3 | q_a \in \mathbb{N}^0, \sum_{a=1}^3 q_a = n\}$, and $\tilde{\mathcal{L}}_{I_j}^{(m)}\left(s,r_{jk}\right) = \mathcal{L}_{I_j}\left(s,r_{jk}\right)\sum_{(p_a)_{a=1}^m\in\mathcal{M}_m}\frac{m!}{\prod_{a=1}^m p_a!}$ $\times\prod_{a=1}^m\left(\frac{2\pi}{\alpha_j}\lambda_j s^{\frac{2}{\alpha_j}}B'\left(1+\frac{2}{\alpha_j},a-\frac{2}{\alpha_j},\frac{1}{1+s r_{jk}^{-\alpha_j}}\right)\right)^{p_a}$ ($j \in \{1,2\}$), where $\mathcal{L}_{I_j}\left(s,r_{jk}\right) = \exp\left(-\frac{2\pi}{\alpha_j}\lambda_j s^{\frac{2}{\alpha_j}}B'\left(\frac{2}{\alpha_j},1-\frac{2}{\alpha_j},\frac{1}{1+s r_{jk}^{-\alpha_j}}\right)\right)$,[17] $\mathcal{M}_m \triangleq \left\{(p_a)_{a=1}^m | p_a \in \mathbb{N}^0, \sum_{a=1}^m a\cdot p_a = m\right\}$, and $B'(a,b,z) \triangleq \int_z^1 u^{a-1}(1-u)^{b-1}\mathrm{d}u$ $(0<z<1)$ [27].

Proof:

1) $k \in \{1, 2\bar{O}, 2OC\}$: When $k \in \{1, 2\bar{O}, 2OC\}$, based on (3), (5), and (7), we have

$$\begin{aligned}&\mathcal{S}_{\mathrm{IN},k,R_{1k},R_{2k}}(r_{1k},r_{2k},\beta)\\ &=\mathrm{E}_{I_1,I_2}\left[\Pr\left(\left|\mathbf{h}_{j_k,00}^{\dagger}\mathbf{f}_{j_k,0}\right|^2>\beta Y_{j_k}^{\alpha_{j_k}}\left(\frac{P_1}{P_{j_k}}I_1+\frac{P_2}{P_{j_k}}I_2\right)\right)\right]\\ &\overset{(a)}{=}\sum_{n=0}^{M_k-1}\frac{\left(-\beta Y_{j_k}^{\alpha_{j_k}}\right)^n}{n!}\sum_{n_1=0}^{n}\binom{n}{n_1}\left(\frac{P_1}{P_{j_k}}\right)^{n_1}\mathcal{L}_{I_1}^{(n_1)}\left(s,r_{1k}\right)\Big|_{s=\beta Y_{j_k}^{\alpha_{j_k}}\frac{P_1}{P_{j_k}}}\\ &\quad\times\left(\frac{P_2}{P_{j_k}}\right)^{n-n_1}\mathcal{L}_{I_2}^{(n-n_1)}\left(s,r_{2k}\right)\Big|_{s=\beta Y_{j_k}^{\alpha_{j_k}}\frac{P_2}{P_{j_k}}}\end{aligned} \tag{34}$$

[17]$\mathcal{L}_{I_j}\left(s,r_{jk}\right)$ is the Laplace transform of the aggregated interference $I_j = \sum_{\ell\in\Phi(\lambda_j)\setminus B(0,r_{jk})}\frac{\left|\mathbf{h}_{j,\ell 0}^{\dagger}\mathbf{f}_{j,\ell}\right|^2}{\left|D_{j,\ell 0}\right|^{\alpha_j}}$ from the jth tier.

where (a) is obtained by noting that $\left|\mathbf{h}_{j_k,00}^{\dagger}\mathbf{f}_{j_k,0}\right|^2 \overset{d}{\sim} \text{Gamma}(M_k,1)$, using binomial theorem, and noting that $\text{E}_{I_j}\left[I_j^n \exp\left(-sI_j\right)\right] = (-1)^n \mathcal{L}_{I_j}^{(n)}(s, r_{jk})$.

We now calculate the Laplace transform $\mathcal{L}_{I_1}(s, r_{1k})$ and its higher order derivative $\mathcal{L}_{I_1}^{(m)}(s, r_{1k})$. Firstly, let $G_{1,\ell} \triangleq \left|\mathbf{h}_{1,\ell 0}^{\dagger}\mathbf{f}_{1,\ell}\right|^2$. Then, $\mathcal{L}_{I_1}(s, r_{1k})$ can be calculated as follows:

$$\begin{aligned}
&\mathcal{L}_{I_1}(s, r_{1k}) \\
=&\text{E}_{\Phi(\lambda_1)}\left[\prod_{\ell\in\Phi(\lambda_1)\setminus B(0,r_{1k})} \text{E}_{G_{1,\ell}}\left[\exp\left(-s\frac{1}{|D_{1,\ell 0}|^{\alpha_j}}G_{1,\ell}\right)\right]\right] \\
\overset{(a)}{=}&\exp\left(-2\pi\lambda_1\int_{r_{1k}}^{\infty}\left(1-\frac{1}{1+sr^{-\alpha_1}}\right)r\text{d}r\right) \\
\overset{(b)}{=}&\exp\left(-\frac{2\pi}{\alpha_1}\lambda_1 s^{\frac{2}{\alpha_1}}\int_{\frac{1}{1+sr_{1k}^{-\alpha_1}}}^{1}(1-w)^{-\frac{2}{\alpha_1}}w^{-1+\frac{2}{\alpha_1}}\text{d}w\right) \quad (35)
\end{aligned}$$

where (a) is obtained by utilizing the probability generating functional of PPP [19], (b) is obtained by first replacing $s^{-\frac{1}{\alpha_1}}r$ with t, and then replacing $\frac{1}{1+t^{-\alpha_1}}$ with w.

Next, we calculate $\mathcal{L}_{I_1}^{(m)}(s, r_{1k})$ based on (35). Utilizing Faà di Bruno's formula [28], we have

$$\begin{aligned}
\mathcal{L}_{I_1}^{(m)}(s, r_{1k}) =& \sum_{(p_a)_{a=1}^m\in\mathcal{M}_m}\frac{\mathcal{L}_{I_1}(s, r_{1k})\, m!}{\prod_{a=1}^m (p_a!(a!)^{m_a})} \\
&\times\prod_{a=1}^m\left(-2\pi\lambda_1\int_{r_{1k}}^{\infty}\left(-\frac{(-1)^a\Gamma(1+a)}{r_{1k}^{a\alpha_1}\left(1+sr_{1k}^{-\alpha_1}\right)^{a+1}}\right)r\text{d}r\right)^{p_a} \quad (36)
\end{aligned}$$

where the integral can be solved using similar method as calculating (35). Similarly, we can calculate $\mathcal{L}_{I_2}(s, r_{2k})$ and its higher order derivative $\mathcal{L}_{I_2}^{(m)}(s, r_{2k})$. Finally, after some algebraic manipulations, we can obtain $\mathcal{S}_{\text{IN},k,R_{1k},R_{2k}}(r_{1k}, r_{2k}, \beta)$ where $k \in \{1, 2\bar{O}, 2OC\}$.

2) $k = 2O\bar{C}$: When $k = 2O\bar{C}$, based on (9), using multinomial theorem, and following similar procedures in calculating (34), we can obtain $\mathcal{S}_{\text{IN},k,R_{1k},R_{2k}}(r_{1k}, r_{2k}, \beta)$. ■

Note that $\mathcal{T}_{k,R_{1k},R_{2k}}(n, r_{1k}, r_{2k}, \beta)$ in (32) can be interpreted as the gain of the SIR coverage probability when the DoF for boosting the desired signal to $u_0 \in \mathcal{U}_k$ at its serving BS is changed from n to $n+1$.

Finally, the expressions in (12)–(15) are obtained by removing the conditions of $\mathcal{S}_{\text{IN},k,R_{1k},R_{2k}}(r_{1k}, r_{2k}, \beta)$ on R_{jk} $(j = 1, 2)$ in (32). Here, $f_{Y_1}(y)$, $f_{Y_2}(y)$, $\mathcal{A}_1$ and $\mathcal{A}_{2\bar{O}}$ are given in [5], and $f_{Y_1,Y_2}(x, y)$ is given in [6].

C. Proof of Lemma 3

1) Proof of $\Delta\bar{\mathcal{R}}_{\text{IN},1}(U,\tau) < 0$: When the design parameter is U, we have

$$\begin{aligned}
\bar{\mathcal{R}}_{\text{IN},1}(U,\tau) =& \int_0^{\infty} f_{Y_1}(y) \\
&\times\left(\sum_{u=0}^{U}\left(\sum_{n=0}^{N_1-u-1}\mathcal{T}_{1,Y_1}(n, y, \hat{\beta})\right)\Pr\left(u_{2OC,0}(U) = u\right)\right)\text{d}y \quad (37)
\end{aligned}$$

where $\hat{\beta} = 2^{\frac{\text{E}\left[L_{0,1}\right]\tau}{W}} - 1$, and $\Pr\left(u_{2OC,0}(U) = u\right) = \begin{cases} \Pr\left(U_{2O_a,0} = u\right), & \text{for } 0 \le u < U \\ \sum_{u=U}^{\infty}\Pr\left(U_{2O_a,0} = u\right), & \text{for } u = U \end{cases}$.

Similarly, when the design parameter is $U-1$, we have

$$\begin{aligned}
\bar{\mathcal{R}}_{\text{IN},1}(U-1,\tau) =& \int_0^{\infty} f_{Y_1}(y) \\
&\times\left(\sum_{u=0}^{U-1}\left(\sum_{n=0}^{N_1-u-1}\mathcal{T}_{1,Y_1}(n, y, \hat{\beta})\right)\Pr\left(u_{2OC,0}(U-1) = u\right)\right)\text{d}y \quad (38)
\end{aligned}$$

where $\Pr\left(u_{2OC,0}(U-1) = u\right) = \begin{cases} \Pr\left(U_{2O_a,0} = u\right), & \text{for } 0 \le u < U-1 \\ \sum_{u=U-1}^{\infty}\Pr\left(U_{2O_a,0} = u\right), & \text{for } u = U-1 \end{cases}$.

Based on (37) and (38), and after some algebraic manipulations, we have

$$\begin{aligned}
\Delta\bar{\mathcal{R}}_{\text{IN},1}(U,\tau) =& -\left(1-\sum_{u=0}^{U-1}\Pr\left(U_{2O_a,0} = u\right)\right) \\
&\times\int_0^{\infty}\mathcal{T}_{1,y}(N_1-U, y, \hat{\beta})f_{Y_1}(y)\text{d}y < 0\,. \quad (39)
\end{aligned}$$

2) Proof of $\Delta\bar{\mathcal{R}}_{\text{IN},2\bar{O}}(\tau) = 0$: Follows by noting that $\bar{\mathcal{R}}_{\text{IN},2\bar{O}}(\tau)$ is independent of U.

3) Proof of $\Delta\bar{\mathcal{R}}_{\text{IN},2O}(U,\tau) > 0$: We first show that $\Pr\left(\mathcal{E}_{2OC,0}(U-1)\right) = \sum_{n=1}^{U-1}\Pr\left(\hat{U}_{2O_a,0} = n\right) + \sum_{n=U}^{\infty}\frac{U-1}{n}\Pr\left(\hat{U}_{2O_a,0} = n\right) < \sum_{n=1}^{U-1}\Pr\left(\hat{U}_{2O_a,0} = n\right) + \Pr\left(\hat{U}_{2O_a,0} = U\right) + \sum_{n=U+1}^{\infty}\frac{U}{n}\Pr\left(\hat{U}_{2O_a,0} = n\right) = \Pr\left(\mathcal{E}_{2OC,0}(U)\right)$, where the inequality is obtained by noting that $\frac{U-1}{n} < \frac{U}{n}$ $(n \in \mathbb{N})$. Next, we show that $\bar{\mathcal{R}}_{\text{IN},2OC}(U,\tau) > \bar{\mathcal{R}}_{\text{IN},2O\bar{C}}(U,\tau)$. From (7) and (9), we note that for any network and channel realizations, since $\frac{\left|\mathbf{h}_{1,10}^{\dagger}\mathbf{f}_{1,1}\right|^2}{Y_1^{\alpha_1}} > 0$, we always have $\text{SIR}_{\text{IN},2OC,0} > \text{SIR}_{\text{IN},2O\bar{C},0}$. Hence, we have $\Pr\left(\text{SIR}_{\text{IN},2OC,0} > \beta\right) > \Pr\left(\text{SIR}_{\text{IN},2O\bar{C},0} > \beta\right)$, i.e., $\bar{\mathcal{R}}_{\text{IN},2OC}(U,\tau) > \bar{\mathcal{R}}_{\text{IN},2O\bar{C}}(U,\tau)$. Finally, since $\Delta\bar{\mathcal{R}}_{\text{IN},2O}(U,\tau) = \left(\Pr\left(\mathcal{E}_{2OC,0}(U)\right) - \Pr\left(\mathcal{E}_{2OC,0}(U-1)\right)\right)\left(\bar{\mathcal{R}}_{\text{IN},2OC}(\tau) - \bar{\mathcal{R}}_{\text{IN},2O\bar{C}}(\tau)\right)$, we obtain $\Delta\bar{\mathcal{R}}_{\text{IN},2O}(U,\tau) > 0$.

D. Proof of Proposition 1

To characterize $\left|\Delta\bar{\mathcal{R}}_{\text{IN},1}(U,\tau)\right|$, by *Corollary 1*, *Theorem 1*, and *Lemma 7*, we first characterize $\mathcal{T}_{k,R_{1k},R_{2k}}\left(n, r_{1k}, r_{2k}, 2^{\text{E}\left[L_{0,j_k}\right]\tau/W} - 1\right)$ in the following lemma.

Lemma 8: When $\tau \to 0$, we have $\mathcal{T}_{k,R_{1k},R_{2k}}\left(n, r_{1k}, r_{2k}, 2^{\text{E}\left[L_{0,j_k}\right]\tau/W} - 1\right) = \Theta\left(\tau^n\right)$.

Proof: Firstly, let $\hat{\beta} = 2^{\frac{\text{E}\left[L_{0,j_k}\right]\tau}{W}} - 1$. It can be easily seen that $\hat{\beta} \to 0$ when $\tau \to 0$. Then, we investigate the asymptotic behavior of $\mathcal{T}_{k,R_{1k},R_{2k}}\left(n, r_{1k}, r_{2k}, \hat{\beta}\right)$ when $\hat{\beta} \to 0$. We note that when $z \to 1$, $B'(a, b, z) = \frac{(1-z)^b}{b} + o\left((1-z)^b\right)$. Then, we have $B'\left(\frac{2}{\alpha}, 1-\frac{2}{\alpha}, \frac{1}{1+c\hat{\beta}}\right) = \frac{\left(c\hat{\beta}\right)^{1-\frac{2}{\alpha}}}{1-\frac{2}{\alpha}} + o\left(\hat{\beta}^{1-\frac{2}{\alpha}}\right)$ and $B'\left(1+\frac{2}{\alpha}, a-\frac{2}{\alpha}, \frac{1}{1+c\hat{\beta}}\right) = \frac{(c\hat{\beta})^{a-\frac{2}{\alpha}}}{a-\frac{2}{\alpha}} + o\left(\hat{\beta}^{a-\frac{2}{\alpha}}\right)$, where $c \in \mathbb{R}^+$. Based on these two asymptotic expressions, and let $s = \tilde{c}\hat{\beta}$ $(\tilde{c} \in \mathbb{R}^+)$ in $\mathcal{L}_{I_j}(s, r_{jk})$ and $\mathcal{L}_{I_j}^{(m)}(s, r_{jk})$, we can obtain: $\mathcal{L}_{I_j}(s, r_{jk}) = 1 - \frac{2\pi\lambda_j\tilde{c}r_{jk}^{2-\alpha_j}}{\alpha_j\left(1-\frac{2}{\alpha_j}\right)}\hat{\beta} + o\left(\hat{\beta}\right)$, and $\mathcal{L}_{I_j}^{(m)}(s, r_{jk}) = \hat{\beta}^m\left(\frac{\tilde{c}}{r_{jk}^{\alpha_j}}\right)^m\sum_{(p_a)_{a=1}^m\in\mathcal{M}_m}\frac{m!}{\prod_a^m(p_a!)}\prod_{a=1}^m\left(\frac{2\pi\lambda_j r_{jk}^2}{\alpha_j\left(a-\frac{2}{\alpha_j}\right)}\right)^{p_a} +$

$o\left(\hat{\beta}^m\right)$. Moreover, when $u_0 \in \mathcal{U}_{2O\bar{C}}$, we have $\left(1+\hat{\beta}\frac{P_1 Y_2^{\alpha_2}}{P_2 Y_1^{\alpha_1}}\right)^{-(q_3+1)} = 1-(q_3+1)\frac{P_1 Y_2^{\alpha_2}}{P_2 Y_1^{\alpha_1}}\hat{\beta} + o(\hat{\beta})$. Substituting the series expansions of $\mathcal{L}_{I_j}(s, r_{jk})$, $\mathcal{L}_{I_j}^{(m)}(s, r_{jk})$, and $\left(1+\hat{\beta}\frac{P_1 Y_2^{\alpha_2}}{P_2 Y_1^{\alpha_1}}\right)^{-(q_3+1)}$ into $\mathcal{T}_{k,R_{1k},R_{2k}}\left(n, r_{1k}, r_{2k}, \hat{\beta}\right)$, and after some algebraic manipulations, we have the final result. ∎

Based on *Lemma 8*, we now show the result in *Proposition 1*. The proof follows by showing the integrand in (39) is upper bounded by an integrable function. In particular, for the integrand in (39), we have $\mathcal{T}_{1,y}(N_1-U, y, \hat{\beta}) f_{Y_1}(y) < \sum_{n_1=0}^{N_1-U}\sum_{(p_a)_{a=1}^{n_1}\in\mathcal{M}_{n_1}}\sum_{(p_b)_{b=1}^{N_1-U-n_1}\in\mathcal{M}_{N_1-U-n_1}} g\left(n_1, \{p_a\}, \{p_b\}, \hat{\beta}\right)\exp\left(-cy^2\right) y^{2\sum_{a=1}^{n_1} p_a + \frac{2\alpha_1}{\alpha_2}\sum_{b=1}^{N_1-U-n_1} p_b + 1}$ where c is a real positive constant and $g\left(n_1, \{p_a\}, \{p_b\}, \hat{\beta}\right)$ is the coefficient (independent of y). Here, the inequality is obtained by noting that $\mathcal{L}_{I_j}(s, r_{jk}) < 1$, $B'(a,b,z) < B(a,b)$ which is the beta function, and $\exp\left(-\pi\lambda_2\left(\frac{P_2 B}{P_1}\right)^{\frac{2}{\alpha_2}} y^{\frac{2\alpha_1}{\alpha_2}} y^{\frac{2\alpha_1}{\alpha_2}}\right) < 1$. It can be easily shown that $y^{2\sum_{a=1}^{n_1} p_a + \frac{2\alpha_1}{\alpha_2}\sum_{b=1}^{N_1-U-n_1} p_b + 1}\exp\left(-cy^2\right)$ is integrable. We know $\mathcal{T}_{k,R_{1k},R_{2k}}\left(N_1-U, r_{1k}, r_{2k}, 2^{\mathrm{E}\left[L_{0,j_k}\right]\tau/W}-1\right) = \Theta\left(\tau^{N_1-U}\right)$ from *Lemma 8*, then using *dominated convergence theorem*, the proof completes.

E. Proof of Proposition 2

When $u_0 \in \mathcal{U}_{2OC}$, since $\left|\mathbf{h}_{2,00}^{\dagger}\mathbf{f}_{2,0}\right|^2 \overset{d}{\sim} \mathrm{Gamma}(N_2, 1)$, we have

$$1-\bar{\mathcal{R}}_{\mathrm{IN},2OC}(\tau) = \tau^{N_2}\frac{\left(\mathrm{E}\left[L_{0,2}\right]\ln(2)\right)^{N_2}}{W^{N_2}N_2!}\sum_{n=0}^{N_2}\binom{N_2}{n}\left(\frac{P_1}{P_2}\right)^n \times \mathrm{E}\left[\left(Y_2^{\alpha_2}I_1\right)^n\right]\mathrm{E}\left[\left(Y_2^{\alpha_2}I_2\right)^{N_2-n}\right] + o\left(\tau^{N_2}\right). \quad (40)$$

In order to show that $1-\bar{\mathcal{R}}_{\mathrm{IN},2OC}(\tau) = \Theta\left(\tau^{N_2}\right)$, we need to show that $\mathrm{E}\left[\left(Y_2^{\alpha_2}I_1\right)^n\right] < \infty$ and $\mathrm{E}\left[\left(Y_2^{\alpha_2}I_2\right)^{N_2-n}\right] < \infty$. This can be proved by noting that $\mathrm{E}\left[\left(Y_2^{\alpha_2}I_2\right)^{N_2-n}\right] < \infty$ [26], and $\mathrm{E}\left[\left(Y_2^{\alpha_2}I_1\right)^n\right] \overset{(a)}{<} \left(\frac{BP_2}{P_1}\right)^n \mathrm{E}\left[\left(Y_1^{\alpha_1}I_1\right)^n\right] < \infty$ [26] where (a) is obtained by following $Y_2^{\alpha_2} < \frac{BP_2}{P_1}Y_1^{\alpha_1}$ when $u_0 \in \mathcal{U}_{2OC}$. Similarly, when $u_0 \in \mathcal{U}_{2O\bar{C}}$, we have

$$\begin{aligned}&1-\bar{\mathcal{R}}_{\mathrm{IN},2O\bar{C}}(\tau)\\ =&\tau^{N_2}\frac{\left(\mathrm{E}\left[L_{0,2}\right]\ln(2)\right)^{N_2}}{W^{N_2}N_2!}\mathrm{E}\left[Y_2^{\alpha_2}\left(\frac{P_1}{P_2}I_1 + I_2 + \frac{1}{Y_1^{\alpha_1}}\frac{P_1}{P_2}g_{1,1}\right)^{N_2}\right]\\ &+o\left(\tau^{N_2}\right)\\ =&\Theta\left(\tau^{N_2}\right). \end{aligned} \quad (41)$$

Finally, by noting that $\frac{1}{N_2!}\mathrm{E}\left[Y_2^{\alpha_2}\left(\frac{P_1}{P_2}I_1 + I_2 + \frac{1}{Y_1^{\alpha_1}}\frac{P_1}{P_2}g_{1,1}\right)^{N_2}\right] > \frac{1}{N_2!}\mathrm{E}\left[Y_2^{\alpha_2}\left(\frac{P_1}{P_2}I_1 + I_2\right)^{N_2}\right]$, we have $\bar{\mathcal{R}}_{\mathrm{IN},2OC}(\tau) - \bar{\mathcal{R}}_{\mathrm{IN},2O\bar{C}}(\tau) = \Theta\left(\tau^{N_2}\right)$. Moreover, since $\Pr\left(\mathcal{E}_{2OC,0}(U)\right) - \Pr\left(\mathcal{E}_{2OC,0}(U-1)\right)$ is independent of τ, we obtain the final result.

F. Proof of Theorem 3

According to (26), *Proposition 1* and *Proposition 2*, and noting that $\mathcal{A}_{2O}$ and $\mathcal{A}_1$ are independent of τ, we have

$$\begin{aligned}\Delta\bar{\mathcal{R}}_{\mathrm{IN}}(U,\tau) &= \Theta\left(\tau^{N_2}\right) - \Theta\left(\tau^{N_1-U}\right)\\ &= \begin{cases}\Theta\left(\tau^{N_2}\right) > 0\,, & \text{when } U < N_1-N_2\\ \Theta\left(\tau^{N_2}\right) - \Theta\left(\tau^{N_2}\right)\,, & \text{when } U = N_1-N_2\\ \Theta\left(\tau^{N_2-U}\right) < 0\,, & \text{when } U > N_1-N_2\end{cases}. \end{aligned} \quad (42)$$

Since $U^*(\tau)$ satisfies $\Delta\bar{\mathcal{R}}_{\mathrm{IN}}(U^*(\tau),\tau) > 0$ and $\Delta\bar{\mathcal{R}}_{\mathrm{IN}}(U^*(\tau)+1,\tau) \le 0$, we see from (42) that $U^*(\tau)$ should be in the set $\{N_1-N_2-1, N_1-N_2\}$, and the exact value of $U^*(\tau)$ depends on whether $\Delta\bar{\mathcal{R}}_{\mathrm{IN}}(U,\tau)$ is positive or not when $U = N_1-N_2$ (i.e., the second case in (42)), i.e., whether the coefficient in $\Theta\left(\tau^{N_2}\right)$ corresponding to $\mathcal{A}_{2O}\Delta\bar{\mathcal{R}}_{\mathrm{IN},2O}(U,\tau)$ (i.e., the first one) is larger than that in $\Theta\left(\tau^{N_2}\right)$ corresponding to $\mathcal{A}_1\left|\Delta\bar{\mathcal{R}}_{\mathrm{IN},1}(U,\tau)\right|$ (i.e., the second one) or not.

G. Proof of Lemma 5

1) Proof of i): First, assuming that $N_1 - U$ DoF are used for IN, we obtain a lower bound of $\bar{\mathcal{R}}_{\mathrm{IN},1}(U,\tau)$, denoted as $\bar{\mathcal{R}}_{\mathrm{IN},1}^{\mathrm{lb}}(U,\tau)$. Following similar procedures in [29, Appendix B], we have the following result for $\bar{\mathcal{R}}_{\mathrm{IN},1}^{\mathrm{lb}}(U,\tau)$ when $N_1, U \to \infty$ with $\frac{U}{N_1} \to \kappa \in (0,1)$ and $\tau \to 0$:

$$\begin{aligned}\bar{\mathcal{R}}_{\mathrm{IN},1}^{\mathrm{lb}}(U,\tau) &\approx \Pr\left(\frac{\frac{P_1}{Y_1^{\alpha_1}}\frac{N_1(1-\kappa)}{2^{\frac{\tau}{W}\mathrm{E}\left[L_{0,1}\right]}-1}}{P_1I_1+P_2I_2} > 1\right)\\ &\overset{(a)}{\approx} \Pr\left(\frac{P_1}{Y_1^{\alpha_1}\left(P_1I_1+P_2I_2\right)} > \frac{\ln(2)\tau\mathrm{E}\left[L_{0,1}\right]}{WN_1(1-\kappa)}\right) \end{aligned} \quad (43)$$

where (a) is obtained by noting that $2^{\frac{\tau}{W}\mathrm{E}[L_{0,1}]} \approx 1 + \ln(2)\mathrm{E}\left[L_{0,1}\right]\frac{\tau}{W}$ as $\tau \to 0$.

Similarly, for ABS, when $\tau \to 0$, we have the following:

$$\bar{\mathcal{R}}_{\mathrm{ABS},1}(\tau) \approx \Pr\left(\frac{P_1}{Y_1^{\alpha_1}\left(P_1I_1+P_2I_2\right)} > \frac{\ln(2)\tau\mathrm{E}\left[L_{0,1}\right]}{WN_1(1-\eta)}\right). \quad (44)$$

From (43) and (44), we see that $\bar{\mathcal{R}}_{\mathrm{IN},1}^{\mathrm{lb}}(U,\tau) > \bar{\mathcal{R}}_{\mathrm{ABS},1}(\tau)$, which is a sufficient condition of $\bar{\mathcal{R}}_{\mathrm{IN},1}(U,\tau) > \bar{\mathcal{R}}_{\mathrm{ABS},1}(\eta,\tau)$, i.f.f. $\frac{1}{1-\kappa} < \frac{1}{1-\eta}$. After some manipulations, we have the final result.

2) Proof of ii): The proof is similar to that of i), and is omitted due to page limit.

3) Proof of iii): Omitted due to page limit.

REFERENCES

[1] Cisco, "Cisco visual networking index: Global mobile data traffic forecast update, 2013-2018," *White paper*, Feb. 2014.

[2] H. S. Jo, Y. J. Sang, P. Xia, and J. G. Andrews, "Heterogeneous cellular networks with flexible cell association: A comprehensive downlink SINR analysis," *IEEE Trans. Wireless Commun.*, vol. 11, no. 10, pp. 3484–3495, Oct. 2012.

[3] J. G. Andrews, S. Singh, Q. Ye, X. Lin, and H. S. Dhillon, "An overview of load balancing in HetNets: Old myths and open problems," *IEEE Wireless Commun.*, vol. 21, no. 2, pp. 18–25, Apr. 2014.

[4] S. Singh, H. S. Dhillon, and J. G. Andrews, "Offloading in heterogeneous networks: modeling, analysis, and design insights," *IEEE Trans. Wireless Commun.*, vol. 12, no. 5, pp. 2484–2497, Mar. 2013.

[5] S. Singh and J. G. Andrews, "Joint resource partitioning and offloading in heterogeneous cellular networks," *IEEE Trans. Wireless Commun.*, vol. 13, no. 2, pp. 888–901, Feb. 2014.

[6] A. H. Sakr and E. Hossain, "Location-aware cross-tier coordinated multipoint transmission in two-tier cellular networks," *to appear in IEEE Trans. Wireless. Commun.*, 2014. [Online]. Available: http://arxiv.org/abs/1405.2876

[7] Y. Cui, V. K. N. Lau, and Y. Wu, "Delay-aware BS discontinuous transmission control and user scheduling for energy harvesting downlink coordinated MIMO systems," *IEEE Trans. Signal Processing*, vol. 60, no. 7, pp. 3786–3795, Jul. 2012.

[8] K. Hosseini, J. Hoydis, S. t. Brink, and M. Debbah, "Massive MIMO and small cells: How to densify heterogeneous networks," in *Proc. of IEEE Int. Conf. on Commun. (ICC)*, Budapest, Jun. 2013, pp. 5442–5447.

[9] M. Kountouris and N. Pappas, "HetNets and massive MIMO: Modeling, potential gains, and performance analysis," in *Proc. of IEEE-APS Topical Conference on APWC*, Torino, Italy, Sep. 2013, pp. 1319–1322.

[10] A. Adhikary, E. A. Safadi, and G. Caire, "Massive MIMO and inter-tier interference coordination," in *Proc. of Information Theory and Applications Workshop (ITA)*, San Diego, CA, Feb. 2014, pp. 1–10.

[11] A. Adhikary, H. S. Dhillon, and G. Caire, "Massive-MIMO meets HetNet: Interference coordination through spatial blanking," *submitted to IEEE J. Select. Areas Commun.*, Jul. 2014. [Online]. Available: http://arxiv.org/abs/1407.5716

[12] P. Xia, C. H. Liu, and J. G. Andrews, "Downlink coordinated multi-point with overhead modeling in heterogeneous cellular networks," *IEEE Trans. Wireless Commun.*, vol. 12, no. 8, pp. 4025–4037, Aug. 2013.

[13] X. Zhang and M. Haenggi, "A stochastic geometry analysis of inter-cell interference coordination and intra-cell diversity," *to appear in IEEE Trans. on Wireless. Commun.*, 2014. [Online]. Available: http://arxiv.org/abs/1403.0012

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

[15] 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, Apr. 2015.

[16] W. Nie, F. C. Zheng, X. Wang, S. Jin, and W. Zhang, "Energy efficiency of cross-tier base station cooperation in heterogeneous cellular networks," *submitted to IEEE Trans. on Wireless Commun.* [Online]. Available: http://arxiv.org/abs/1406.1867

[17] Y. Lin and W. Yu, "Joint spectrum partition and user association in multi-tier heterogeneous networks," in *Proc. of Conference on Information Science and Systems (CISS)*, Princeton, NJ, Mar. 2014, pp. 1–6.

[18] S. Sesia, I. Toufik, and M. Baker, *LTE–the UMTS Long Term Evolution: from Theory to Practice*, 1st ed. United Kingdom: John Wiley and Sons, 2009.

[19] M. Haenggi and R. K. Ganti, "Interference in large wireless networks," *Foundations and Trends in Networking*, vol. 3, no. 2, pp. 127–248, 2009.

[20] R. W. Heath Jr., M. Kountouris, and T. Bai, "Modeling heterogeneous network interference using Poisson point processes," *IEEE Trans. Signal Processing*, vol. 61, no. 16, pp. 4114–4126, Aug. 2013.

[21] Y. Cui, Q. Huang, and V. K. N. Lau, "Queue-aware dynamic clustering and power allocation for network MIMO systems via distributed stochastic learning," *IEEE Trans. Signal Processing*, vol. 59, no. 3, pp. 1229–1238, Mar. 2011.

[22] B. Clerckx and C. Oestges, *MIMO Wireless Networks: Channels, Techniques and Standards for Multi-Antenna, Multi-User and Multi-Cell Systems*. Oxford, UK: Academic Press (Elsevier), 2013.

[23] 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, Nov. 2011.

[24] T. Bai and R. W. Heath Jr., "Asymptotic coverage probability and rate in massive MIMO networks," 2013. [Online]. Available: http://arxiv.org/abs/1305.2233

[25] S. M. Yu and S. L. Kim, "Downlink capacity and base station density in cellular networks," in *Workshop in Spatial Stochastic Models for Wireless Networks*, Tsukuba Science City, Japan, May 2013, pp. 1–7.

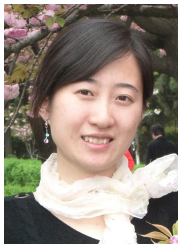
[26] M. Haenggi, "The mean interference-to-signal ratio and its key role in cellular and amorphous networks," *to appear in IEEE Wireless Commun. Lett.*, 2014. [Online]. Available: http://arxiv.org/abs/1406.2794

[27] A. K. Gupta, H. S. Dhillon, S. Vishwanath, and J. G. Andrews, "Downlink multi-antenna heterogeneous cellular network with load balancing," *IEEE Trans. Comm.*, vol. 62, no. 11, pp. 4052–4067, Nov. 2014.

[28] W. P. Johnson, "The curious history of Faa di Bruno's formula," *The American Mathematical Monthly*, vol. 109, no. 3, pp. 217–234, Mar. 2002.

[29] Y. Wu, R. H. Y. Louie, M. R. McKay, and I. B. Collings, "Generalized framework for the analysis of linear MIMO transmission schemes in decentralized wireless ad hoc networks," *IEEE Trans. Wireless Commun.*, vol. 11, no. 8, pp. 2815–2827, Aug. 2012.

Yueping Wu (S'09-M'13) received the B.Eng degree in Information and Communication Engineering from Xi'an Jiaotong University, China, in 2008 and the Ph.D degree in Electronic and Computer Engineering from the Hong Kong University of Science and Technology (HKUST) in 2013. From Nov. 2013 to Feb. 2015, she was a Postdoctoral Research Associate at Imperial College London, United Kingdom. Since Mar. 2015, she is a Postdoctoral Research Fellow at the University of Hong Kong. From June 2012 to December 2012, she was a visiting scholar at the Wireless Networking and Communications Group (WNCG), the University of Texas at Austin, TX, USA. Her current research interests include mmWave communications, heterogeneous networks, MIMO systems, and the applications of stochastic geometry and random matrix theory. Dr. Wu was awarded a Best Paper Award at IEEE Globecom 2010.



Ying Cui (S'08-M'12) received her B.E. degree in Electronic and Information Engineering from Xi'an Jiao Tong University, China, in 2007 and her Ph.D. degree in Electronic and Computer Engineering from the Hong Kong University of Science and Technology (HKUST), Hong Kong, in 2011. From Janurary 2011 to July 2011, she was a Visiting Assistant in Research in the Department of Electrical Engineering at Yale University, US. From March 2012 to June 2012, she was a Visiting Scholar in the Department of Electronic Engineering at Macquarie University, Australia. From June 2012 to June 2013, she was a Postdoctoral Research Associate in the Department of Electrical and Computer Engineering at Northeastern University, US. From July 2013 to December 2014, she was a Postdoctoral Research Associate in the Department of Electrical Engineering and Computer Science at Massachusetts Institute of Technology (MIT), US. Since Janurary 2015, she has been an Associate Professor in the Department of Electronic Engineering at Shanghai Jiao Tong University, China. Her current research interests include delay-sensitive cross-layer control and optimization, future Internet architecture and network coding. She was selected into China's 1000Plan Program for Young Talents in 2013.

Bruno Clerckx (M'08) is a Senior Lecturer (Associate Professor) in the Electrical and Electronic Engineering Department at Imperial College London (London, United Kingdom). He received his M.S. and Ph.D. degree in applied science from the Université catholique de Louvain (Louvain-la-Neuve, Belgium) in 2000 and 2005, respectively. He held visiting research positions at Stanford University (CA, USA) in 2003 and Eurecom Institute (Sophia-Antipolis, France) in 2004. In 2006, he was a Post-Doc at the Universit catholique de Louvain. From 2006 to 2011, he was with Samsung Electronics (Suwon, South Korea) where he actively contributed to 3GPP LTE/LTE-A and IEEE 802.16m and acted as the rapporteur for the 3GPP Coordinated Multi-Point (CoMP) Study Item. He is the author or coauthor of two books on MIMO wireless communications and networks and numerous research papers, standard contributions and patents. He received the Best Student Paper Award at the IEEE SCVT 2002 and several Awards from Samsung in recognition of special achievements. Dr. Clerckx has served as an editor for IEEE TRANSACTIONS ON COMMUNICATIONS.