Performance Analysis of SDMA with Inter-tier Interference Nulling in HetNets

Yamuna Dhungana, *Student Member, IEEE*, and Chintha Tellambura, *Fellow, IEEE*

***Abstract*— The downlink performance of two-tier (macro/pico) multi-antenna cellular heterogeneous networks employing space division multiple access (SDMA) technique with zero-forcing precoding is analyzed in this paper. The number of users simultaneously served with SDMA by a base-station (BS) depends on the number of active users in its cell, with the maximum served users limited to $L_{\max}$. To protect the pico users from severe macro-interference, part of the antennas at each macro BS is proposed to be utilized toward interference nulling to pico users. The partitioning of macro antenna resources to serve macro-users and to null interference to pico users for optimal performance is investigated in this paper. Biased-nearest-distance-based user association scheme is proposed, where the bias value accounts for the natural bias due to the differences in multi-antenna transmission schemes across tiers, as well as the artificial bias for load balancing. The signal-to-interference-ratio coverage probability, rate distribution, and average rate of a typical user are then derived. Our results demonstrate that the proposed interference nulling scheme has strong potential for improving performance if the macro antennas partitioning is carefully done. The optimal $L^*_{\max}$ for both macro and pico-tier, which maximize the average data rate, is also investigated and it is found to outperform both single-user beamforming and full-SDMA. Finally, the impact of imperfect channel state information due to limited feedback is analyzed.**

***Index Terms*— Heterogeneous networks (HetNets), interference nulling, limited feedback, Poisson point process (PPP), space division multiple access (SDMA), stochastic geometry.**

I. Introduction

NETWORK densification (dense deployment of base-stations (BSs)) and multi-antenna techniques are well-known for their tremendous potential to increase spectral efficiency of wireless networks. In a conventional macro only cellular network, where the locations of high-power macro BSs are strictly planned, adding more BSs can be very challenging for dense urban areas due to extremely high site acquisition cost. Thus, the cost-effective way of network densification is to deploy a diverse set of low-power BSs within the areas covered by macro cellular infrastructure [1]. The resulting network of mixed types of BSs is known as heterogeneous network (HetNet). If the BSs are equipped with multiple antennas, the additional degrees of freedom (DoF) in the spatial dimension can be utilized in a number of ways, for example, to improve the spectral efficiency, and to enhance the link reliability. The diversity and spatial multiplexing gains have been extensively studied in general for point-to-point links without interference. Some examples of diversity techniques are space-time coding [2], [3] and coherent processing known as beamforming [4]. The spatial multiplexing which utilizes the multiple antennas to transmit independent data streams simultaneously over spatial sub-channels, has been explored in [5]. Space division multiple access (SDMA) which allows multiple users to be served simultaneously on the same time-frequency resource has also been analyzed [6], [7]. However, in interference-prone cellular networks, for example, a dense deployed HetNet, where complex interference scenarios may arise due to power disparities between the BSs, the effectiveness of spatial multiplexing may diminish [8]. Nevertheless, if the available spatial DoF are intelligently utilized to suppress/mitigate interference as well as to harvest diversity and multiplexing gain, the performance of cellular networks can be improved. In this paper, we develop a tractable framework to analyze the downlink performance of zero-forcing (ZF) precoding based joint SDMA and inter-tier interference-nulling scheme in HetNets.

A. Related Work and Contributions of the Paper

Although multiple antenna in wireless communications is a mature technology, its incorporation into cellular networks, traditional single tier, as well as HetNets, has received much momentum both in academic research and standardization efforts only recently with the introduction of massive-MIMO concept [9]–[12]. By utilizing the stochastic geometry framework which enables systematic modeling of interference, several studies on the modeling and analysis of downlink single-tier multi-antenna cellular networks have been reported in the literature. For example, error probability analysis by using the equivalent-in-distribution approach in [13], coverage and rate analyses using the Gil-Pelaez inversion theorem in [14], and a unified approach to error probability, outage and rate analyses for different multi-antenna configurations with retransmissions in [15]. Apart from single-tier networks, stochastic geometric modeling of downlink multi-antenna HetNets have been significantly explored as well. Reference [16] compared the signal-to-interference-and-noise ratio (SINR) coverage of SU-BF with that of ZF SDMA for a two-tier multi-antenna HetNet by considering a single fixed-radius circular

Manuscript received June 12, 2016; revised November 12, 2016; accepted January 4, 2017. This work was supported by the Alberta Innovates - Technology Futures (AITF) Graduate Student Scholarship. This paper was presented in part at IEEE ICC, Kuala Lampur, Malaysia, May 2016. The associate editor coordinating the review of this paper and approving it for publication was P. Salvo Rossi.

Y. Dhungana is with Ericsson Canada Inc., Ottawa, ON K2K 2V6, Canada (e-mail: dhungana@ualberta.ca).

C. Tellambura is with the Department of Electrical and Computer Engineering, University of Alberta, Edmonton, AB T6G 1H9, Canada (e-mail: chintha@ece.ualberta.ca).

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TWC.2017.2656083

macro cell with multiple femto cells of fixed radii, distributed according to a Poisson point process (PPP) within the macro cell. However, since BS-user association and macro-tier interference are ignored, the insights in [16] may not be accurate for practical HetNets. The coverage probability and average link spectral efficiency of ZF precoding in multi-antenna HetNet, spatially averaged over a given cell of known radius and guard region are derived in [17]. Unlike the spatial averaging over a given cell in [17], system-wide spatial averaging is considered in [18] and the upper bounds on coverage probability of ZF SDMA and SU-BF are derived. The ordering results for the coverage probability and rate per user performance of SDMA, SU-BF and single-antenna transmission are also derived in [18] by using tools from stochastic orders. While the analysis in [18] is based on maximum instantaneous SINR based BS-user association, association rules intended to maximize the average receive SINR (and thus, the SINR coverage), and biased association for optimal rate coverage are proposed for multi-antenna HetNets in [19]. Closed form expressions for the signal-to-interference ratio (SIR) of ZF SDMA and SU-BF are derived in [20] for user association based on the received power of the reference signal transmitted from a single-antenna with total power. In all of these downlink multi-antenna HetNet analyses [16]–[20], each cell of a tier is assumed to be spatially multiplexing to the same number of users, say L, and it can be any arbitrary integer in the interval $[1, K_i]$, where K_i is the number of antennas at a BS of the ith tier. This assumption, however, is not suitable for cellular networks because the number of users, which depends on user distribution, is generally different from one cell to another. An open-loop SDMA with each antenna serving an independent data stream to its user with the limiting requirement that the number of users in each cell must be at least equal to the number of transmit antennas is analyzed in [21] for single-tier cellular networks with ZF and MMSE receivers. In this paper, we consider user-distribution dependent SDMA scheme, i.e., the number of users simultaneously served with SDMA in each cell depends on the total number of users in that cell. If the number of users in a cell is less than the maximum number of users served per resource block (RB), say $L_{\max}$, all the users are simultaneously served; otherwise only $L_{\max}$ users chosen randomly are served.

One of the key challenges in downlink cellular HetNets is inter-tier interference management. Due to large transmit power disparities between macro and small-cell nodes such as picos and femtos, and proactive user offloading from macro to small cells, interference management between the macro and pico/femto tiers is very important because the performance of small-cell cell-edge users could be severely degraded. While almost blank subframes (ABSF) [22], [23] and frequency-domain resource partitioning [24], [25] can be used, inter-tier interference can be more efficiently managed without compromising time/frequency resources by using multiple antennas. Inter-tier interference mitigation by using multiple receive antennas at the user devices is analyzed in [26]. In this paper, we analyze ZF-precoding based *interference-nulling* method by using BS antennas to suppress the interference from the macro tier to small-cell users. Compared to other potential techniques such as joint transmission [27] and transmission point selection [28], which require both user data and channel state information (CSI) to be shared between the coordinating BSs, interference nulling requires only CSI to be shared. Joint transmission with local precoding, which requires no CSI exchange between the coordinating BSs, is studied in [12]. However, it stills requires user data sharing, which could be very challenging due to backhaul overhead. In [29], interference nulling to U offloadedpico users by each macro BS is analyzed, where the optimal U for maximum rate coverage is also investigated. However, unlike [29] which considers a single served user per RB in each cell, we consider a user-distribution dependent SDMA scheme. SU-BF with interference nulling to a fixed number of neighboring-cells users at each BS of any tier for general multi-tier HetNets is analyzed in [30], without specifying how these users are selected. SU-BF with interference nulling in single-tier cellular networks is studied in [31] and [32]. Although SU-BF with interference nulling has been relatively well analyzed, to the best of our knowledge, this paper is the first work to analyze a user-distribution dependent SDMA scheme with inter-tier interference nulling in cellular HetNets. The main contributions of this paper are summarized as follows.

1) We develop a tractable framework to analyze a user-distribution dependent SDMA scheme in a two-tier (macro/pico) multi-antenna HetNet with ZF precoding, in which the number of users simultaneously served by a BS in an RB depends on the number of active users in its cell. The framework also allows the analysis of SU-BF and full-SDMA by setting the limit on the number of users served per RB to one, and the total number of transmit antennas, respectively.
2) To suppress the detrimental macro-to-pico interference, interference-nulling precoding, jointly with user-distribution dependent SDMA, is proposed. That is, the precoding matrix at each macro BS is designed to null interference to a set of activepico users while spatially multiplexing the macro-users in the cell. In the proposed interference-nulling scheme, the candidatepico users for interference nulling from a macro BS, say b, are the ones which have b as their nearest interfering macro BS.
3) Considering the complexity of BS-user association in multi-antenna HetNets, a simple biased-nearest-distance based association rule is introduced, in which the bias value accounts for the natural bias required for SINR maximization in multi-antenna HetNets, as well as the artificial bias for load balancing.
4) By considering interference limited scenario, we derive analytical expressions for the SIR and rate distributions, as well as the average rate of a typical user. We then perform comprehensive analysis to investigate the optimal association bias, and the inherent trade-off between interference cancellation, signal power boosting and spatial multiplexing. The following useful network design insights are obtained from these analyses:
   a) By optimizing the maximum number of users simultaneously served per RB, SDMA can achieve

significantly higher average data rate than both SU-BF and full SDMA.

b) If the number of users in a typical cell is less than the maximum number of users served per RB, say L_{max}, the optimal number of antennas towards spatial multiplexing and signal power boosting of local users is found to be L_{max}.Thus, rather than allocating additional antennas to these users, the average data rate can be significantly increased if the surplus antennas are used towards interference nulling topico users.

c) The optimal number of antennas towards interference nulling topico users increases with the increase in pico cell density, as well as association bias.

5) Finally, the impact of the CSI quantization error due to limited feedback on interference nulling is also investigated.

The paper is organized as follows. The system model and the proposed multi-antenna technique are presented in Section II. Section III derives the SIR distribution. The rate coverage and the average rate are derived in Section IV. In Section V, the impact of limited feedback is analyzed. The numerical results are presented in Section VI, and the concluding remarks in Section VII.

II. System Model

We consider the downlink of a two-tier multi-antenna HetNet comprising macro and pico BSs spatially distributed on $\mathbb{R}^2$ plane as independent homogeneous PPPs Φ_m with density λ_m and Φ_p with density λ_p, respectively. The macro BSs are equipped with K_m transmit antennas, and the pico BSs with K_p antennas. Similarly, users are assumed to be distributed according to an independent PPP Φ_u with density λ_u, and each has a single receive antenna. The two network tiers share the same spectrum with the universal frequency reuse.

The transmission scheme is SDMA with ZF precoding applied at each BS to serve multiple users simultaneously in each RB. We assume only one RB per time slot. As the BSs and users are independently distributed on the $\mathbb{R}^2$ plane, the number of users varies across cells. Thus, in our proposed SDMA scheme, a typical active macro cell with $N_m \geq 1$ users serves $M_m = \min(N_m, L^M_{\max})$ users simultaneously in a given time slot, where $L^M_{\max}$ is the maximum number of users it can serve. If $N_m > L^M_{\max}$, the BS choses $L^M_{\max}$ users for service randomly, else, all N_m users are served. Similarly, $M_p = \min(N_p, L^P_{\max})$ users are simultaneously served by a typical active pico cell in a given time slot, which has $N_p \geq 1$ users, and $L^P_{\max}$ is the maximum number the pico cell can serve. The macro and pico BSs transmit to each of their users with power P_m and P_p, respectively.

A. User Association

According to the user association rule introduced in [19] for average SINR maximization, a typical user at the origin is associated with the nearest pico BS if $P_p\sqrt{\Delta_p \tau_p} X_p^{-\alpha} \geq P_m \sqrt{\Delta_m \tau_m} X_m^{-\alpha}$, and otherwise, is associated with the nearest macro BS, where $X_m = \min_{x_m \in \Phi_m} \|x_m\|$ and $X_p = \min_{x_m \in \Phi_p} \|x_p\|$ are the distances from the origin to the nearest macro and pico BSs, respectively. If associated with the macro tier, Δ_m is the average desired channel gain from the nearest macro BS, and τ_p is the average interference channel gain from the nearest pico BS. Similarly, Δ_p and τ_m are the corresponding values, if associated with the pico tier. These channel gains depend on the number of users served with SDMA. This association rule is thus not suitable for our proposed SDMA scheme, where the number of users served with SDMA in each cell is a function of the number of users in that cell. The number of users, on the other hand, is determined by the association rule. The above rule however can be equivalently expressed as follows: a user is associated with the pico tier only if

$$X_m \geq \left(\frac{P_m}{P_p}\right)^{\frac{1}{\alpha}} \left(\frac{1}{\varrho}\right)^{\frac{1}{\alpha}} X_p, \tag{1}$$

where $\varrho = \sqrt{\frac{\Delta_p \tau_p}{\Delta_m \tau_m}}$. If we compare (1) with the popular received power based association in HetNets [24], [33], ϱ can be interpreted as the natural bias required for average SINR maximization in multi-antenna HetNets due to the differences in transmission schemes. This coverage maximization bias, however, may not always achieve optimum load balancing for maximum rate. Thus, by further introducing an artificial bias B for load balancing, the resultant condition for pico tier association becomes $X_m \geq \rho X_p$, which can be perceived as biased nearest distance association with bias value $\rho = (\frac{P_m}{P_p}\frac{1}{\eta})^{\frac{1}{\alpha}}$, where $\eta = B\varrho$. We investigate the optimal value of η for the average data rate in Section VI, which determines the optimal ρ.

As X_m and X_p follow Rayleigh distributions with mean $(2\sqrt{\lambda_m})^{-1}$ and $(2\sqrt{\lambda_p})^{-1}$, respectively [34], the probability that a typical user at the origin is associated with the pico tier is

$$A_p = \mathbb{P}(X_m \geq \rho X_p) = \frac{\lambda_p}{\lambda_p + \lambda_m \rho^2}, \tag{2}$$

and the probability that this user is associated with the macro tier is $A_m = 1 - A_p$. These tier association probabilities are also valid for any randomly selected user. Thus, the total set of users in the network, Φ_u can be divided into two disjoint subsets: Φ_u^m and Φ_u^p, the set of macro- and pico-users, respectively. A_m and A_p can be interpreted as the average fraction of users belonging to Φ_u^m and Φ_u^p, respectively. As we are interested in the number of users in a typical cell, rather than the actual locations of the users, Φ_u^m and Φ_u^p can be equivalently modeled as independent PPPs with density $A_m\lambda_u$ and $A_p\lambda_u$, respectively. Since each macro-user is always associated with the nearest macro BS and each pico-userr with the nearest pico BS, the network can be viewed as a superposition of two independent Voronoi tessellations of the macro and pico tiers. Let the number of users in a randomly chosen macro and pico cell be denoted by U_m and U_p, respectively. Their approximate[1] probability mass

[1] The PDF of the normalized Poisson-Voronoi cell area is approximated as Gamma(3.5, 3.5) [35] while deriving the PMFs.

function (PMFs) are given by [24, Lemma 2]

$$\mathbb{P}(U_l = n) = \frac{3.5^{3.5}\Gamma(3.5+n)(A_l\lambda_u/\lambda_l)^n}{\Gamma(3.5)n!(A_l\lambda_u/\lambda_l+3.5)^{n+3.5}}, n \geq 0, \quad \forall l \in \{m, p\}. \tag{3}$$

A BS without any user associated does not transmit at all and is inactive. The PMFs of the number of users in a randomly chosen active cell of the macro and pico tiers are given by

$$\mathbb{P}(N_l = n) = \frac{\mathbb{P}(U_l = n)\mathbf{1}(n \geq 1)}{p_l}, \ \forall l \in \{m, p\}, \tag{4}$$

where p_m and p_p are the probabilities that a typical BS of the macro and pico tiers, respectively, is active, and are given by

$$p_l = 1 - \mathbb{P}(U_l = 0) = 1 - \left(1 + 3.5^{-1}\frac{A_l\lambda_u}{\lambda_l}\right)^{-3.5}, \quad \forall l \in \{m, p\}. \tag{5}$$

Let the sets of active macro and active pico BSs be denoted by Ψ_m and Ψ_p, respectively. Ψ_m and Ψ_p are thinned versions of the original PPPs Φ_m and Φ_p, respectively, and hence are independent PPPs with densities $p_m\lambda_m$ and $p_p\lambda_p$, respectively.

By using the PMFs in (4), the PMFs of the number of users simultaneously served by a typical active BS of macro and pico tiers in a given time slot for $L^l_{\max} > 1$ can be obtained as

$$\mathbb{P}(M_l = n) = \begin{cases} \mathbb{P}(N_l = n), & 1 \leq n < L^l_{\max} \\ 1 - \sum_{k=1}^{L^l_{\max}-1} \mathbb{P}(N_l = k), & n = L^l_{\max}, \end{cases} \quad \forall l \in \{m, p\}. \tag{6}$$

For the special case of $L^l_{\max} = 1, \mathbb{P}(M_l = 1) = 1, \forall l \in \{m, p\}$.

B. Interference Nulling

We assume K_m to be typically much larger than K_p. By using the interference nulling strategy, the additional spatial DoF of macro BSs can be utilized to suppress the strong macro interference topico users. Thus, we propose that each served pico-user requests its nearest active macro BS to perform interference nulling. However, as nulling costs macro BSs their available DoF for their own users, we assume that each macro BS can handle at most $K_m - T_{\min}$ requests only. This limit ensures that each macro BS has at least $T_{\min} \geq L^M_{\max}$ antennas dedicated for serving its own users. Hence, if Q_m requests are received by a typical active macro BS, it will perform interference nulling to $o = \min(Q_m, K_m - T_{\min})$pico users. For $Q_m > (K_m - T_{\min})$, the BS will randomly choose $K_m - T_{\min}$pico users.

The number of interference-nulling requests Q_m received by a typical active macro BS is equal to the number of servedpico users within a typical Voronoi cell Υ of the tessellation formed by Ψ_m. Although the number ofpico users served by a typical active pico BS cannot exceed $L^p_{\max}$, Q_m is unbounded because the number of active pico BSs within Υ is Poisson distributed with mean $p_p\lambda_p/(p_m\lambda_m)$. To derive the PMF of Q_m, we first derive $\mathbb{E}[M_p] = A_p\vartheta_p\lambda_u/(p_p\lambda_p)$, where

$$\vartheta_p = \frac{L^p_{\max}p_p\lambda_p}{A_p\lambda_u} - \frac{3.5^{3.5}}{\Gamma(3.5)}\sum_{k=1}^{L^p_{\max}-1} \times \left[\frac{\Gamma(3.5+n)}{n!}\frac{(A_p\lambda_u/\lambda_p)^{n-1}(L^p_{\max}-k)}{(A_p\lambda_u/\lambda_p+3.5)^{n+3.5}}\right]. \tag{7}$$

Note that for $L^p_{\max} = 1$, $\vartheta_p = \frac{p_p\lambda_p}{A_p\lambda_u}$. Next, let us denote the set ofpico users requesting interference nulling by Ψ^p_u. Because we are only interested in the number of such users in a typical Voronoi cell Υ, and not their actual locations, and we know that $\mathbb{E}[Q_m] = A_p\vartheta_p\lambda_u/(p_m\lambda_m)$, Ψ^p_u can be assumed to be a PPP with density $A_p\vartheta_p\lambda_u$. The PMF of Q_m can then be obtained as

$$\mathbb{P}(Q_m = n) = \frac{3.5^{3.5}\Gamma(3.5+n)\left(\frac{A_p\vartheta_p\lambda_u}{p_m\lambda_m}\right)^n}{\Gamma(3.5)n!\left(\frac{A_p\vartheta_p\lambda_u}{p_m\lambda_m}+3.5\right)^{n+3.5}}, n \geq 0. \tag{8}$$

Due to the limited resources as discussed earlier, not all interference-nulling requests received by an active macro BS are satisfied. Let χ denotes the set ofpico users whose interference-nulling requests to their corresponding nearest active macro BSs are satisfied. In the following lemma, we derive the probability that a randomly chosen pico-user in service belongs to χ.

Lemma 1: The probability φ that the interference-nulling request made by a randomly chosen pico-user to its nearest active macro BS is fulfilled is given by

$$\varphi = \frac{(K_m - T_{min})p_m\lambda_m}{A_p\vartheta_p\lambda_u}\left(1 - \left(1 + 3.5^{-1}\frac{A_p\vartheta_p\lambda_u}{p_m\lambda_m}\right)^{-3.5}\right) - \frac{3.5^{3.5}}{\Gamma(3.5)}\sum_{n=1}^{K_m-T_{min}}\frac{\Gamma(3.5+n)\left(\frac{A_p\vartheta_p\lambda_u}{p_m\lambda_m}\right)^{n-1}(K_m-T_{min}-n)}{n!\left(\frac{A_p\vartheta_p\lambda_u}{p_m\lambda_m}+3.5\right)^{n+3.5}}. \tag{9}$$

Proof: Let Q'_m denotes the number of other requests received by the macro BS, which received nulling request from a randomly chosen pico-user. Then, conditioned on Q'_m, $\varphi = 1$ if $Q'_m+1 \leq K_m - T_{\min}$; otherwise, $\varphi = (K_m - T_{\min})/(Q'_m+1)$. Thus, φ can be expressed as

$$\begin{aligned}\varphi &= \sum_{n=0}^{K_m-T_{\min}-1}\mathbb{P}(Q'_m = n) + \sum_{n=K_m-T_{\min}}^{\infty}\frac{K_m - T_{\min}}{n+1}\mathbb{P}(Q'_m = n) \\ &= \sum_{n=1}^{\infty}\frac{K_m - T_{\min}}{n}\mathbb{P}(Q'_m = n-1) \\ &\quad - \sum_{n=1}^{K_m-T_{\min}}\left(\frac{K_m - T_{\min}}{n} - 1\right)\mathbb{P}(Q'_m = n-1).\end{aligned} \tag{10}$$

By using the fact that the conditional probability density function (PDF) $f'_Y(y)$ of the area of a Voronoi cell given that a randomly chosen user belongs to it is equal to $cyf_Y(y)$, where $f_Y(y)$ is the unconditional PDF and c is a constant such that $\int_o^\infty f'_Y(y)dy = 1$ [22], the PMF of Q'_m can be derived as $\mathbb{P}(Q'_m = n) = (n+1)\mathbb{P}(Q_m = n+1)/\mathbb{E}[Q_m]$, $n \geq 0$. Theorem 1 is then obtained by substituting the PMF of Q'_m in (10), and then using $\sum_{n=1}^{\infty} \mathbb{P}(Q_m = n) = 1 - \mathbb{P}(Q_m = 0)$. ■

C. Channel Model and Precoding Matrices

Assuming standard power law path-loss with exponent α, linear precoding and frequency-flat fading, the received signal z_m at a typical user u located at the origin if $u \in \Phi_u^m$ is given by

$$\begin{aligned} z_m &= \sqrt{P_m} D_m^{-\frac{\alpha}{2}} \mathbf{h}_{b_m,1}^* \mathbf{W}_{b_m} \mathbf{s}_{b_m} \\ &\quad + \sum_{q \in \{m,p\}} \sqrt{P_q} \sum_{x_q \in \Psi_q \setminus b_m} ||x_q||^{-\frac{\alpha}{2}} \mathbf{g}_{x_q,1}^* \mathbf{W}_{x_q} \mathbf{s}_{x_q} + n_m, \end{aligned} \tag{11}$$

where b_m is the serving macro BS at a distance D_m, which is serving M'_m other users simultaneously; $\mathbf{h}_{b_m,1} \sim \mathcal{CN}(\mathbf{0}_{K_m \times 1}, \mathbf{I}_{K_m})$ and $\mathbf{g}_{x_q,1} \sim \mathcal{CN}(\mathbf{0}_{K_q \times 1}, \mathbf{I}_{K_q})$ are the desired and interference complex Gaussian channel vectors from the tagged BS b_m and the interfering BS at x_q, respectively, with independent and identically distributed (i.i.d.) unit variance components; $n_m \sim \mathcal{CN}(0, \sigma^2)$ is complex Gaussian noise with variance σ^2; $\mathbf{s}_{b_m} = [s_{b_m,i}]_{1 \leq i \leq M'_m+1} \in \mathbb{C}^{(M'_m+1) \times 1}$ is the complex-valued signal vector transmitted from b_m to its $M'_m + 1$ served users with the symbol $s_{b_m,1}$ intended for u and $\mathbf{W}_{b_m} = [\mathbf{w}_{b_m,i}]_{1 \leq i \leq (M'_m+1)} \in \mathbb{C}^{K_m \times (M'_m+1)}$ is the corresponding ZF precoding matrix.

Let the channel vectors from the tagged BS b_m to its M'_m users other than u be represented by $[\, \mathbf{h}_{b_m,i} \,]_{2 \leq i \leq M'_m+1}$, and the interference channel vector from the tagged BS to $O = \min(Q_m, K_m - T_{\min})$pico users chosen for interference nulling by $\mathbf{F} = [\, \mathbf{f}_i \,]_{1 \leq i \leq O} \in \mathbb{C}^{K_m \times O}$. Under the perfect CSI assumption, the ZF precoding matrix $\mathbf{W}_{b_m} = [\mathbf{w}_{b_m,i}]_{1 \leq i \leq (M'_m+1)}$ is designed such that $|\mathbf{h}_{b_m,j}^* \mathbf{w}_{b_m,j}|^2$ is maximized for each $j = 1, 2, \ldots, M'_m + 1$, while satisfying the orthogonality conditions $\mathbf{h}_{b_m,j}^* \mathbf{w}_{b_m,i} = 0$ for $\forall i \neq j$ and $\mathbf{f}_i^* \mathbf{w}_{b_m,j} = 0, \forall i = 1, 2, \ldots, O, \forall j = 1, 2, \ldots, M'_m + 1$. It can be achieved by choosing $\mathbf{w}_{b_m,i}$ in the direction of the projection of $\mathbf{h}_{b_m,i}$ on $\text{Null}\big([\, \mathbf{h}_{b_m,j} \,]_{1 \leq j \leq (M'_m+1), j \neq i}, [\, \mathbf{f}_i \,]_{1 \leq i \leq O}\big)$. The nullspace is $K_m - M'_m - O$ dimensional and thus, the desired channel power gain $\beta_{b_m} = |\mathbf{h}_{b_m,1}^* \mathbf{w}_{b_m,1}|^2 \sim \text{Gamma}(\Delta_m, 1)$, where $\Delta_m = K_m - M'_m - O$ [36]. Given that an interfering macro BS at x_m is serving M_m users simultaneously, $\mathbf{W}_{x_m} = [\mathbf{w}_{x_m,i}]_{1 \leq i \leq M_m} \in \mathbb{C}^{K_m \times M_m}$, which is designed independent of $\mathbf{g}_{x_m,1}$. Assuming that the precoding matrix has linearly independent unit norm columns, $\mathbf{g}_{x_m,1}^* \mathbf{w}_{x_m,1}$, $\mathbf{g}_{x_m,1}^* \mathbf{w}_{x_m,2}$, $\ldots$, $\mathbf{g}_{x_m,1}^* \mathbf{w}_{x_m,M_m}$ are i.i.d. complex Gaussian random variables (RVs), and their squared norms are i.i.d. exponential RVs. Thus, the interference channel power gain $\zeta_{x_m} = ||\mathbf{g}_{x_m,1}^* \mathbf{W}_{x_m}||^2 \sim \text{Gamma}(M_m, 1)$, as it is a sum of M_m i.i.d. exponential RVs [18].

A feasible choice of the precoding matrix $\mathbf{W}_{b_m} = [\mathbf{w}_{b_m,i}]_{1 \leq i \leq (M'_m+1)}$ is the pseudo inverse[2] of $\tilde{\mathbf{H}}_{b_m}^*$, i.e., $\mathbf{W}_{b_m} = \tilde{\mathbf{H}}_{b_m} (\tilde{\mathbf{H}}_{b_m}^* \tilde{\mathbf{H}}_{b_m})^{-1}$ with normalized columns, where $\tilde{\mathbf{H}}_{b_m} = [\tilde{\mathbf{h}}_{b_m,i}]_{1 \leq i \leq (M'_m+1)} \in \mathbb{C}^{K_m \times (M'_m+1)}$, $\tilde{\mathbf{h}}_{b_m,i} = (\mathbf{I}_{K_m} - \mathbf{F}(\mathbf{F}^*\mathbf{F})^{-1}\mathbf{F}^*)\mathbf{h}_{b_m,i}$ being the projection of $\mathbf{h}_{b_m,i}$ on the nullspace of $\mathbf{F} = [\, \mathbf{f}_i \,]_{1 \leq i \leq O}$ [31], [36].

[2]Pseudo inversion of the channel matrix is an easy choice of ZF precoding [7].

Similarly, the received signal z_p at u when $u \in \Phi_u^p$ is

$$\begin{aligned} z_p &= \sqrt{P_p} D_p^{-\frac{\alpha}{2}} \mathbf{h}_{b_p,1}^* \mathbf{W}_{b_p} \mathbf{s}_{b_p} + \xi \\ &\quad + \sum_{q \in \{m,p\}} \sqrt{P_q} \sum_{x_q \in \Psi_q \setminus \{v_m, b_p\}} ||x_q||^{-\frac{\alpha}{2}} \mathbf{g}_{x_q,1}^* \mathbf{W}_{x_q} \mathbf{s}_{x_q} + n_p, \end{aligned} \tag{12}$$

where

$$\xi = \begin{cases} 0, & \text{if } u \in \chi \\ \sqrt{P_m} V_m^{-\frac{\alpha}{2}} \mathbf{g}_{v_m,1}^* \mathbf{W}_{v_m} \mathbf{s}_{v_m}, & \text{if } u \notin \chi; \end{cases} \tag{13}$$

b_p is the serving pico BS at a distance D_p, which is serving M'_p other users simultaneously; $n_p \sim \mathcal{CN}(0, \sigma^2)$ is complex Gaussian noise, v_m is the nearest active macro BS to u at a distance V_m, which receives an interference-nulling request from u. The ZF precoding matrix $\mathbf{W}_{b_p} = [\mathbf{w}_{b_p,i}]_{1 \leq i \leq (M'_p+1)}$ is given by $\mathbf{H}_{b_p} (\mathbf{H}_{b_p}^* \mathbf{H}_{b_p})^{-1}$ with normalized columns, where $\mathbf{H}_{b_p} = [\mathbf{h}_{b_p,i}]_{1 \leq i \leq (M'_p+1)} \in \mathbb{C}^{K_p \times (M'_p+1)}$ is the channel matrix from the tagged BS b_p to its $M'_p + 1$ servedpico users. The desired channel power gain $\beta_{b_p} = ||\mathbf{h}_{b_p,1}^* \mathbf{W}_{b_p}||^2 = |\mathbf{h}_{b_p,1}^* \mathbf{w}_{b_p,1}|^2 \sim \text{Gamma}(\Delta_p, 1)$, where $\Delta_p = K_m - M'_p$, and the interference channel power gain $\zeta_{x_p} = ||\mathbf{g}_{x_p,1}^* \mathbf{W}_{x_p}||^2 \sim \text{Gamma}(M_p, 1)$ given that the interfering pico BS at x_p is serving M_p users simultaneously.

D. Distance to the Serving BS and the BS Receiving Interference Nulling Request

The distance D_l to the serving BS from a typical user $u \in \Phi_u^l$ is a RV, and the corresponding PDFs for each $l \in \{m, p\}$ are derived in the following lemma.

Lemma 2: The PDF $f_{D_m}(r)$ of the distance D_m between the serving macro BS and a typical user u when $u \in \Phi_u^m$ is given by

$$f_{D_m}(r) = \frac{2\pi \lambda_m}{A_m} r \exp(-\pi(\lambda_m + \lambda_p/\rho^2) r^2), \tag{14}$$

and the PDF $f_{D_p}(r)$ of the distance D_p between the serving pico BS and a typical user u when $u \in \Phi_u^p$ is given by

$$f_{D_p}(r) = \frac{2\pi \lambda_p}{A_p} r \exp(-\pi(\lambda_m \rho^2 + \lambda_p) r^2). \tag{15}$$

Proof: Given that $u \in \Phi_u^m$, D_m is the distance to the nearest macro BS from u. The cumulative distribution function (CDF) $F_{D_m}(r) = \mathbb{P}(D_m \leq r)$ is thus given by

$$\begin{aligned} F_{D_m}(r) &= \mathbb{P}(X_m \leq r | u \in \Phi_u^m) = \frac{\mathbb{P}(X_m \leq r, u \in \Phi_m^m)}{\mathbb{P}(u \in \Phi_u^m)} \\ &= \frac{1}{A_m} \int_0^r \mathbb{P}\Big(X_p > \frac{y}{\rho}\Big) f_{X_m}(y) dy. \end{aligned} \tag{16}$$

The PDF $f_{D_m}(r)$ in (14) is obtained by differentiating (16) with respect to r and then applying the probability distributions of Rayleigh RVs X_m and X_p. The PDF $f_{D_p}(r)$ is similarly derived. ■

Another quantity of interest is the distance V_m between a typical pico-user in service and its nearest active macro BS to which it requests interference nulling.

Lemma 3: The conditional PDF of the distance V_m between a typical user $u \in \Phi_u^p$ and the macro BS to which it request interference nulling, given that its distance to the serving pico BS is $D_p = r$, is given by

$$f_{V_m|D_p}(r_1|r) = 2\pi p_m \lambda_m r_1 \exp\left(-\pi p_m \lambda_m (r_1^2 - \rho^2 r^2)\right), \quad r_1 > \rho r. \tag{17}$$

Proof: Given that $u \in \Phi_u^p$, V_m is the distance to the nearest active macro BS. The conditional complementary cumulative distribution function (CCDF) of V_m is thus given by

$$\begin{aligned}\bar{F}_{V_m|D_p}(r_1|r) &= \mathbb{P}(X'_m \geq r_1 | u \in \Phi_u^p, D_p = r)\\ &= \mathbb{P}(X'_m \geq r_1 | X_m > \rho r),\end{aligned} \tag{18}$$

where $X'_m = \min_{x_m \in \Psi_m} \|x_m\|$ is the distance from the origin to the nearest active macro BS. The condition $X_m > \rho r$ implies that no points of Φ_m are within a circle of radius ρr. Thus, no points of Ψ_m as well are within ρr because Ψ_m is the thinned version of Φ_m. Thus, given that no active macro BS is closer than ρr, the probability of no active macro BS closer than r_1 is equal to the probability that no points of Ψ_m are within an annulus centered at the origin with inner radius ρr and outer radius r_1. The conditional CCDF $\bar{F}_{V_m|Dp}(r_1|r)$ is thus given by

$$\bar{F}_{V_m|D_p}(r_1|r) = \exp\left(-\pi p_m \lambda_m (r_1^2 - \rho^2 r^2)\right). \tag{19}$$

The conditional PDF of V_m in (17) is obtained by differentiating (19) with respect to r_1. ■

III. SIR Coverage Analysis

We consider interference-limited scenario, and thus derive the SIR coverage probability in this section. The SIR coverage, i.e., the probability that the SIR of a typical user is greater than a given threshold γ is defined as $\mathrm{P}(\gamma) = \mathbb{P}(\mathrm{SIR} > \gamma)$, where $\mathrm{SIR} = \sum_{l \in \{m,p\}} \mathbf{1}(u \in \Phi_u^l)\, \mathrm{SIR}_l$. From (11) and (12) and the discussion that follows, the SIR of a typical user u at the origin when it belongs to Φ_u^l can be expressed as

$$\mathrm{SIR}_l = \frac{P_l \beta_{b_l} D_l^{-\alpha}}{I_{b_l,m} + I_{b_l,p}}, \quad \forall l \in \{m, p\}, \tag{20}$$

where $I_{b_l,m}$ and $I_{b_l,p}$ are the interference powers from the macro and pico tiers, respectively when $u \in \Phi_u^l$, $l \in \{m, p\}$, and are given by

$$\begin{aligned}
I_{b_p,p} &= P_p \sum_{x_p \in \Psi_p \setminus b_p} \zeta_{x_p} ||x_p||^{-\alpha}\\
I_{b_p,m} &= \begin{cases} P_m \sum_{x_m \in \Psi_m \setminus \upsilon_m} \zeta_{x_m} ||x_m||^{-\alpha} & \text{if } u \in \chi\\ P_m \sum_{x_m \in \Psi_m} \zeta_{x_m} ||x_m||^{-\alpha} & \text{if } u \notin \chi,\end{cases}\\
I_{b_m,p} &= P_p \sum_{x_p \in \Psi_p} \zeta_{x_p} ||x_p||^{-\alpha}\\
I_{b_m,m} &= P_m \sum_{x_m \in \Psi_m \setminus b_m} \zeta_{x_m} ||x_m||^{-\alpha}.
\end{aligned} \tag{21}$$

By using the law of total probability, the SIR coverage probability of a typical user u is

$$\mathrm{P}(\gamma) = \mathrm{P}_m(\gamma) A_m + \mathrm{P}_p(\gamma) A_p, \tag{22}$$

where $A_l = \mathbb{P}(u \in \Phi_u^l)$, $l \in \{m, p\}$ is the tier association probability, and $\mathrm{P}_m(\gamma) = \mathbb{P}(\mathrm{SIR}_m > \gamma | u \in \Phi_u^m)$, and $\mathrm{P}_p(\gamma) = \mathbb{P}(\mathrm{SIR}_p > \gamma | u \in \Phi_u^p)$ are the conditional coverage probabilities of the user u when associated with the macro and pico tiers, respectively. To evaluate (22), we first derive the Laplace transform (LT) of the total interference power received by u.

Lemma 4: The LT $\mathcal{L}_{I_{b_p}}(s)$ of the total interference power $I_{b_p} = I_{b_p,m} + I_{b_p,p}$ received by u when $u \in \Phi_u^p$ conditional on $D_p = r$ and $V_m = r_1$ is given by

$$\mathcal{L}_{I_{b_p}}(s) = \left(\varphi \mathcal{L}^1_{I_{b_p,m}}(s) + (1-\varphi)\mathcal{L}^2_{I_{b_p,m}}(s)\right) \mathcal{L}_{I_{b_p,p}}(s), \tag{23}$$

where $\mathcal{L}_{I_{b_p,p}}(s)$ is the LT of $I_{b_p,p}$; $\mathcal{L}^1_{I_{b_p,m}}(s) = \mathcal{L}_{I_{b_p,m}}(s|u \in \chi)$, and $\mathcal{L}^2_{I_{b_p,m}}(s) = \mathcal{L}_{I_{b_p,m}}(s|u \notin \chi)$ are the LTs of $I_{b_p,m}$ conditional on $u \in \chi$ and $u \notin \chi$, respectively. These LTs are given by (24)-(26)*, as shown at the top of the next page, where ${}_2F_1(a, b, c, z)$ is the Gauss Hypergeometric function [37].*

Proof: The LT $\mathcal{L}_{I_{b_p,q}}(s) = \mathbb{E}[e^{-sI_{b_p,q}}]$, $\forall q \in \{m, p\}$ can be derived as

$$\mathcal{L}_{I_{b_p,q}}(s) = \mathbb{E}_{\hat{\Psi}_q} \prod_{x_l \in \hat{\Psi}_q} \mathbb{E}_{\zeta_{x_q}}\left[\exp(-sP_q \zeta_{x_q} ||x_q||^{-\alpha})\right], \tag{28}$$

where $\hat{\Psi}_p = \Psi_p \setminus b_p$, and $\hat{\Psi}_m = \Psi_m \setminus \upsilon_m$ if $u \in \chi$, else $\hat{\Psi}_m = \Psi_m$. By performing the expectation over the distribution of $\zeta_{x_q} \sim \mathrm{Gamma}(M_q, 1)$ conditioned on M_q, and then applying the probability generating functional of PPP with density $p_q \lambda_q$ [34], and finally taking the expectation over the PMF of M_q, we have

$$\begin{aligned}\mathcal{L}_{I_{b_p,q}}(s) = \exp\Bigg\{ -\pi p_q \lambda_q \varpi_{p,q}^2 \Bigg(\sum_{i=1}^{L^q_{\max}} \mathbb{P}(M_q = i)\\ \times {}_2F_1\Big[i, -\frac{2}{\alpha}, \frac{\alpha-2}{\alpha}, -\frac{P_q}{\varpi_{p,q}^\alpha}s\Big] - 1\Bigg)\Bigg\},\end{aligned} \tag{29}$$

where $\varpi_{p,q}$ is the lower bound on the distance to the closest interferer from u in the tier $q \in \{m, p\}$. Thus, $\varpi_{p,p} = r$, and $\varpi_{p,m} = r_1$ if $u \in \chi$; otherwise, $\varpi_{p,m} = \rho r$. ■

Similarly, the LT of $I_{b_m} = I_{b_m,m} + I_{b_m,p}$ conditional on $D_m = r$ can be derived as $\mathcal{L}_{I_{b_m}}(s) = \mathcal{L}_{I_{b_m,m}}(s)\mathcal{L}_{I_{b_m,p}}(s)$, where $\mathcal{L}_{I_{b_m,q}}(s)$, $q \in \{m, p\}$ is given by (27) shown at the top of the next page, with $\varpi_{m,m} = r$ and $\varpi_{m,p} = r/\rho$.

Having derived the LTs, we now evaluate $\mathrm{P}_l(\gamma) = \mathbb{P}\left(P_l \beta_{b_l} D_l^{-\alpha} > \gamma I_{b_l} | u \in \Phi_u^l\right)$, $\forall l \in \{m, p\}$. Conditional on $D_l = r$, $V_m = r_1$ and $\Delta_l = n$, we have

$$\mathrm{P}_l(\gamma | r, r_1, \Delta_l = n) = \sum_{l=0}^{n-1} \frac{(-s)^l}{l!} \frac{\mathrm{d}^l}{\mathrm{d}s^l}\left(\mathcal{L}_{I_{b_l}}(s)\right)\Big|_{s=\frac{\gamma r^\alpha}{P_l}}, \tag{30}$$

which follows from the distribution $\mathrm{Gamma}(n, 1)$ of β_{b_l} for a given $\Delta_l = n$, and the differentiation property of LT. Since the LTs in (24)-(27) are composite functions, (30) requires evaluating lth derivatives of composite functions. These derivatives

$$\mathcal{L}^1_{I_{b_p,m}}(s) = \exp\left\{-\pi p_m \lambda_m r_1^2\left(\sum_{i=1}^{L_{\max}^m}\mathbb{P}(M_m=i)\,{}_2F_1\left[i,-\frac{2}{\alpha},\frac{\alpha-2}{\alpha},-\frac{P_m s}{r_1^{\alpha}}\right]-1\right)\right\} \tag{24}$$

$$\mathcal{L}^2_{I_{b_p,m}}(s) = \exp\left\{-\pi p_m \lambda_m \rho^2 r^2\left(\sum_{i=1}^{L_{\max}^m}\mathbb{P}(M_m=i)\,{}_2F_1\left[i,-\frac{2}{\alpha},\frac{\alpha-2}{\alpha},-\frac{P_m s}{\rho^{\alpha} r^{\alpha}}\right]-1\right)\right\} \tag{25}$$

$$\mathcal{L}_{I_{b_p,p}}(s) = \exp\left\{-\pi p_p \lambda_p r^2\left(\sum_{i=1}^{L_{\max}^p}\mathbb{P}(M_p=i)\,{}_2F_1\left[i,-\frac{2}{\alpha},\frac{\alpha-2}{\alpha},-\frac{P_p s}{r^{\alpha}}\right]-1\right)\right\} \tag{26}$$

$$\mathcal{L}_{I_{b_m,q}}(s) = \exp\left\{-\pi p_q \lambda_q \varpi_{m,q}^2\left(\sum_{i=1}^{L_{\max}^q}\mathbb{P}(M_q=i)\,{}_2F_1\left[i,-\frac{2}{\alpha},\frac{\alpha-2}{\alpha},-\frac{P_q}{\varpi_{m,q}^{\alpha}}s\right]-1\right)\right\} \tag{27}$$

are computed by using Faà di Bruno's formula expressed in terms of integer partition, which is introduced in the following section.

A. Integer Partition and Faà di Bruno's Formula

Integer partition is a partition of a positive integer n as a sum of positive integers. The set of all possible partitions of n is represented by $\mathbf{\Omega}_n$ with the number of partitions denoted by $\mathcal{P}(n)$. The integer 4, for example, can be partitioned as $\mathbf{\Omega}_4 = \{\{4\},\{3,1\},\{2,2\},\{2,1,1\},\{1,1,1\}\}$. Thus, $\mathcal{P}(4)=5$. Let ω_i^n denotes the number of elements in the ith partition p_i^n of n. Also, let μ_{ij}^n denotes the number of positive integer $j \in \{1,2,\ldots,n\}$ in that partition, and a_{ik}^n denotes the kth element ($k \in \{1,2,\ldots,\omega_i^n\}$). *Example*: for the second partition of integer 4 in $\mathbf{\Omega}_4$, i.e., $p_2^4 = \{3,1\}$, we have $\omega_2^4 = 2$, $\mu_{21}^4 = 1$, $\mu_{22}^4 = 0$, $\mu_{23}^4 = 1$, $\mu_{24}^4 = 0$, $a_{21}^4 = 3$, $a_{22}^4 = 1$. For any partition p_i^n, we have the properties $\sum_{j=1}^n j\mu_{ij}^n = n$ and $\sum_{j=1}^n \mu_{ij}^n = \omega_i^n$.

Faà di Bruno's formula for the lth derivative of the composite function $y(t(s))$ in terms of integer partition can be expressed as

$$y_s^{(l)}(t(s)) = \sum_{o=1}^{\mathcal{P}(l)} c_o^l\, y_{t(s)}^{(\omega_o^l)}(t(s)) \prod_{q=1}^{l}\left(t_s^{(q)}(s)\right)^{\mu_{oq}^l}, \tag{31}$$

where

$$c_o^l = \frac{l!}{\prod_{k=1}^{\omega_o^l} a_{ok}^l! \prod_{q=1}^{l}\mu_{oq}^l!},$$

and $y_{t(s)}^{(k)}(t(s))$ is the kth derivative of the function $y(t(s))$ with respect to $t(s)$. Note that the integer partition version has much lesser number of summations as compared to the set partition version used in [21]. The complexity of the numerical computation is thus significantly reduced.

Theorem 1: The SIR coverage probability of a typical pico-user u is given by

$$\mathrm{P}_p(\gamma) = \varphi \mathrm{T}_1(\gamma) + (1-\varphi)\mathrm{T}_2(\gamma), \tag{32}$$

where $\mathrm{T}_1(\gamma) = \mathbb{P}(SIR_p > \gamma | u \in \Phi_u^p, u \in \chi)$ *and* $\mathrm{T}_2(\gamma) = \mathbb{P}(SIR_p > \gamma | u \in \Phi_u^p, u \notin \chi)$ *are the conditional coverage probabilities of a typical pico-user u when* $u \notin \chi$ *and* $u \in \chi$*, respectively. These conditional probabilities can be computed by using* (33) *and* (34)*, as shown at the top of the next page, where* $\delta = P_m/P_p$*, and the function* $\Xi_q^l(\varsigma,\kappa,\varepsilon)$ *is defined as*

$$\Xi_q^l(\varsigma,\kappa,\varepsilon) = \sum_{i=1}^{L_{max}^l}\left(\frac{(i)_q(-\frac{2}{\alpha})_q}{(\frac{\alpha-2}{\alpha})_q}\mathbb{P}(M_l=i) \times {}_2F_1\left[i+q,-\frac{2}{\alpha}+q,\frac{\alpha-2}{\alpha}+q,-\varsigma\kappa^{\alpha}\varepsilon\right]\right), \tag{35}$$

where $(a)_q$ *is a Pochhammer symbol.*

Proof: The proof is given in Appendix A. ■

Remark 1: The number of other users served by the BS which is serving the typical user $u \in \Phi_u^l$ *is given by* $M_l' = \min(U_l', L_{max}^l - 1)$*, where* U_l' *is the number of other users in the Voronoi cell to which the user u belongs. The PMF of* U_l' *can be derived as* $\mathbb{P}(U_l' = n) = (n+1)\mathbb{P}(U_l = n+1)/\mathbb{E}[U_l]$*. The PMF of* M_l' *for* $L_{max}^l > 1$ *is thus given by*

$$\mathbb{P}(M_l' = n) = \begin{cases}\mathbb{P}(U_l' = n), & 0 \le n < L_{max}^l - 1\\ 1 - \displaystyle\sum_{k=1}^{L_{max}^l-2}\mathbb{P}(U_l' = k), & n = L_{max}^l - 1,\end{cases}$$
$$\forall l \in \{m,p\}. \tag{36}$$

For $L_{max}^l = 1$, $\mathbb{P}(M_l' = 0) = 1, \forall l \in \{m,p\}$.

Corollary 1: The coverage probability of a typical macro-user $\mathrm{P}_m(\gamma)$ *is given by* (37)*, as shown at the top of the next page, where the PMF of* Δ_m *conditional on* $M_m' = k$ *for* $T_{min} < K_m$ *is given by*

$$\mathbb{P}(\Delta_m = n | M_m' = k)$$
$$= \begin{cases} 1 - \displaystyle\sum_{\upsilon=0}^{K_m - T_{min} - 1}\mathbb{P}(Q_m = \upsilon), & n = T_{min} - k\\ \mathbb{P}(Q_m = K_m - k - n), & T_{min} - k + 1 \le n \le K_m - k.\end{cases} \tag{38}$$

For the special case of $T_{min} = K_m$ *which implies no interference nulling,* $\Delta_m = K_m - M_m'$*, thus* $\mathbb{P}(\Delta_m = K_m - k | M_m' = k) = 1$.

Proof: $\mathrm{P}_m(\gamma)$ is derived in the same way as $\mathrm{T}_2(\gamma)$. However, since $\Delta_m = K_m - M_m' - \min(Q_m, K_m - T_{\min})$ is a function of the two RVs M_m' and Q_m, deconditioning with respect to Δ_m is achieved in two steps, first averaging over the

$$\mathrm{T}_1(\gamma) = 2p_m\lambda_m\frac{\lambda_p}{A_p}\int_{\theta=0}^{\frac{1}{\rho}}\Bigg[\sum_{k=0}^{L^p_{\max}-1}\mathbb{P}(M_p^{'}=k)\sum_{l=0}^{K_p-k-1}\frac{\gamma^l}{l!}\theta^{\alpha l+1}\sum_{o=1}^{\mathcal{P}(l)}c_o^l(-1)^{\omega_o^l}\Big(p_m\lambda_m\,\Xi_0^m(\delta,\theta,\gamma)+p_p\lambda_p\theta^2\,\Xi_0^p(1,1,\gamma)$$
$$+(1-p_m)\lambda_m\rho^2\theta^2+(1-p_p)\lambda_p\theta^2\Big)^{-(\omega_o^l+2)}\Gamma(\omega_o^l+2)\prod_{q=1}^{l}\Big(p_m\lambda_m\delta^q\,\Xi_q^m(\delta,\theta,\gamma)+\frac{p_p\lambda_p}{\theta^{\alpha q-2}}\,\Xi_q^p(1,1,\gamma)\Big)^{\mu_{oq}^l}\Bigg]\mathrm{d}\theta \quad (33)$$

$$\mathrm{T}_2(\gamma) = \frac{\lambda_p}{A_p}\sum_{k=0}^{L^p_{\max}-1}\mathbb{P}(M_p^{'}=k)\sum_{l=0}^{K_p-k-1}\frac{\gamma^l}{l!}\sum_{o=1}^{P(l)}c_o^l(-1)^{\omega_o^l}\Gamma(\omega_o^l+1)\prod_{q=1}^{l}\left(\frac{p_m\lambda_m\delta^q}{\rho^{\alpha q-2}}\,\Xi_q^m\left(\delta,\frac{1}{\rho},\gamma\right)+p_p\lambda_p\,\Xi_q^p(1,1,\gamma)\right)^{\mu_{oq}^l}$$
$$\times\left(p_p\lambda_p\,\Xi_0^p(1,1,\gamma)+(1-p_m)\lambda_m\rho^2+(1-p_p)\lambda_p+p_m\lambda_m\rho^2\,\Xi_0^m\left(\delta,\frac{1}{\rho},\gamma\right)\right)^{-(\omega_o^l+1)} \quad (34)$$

$$\mathrm{P}_m(\gamma) = \frac{\lambda_m}{A_m}\sum_{k=0}^{L^m_{\max}-1}\mathbb{P}(M_m'=k)\sum_{n=T_{\min}-k}^{K_m-k}\mathbb{P}(\Delta_m=n|M_m'=k)\sum_{l=0}^{n-1}\frac{\gamma^l}{l!}\sum_{o=1}^{\mathcal{P}(l)}c_o^l(-1)^{\omega_o^l}\Gamma(\omega_o^l+1)\Bigg((1-p_m)\lambda_m+(1-p_p)\frac{\lambda_p}{\rho^2}$$
$$+p_m\lambda_m\,\Xi_0^m(1,1,\gamma)+\frac{p_p\lambda_p}{\rho^2}\,\Xi_0^p\left(\frac{1}{\delta},\rho,\gamma\right)\Bigg)^{-(\omega_o^l+1)}\prod_{q=1}^{l}\left(p_m\lambda_m\,\Xi_q^m(1,1,\gamma)+p_p\lambda_p\frac{\rho^{\alpha q-2}}{\delta^q}\,\Xi_q^p\left(\frac{1}{\delta},\rho,\gamma\right)\right)^{\mu_{oq}^l} \quad (37)$$

---

conditional PMF of Δ_m for the given M_m', and then averaging over the PMF of M_m'. ■

Remark 2: For the special case of $L^m_{max} = L^p_{max} = 1$,

$$\begin{aligned}\mathrm{P}_p(\gamma) &= \mathrm{P}_p(\gamma|M_p'=0)\\ &= \varphi\mathrm{T}_1(\gamma|M_p'=0)+(1-\varphi)\mathrm{T}_2(\gamma|M_p'=0), \quad (39)\\ \mathrm{P}_m(\gamma) &= \mathrm{P}_m(\gamma|M_m'=0), \quad (40)\end{aligned}$$

where for each $l \in \{m, p\}$,

$$\Xi_q^l(\varsigma,\kappa,\varepsilon) = \Xi_q(\varsigma,\kappa,\varepsilon) = \frac{(1)_q(-\frac{2}{\alpha})_q}{(\frac{\alpha-2}{\alpha})_q} \times {}_2F_1\left(1+q,-\frac{2}{\alpha}+q,\frac{\alpha-2}{\alpha}+q,-\varsigma\kappa^\alpha\varepsilon\right). \quad (41)$$

IV. Rate Analysis

In this section, we analyze the achievable downlink rate of a typical user. We derive the CCDF of downlink rate, also defined as the rate coverage, and the average rate of a typical user.

Assuming adaptive transmission scheme such that the Shannon limit is achieved, and treating the interference as noise, the data rate of a typical user u is given by

$$R = \sum_{l\in\{m,p\}} S_l W\log_2(1+\mathrm{SIR}_l)\mathbf{1}(u\in\Phi_u^l), \quad (42)$$

where S_l is the fraction of resources received by u when $u \in \Phi_u^l$. For each $l \in \{m, p\}$, given that U_l' is the number of other users in the cell to which the user u belongs, the total users in the tagged cell are $U_l'+1$. We assume one RB per time slot with total bandwidth W, and at most $L^l_{\max}$ users served simultaneously in each RB through spatial multiplexing. Thus, if the total number of users in the tagged cell is less than $L^l_{\max}$ (i.e., $U_l' + 1 < L^l_{\max}$), each user can utilize the entire bandwidth W without sharing; thus, $S_l = 1$. However, if $U_l'+1$ is no less than $L^l_{\max}$ (i.e., $U_l' + 1 \geq L^l_{\max}$), we assume that the time-frequency resources are shared equally among the total users; thus, $S_l = L^l_{\max}/(U_l' + 1)$. Hence, the fraction of resources received by $u \in \Phi_u^l$ can be expressed as

$$S_l = \min\left(\frac{L^l_{\max}}{U_l^{'}+1}, 1\right).$$

Theorem 2: The CCDF of the downlink rate of a typical user u, $\mathcal{R}(\upsilon) = \mathbb{P}(R > \upsilon)$ can be expressed as $\mathcal{R}(\upsilon) = A_m\mathcal{R}_m(\upsilon) + A_p\mathcal{R}_p$, where $A_l = \mathbb{P}(u \in \Phi_u^l)$ and $\mathcal{R}_l(\upsilon) = \mathbb{P}(S_l W \log_2(1+SIR_l) > \upsilon)$ is the rate distribution of $u \in \Phi_u^l$. $\mathcal{R}_l(\upsilon)$ for each $l \in \{m, p\}$ is given by (43), as shown at the top of the next page, where $\mathrm{P}_l(\gamma|M_l' = k)$ is the conditional SIR coverage probability of $u \in \Phi_u^l$ for given $M_l' = k$.

Proof: From (42),

$$\mathbb{P}(R>\upsilon) = \sum_{l\in\{m,p\}}\mathbb{P}(u\in\Phi_u^l)\underbrace{\mathbb{P}(S_lW\log_2(1+\mathrm{SIR}_l)>\upsilon)}_{\mathcal{R}_l(\upsilon)}$$

where

$$\begin{aligned}\mathcal{R}_l(\upsilon) &= \mathbb{P}(W\log_2(1+\mathrm{SIR}_l)>\upsilon, U_l'\leq L^l_{\max}-2)\\ &\quad+\mathbb{P}\Big(\frac{L^l_{\max}}{U_l^{'}+1}W\log_2(1+\mathrm{SIR}_l)>\upsilon, U_l'\geq L^l_{\max}-1\Big)\\ &= \sum_{k=0}^{L^l_{\max}-2}\mathbb{P}(\mathrm{SIR}_l>2^{\upsilon/W}-1|U_l'=k)\mathbb{P}(U_l'=k)\\ &\quad+\sum_{k\geq L^l_{\max}-1}\mathbb{P}(\mathrm{SIR}_l>2^{\frac{\upsilon}{W}\frac{(k+1)}{L^l_{\max}}}-1|U_l'=k)\mathbb{P}(U_l'=k).\end{aligned} \quad (44)$$

The conditional SIR coverage probabilities in (44) can be conditioned on the given value of M_l' by using $M_l' = \min(U_l', L^l_{\max} - 1)$. ■

$$\mathcal{R}_l(v) = \sum_{k=0}^{L^l_{\max}-2} \mathrm{P}_l\left(2^{v/W}-1 \,\middle|\, M'_l = k\right)\mathbb{P}(U'_l=k) + \sum_{k\geq L^l_{\max}-1} \mathrm{P}_l\left(2^{\frac{v}{W}\frac{(k+1)}{L^l_{\max}}}-1 \,\middle|\, M'_l = L^l_{\max}-1\right)\mathbb{P}(U'_l=k) \tag{43}$$

$$\bar{R}_l = \frac{W}{\ln 2}\int_0^\infty \frac{1}{1+y}\left[\sum_{k=0}^{L^l_{\max}-2} \mathrm{P}_l\left(y \,\middle|\, M'_l = k\right)\mathbb{P}(U'_l=k) + O_l\, \mathrm{P}_l\left(y \,\middle|\, M'_l = L^l_{\max}-1\right)\right]dy \tag{46}$$

$$O_l = \frac{L^l_{\max}\lambda_l}{A_l\lambda_u}\left(1-\left(1+3.5^{-1}A_l\lambda_u/\lambda_l\right)^{-3.5}\right) - \frac{3.5^{3.5}}{\Gamma(3.5)}\sum_{k=1}^{L^l_{\max}-1}\frac{\Gamma(3.5+k)\left(\frac{A_l\lambda_u}{\lambda_l}\right)^{k-1}L^l_{\max}}{k!(\frac{A_l\lambda_u}{\lambda_l}+3.5)^{3.5+k}} \tag{47}$$

For the special case of $L^l_{\max}=1$, the rate distribution of $u \in \Phi^l_u$ further simplifies to

$$\mathcal{R}_l(v) = \sum_{k\geq 0} \mathrm{P}_l\left(2^{\frac{v}{W}(k+1)}-1\right)\mathbb{P}(U'_l=k). \tag{45}$$

After the rate coverage, we next derive the average data rate of any randomly chosen user.

Theorem 3: The average rate $\bar{R} = \mathbb{E}[R]$ of a typical user u is given by $\bar{R} = A_m\bar{R}_m + A_p\bar{R}_p$, where $\bar{R}_l = \mathbb{E}[S_l W \log_2(1+\mathrm{SIR}_l)]$ is the average rate of $u \in \Phi^l_u$, $l \in \{m,p\}$. $\bar{R}_l$ is given by (46), as shown at the top of this page, where O_l is computed according to (47), shown at the top of this page.

Proof: From (42),

$$\mathbb{E}[R] = \sum_{l\in\{m,p\}} \mathbb{P}(u\in\Phi^l_u)\,\underbrace{\mathbb{E}[S_l W\log_2(1+\mathrm{SIR}_l)]}_{\bar{R}_l},$$

where

$$\begin{aligned}\bar{R}_l &= W\sum_{k=0}^{L^l_{\max}-2}\mathbb{E}\left[\log_2(1+\mathrm{SIR}_l)|U'_l=k\right]\mathbb{P}(U'_l=k)\\ &\quad + W\sum_{k\geq L^l_{\max}-1}\frac{L^l_{\max}}{k+1}\mathbb{E}\left[\log_2(1+\mathrm{SIR}_l)|U'_l=k\right]\mathbb{P}(U'_l=k).\end{aligned} \tag{48}$$

The computation of $\mathbb{E}[\log_2(1+\mathrm{SIR}_l)]$ requires integrating $\log_2(1+\mathrm{SIR}_l)$ with respect to the PDF of SIR_l. However, the integral can be transformed into $1/(\ln 2)\int_0^\infty \mathrm{P}_l(y)(1+y)^{-1}\,dy$ by applying integration by parts, along with the fact that PDF is the negative differentiation of CCDF. Also, we have $M'_l = \min(U'_l, L^l_{\max}-1)$. Equation (48) thus can be simplified to (46), where

$$\begin{aligned}O_l &= \sum_{k\geq L^l_{\max}-1}\frac{L^l_{\max}}{k+1}\mathbb{P}(U'_l=k)\\ &= \sum_{k=1}^{\infty}\frac{L^l_{\max}}{k}\mathbb{P}(U'_l=k-1) - \sum_{k=1}^{L^l_{\max}-1}\frac{L^l_{\max}}{k}\mathbb{P}(U'_l=k-1).\end{aligned}$$

Equation (47) is obtained by substituting $\mathbb{P}(U'_l=k) = (k+1)\mathbb{P}(U_l=k+1)/\mathbb{E}[U_l]$, $k\geq 0$ and further simplifying by using $\sum_{k=1}^{\infty}\mathbb{P}(U_l=k) = 1-\mathbb{P}(U_l=0)$. ∎

For the special case of $L^l_{\max}=1$, the average data rate of $u\in\Phi^l_u$ simplifies to

$$\bar{R}_l = O_l\frac{W}{\ln 2}\int_0^\infty \frac{\mathrm{P}_l(y)}{1+y}\,dy. \tag{49}$$

V. Impact of Limited Feedback on Interference Nulling

The results so far have been derived based on the perfect CSI assumption. However, in practical systems, the CSI is never perfectly accurate. In frequency division duplex systems, the downlink CSI is fed back by the users to serving BSs. Due to the limited feedback, the BSs receive quantized CSI. In this section, we analyze the impact of the quantization error due to limited feedback on the performance of interference nulling. As the focus is on interference-nulling performance, we consider $L^m_{\max} = L^p_{\max} = 1$.

The feedback model is similar to the one used in [31] and [32]. The quantized channel direction information (CDI) is fed back by using a quantization codebook of 2^B unit norm vectors, where B is the number of feedback bits. The codebook is known at both the transmitter and the receiver. Each user feeds back the index of the codeword closest to its channel direction, measured by the inner product. For example, a typical user, when it belongs to the macro tier, uses the codebook $\mathcal{C}_m = \{\mathbf{c}_{m,j} : j = 1,2,\ldots,2^{B_m}\}$ of size 2^{B_m} to quantize the channel direction $\tilde{\mathbf{h}}_{b_m,1} = \frac{\mathbf{h}_{b_m,1}}{||\mathbf{h}_{b_m,1}||}$ from its serving maco BS b_m. The quantized channel direction is

$$\hat{\mathbf{h}}_{b_m,1} = \arg\max_{\mathbf{c}_{m,j}\in\mathcal{C}_m}\left|\tilde{\mathbf{h}}^*_{b_m,1}\mathbf{c}_{m,j}\right|.$$

Similarly, the typical user, when it belongs to the pico tier, uses the codebook $\mathcal{C}_p = \{\mathbf{c}_{p,j} : j = 1,2,\ldots,2^{B_p}\}$ of size 2^{B_p} to quantize the channel direction from its serving pico BS b_p, and the codebook $\mathcal{C}_m = \{\mathbf{c}_{m,j} : j = 1,2,\ldots,2^{B_m}\}$ to quantize the channel direction from its nearest active macro BS v_m. Otherpico users which request v_m for interference nulling, as well as the user served by v_m, also employ codebooks of size 2^{B_m}, but the codebooks differ from user to user to avoid the possibility of receiving the same quantization vector index from different users. The codebooks are generated by using random vector quantization [38], [39], where each vector $\mathbf{c}_{m,j}$ of $\mathcal{C}_m$ and $\mathbf{c}_{p,j}$ of $\mathcal{C}_p$ are independently chosen from the isotropic distribution on the K_m- dimensional and

$$
\begin{aligned}
\mathrm{T}_{1,LF}(\gamma) = {}& 2p_m\lambda_m \frac{\lambda_p}{A_p}\int_{\theta=0}^{\frac{1}{\rho}} \sum_{l=0}^{K_p-1}\left(\frac{\gamma}{\kappa_p}\right)^{l}\theta^{\alpha l+1}\sum_{\upsilon=0}^{l}\frac{(\delta\kappa_I)^{l-\upsilon}}{\upsilon!(1+\delta\kappa_I\gamma/\kappa_p\theta^{\alpha})^{l-\upsilon+1}}\sum_{o=1}^{\mathcal{P}(\upsilon)}c_o^{\upsilon}(-1)^{\omega_o^{\upsilon}}\Gamma(\omega_o^{\upsilon}+2)\\
&\times\left(p_m\lambda_m\,\Xi_0\left(\delta,\theta,\frac{\gamma}{\kappa_p}\right)+p_p\lambda_p\theta^2\,\Xi_0\left(1,1,\frac{\gamma}{\kappa_p}\right)+(1-p_m)\lambda_m\rho^2\theta^2+(1-p_p)\lambda_p\theta^2\right)^{-(\omega_o^{\upsilon}+2)}\\
&\times\prod_{q=1}^{l}\left(p_m\lambda_m\delta^q\,\Xi_q\left(\delta,\theta,\frac{\gamma}{\kappa_p}\right)+\frac{p_p\lambda_p}{\theta^{\alpha q-2}}\,\Xi_q\left(1,1,\frac{\gamma}{\kappa_p}\right)\right)^{\mu_{oq}^{\upsilon}}\mathrm{d}\theta
\end{aligned}
\tag{53}
$$

K_p- dimensional unit spheres, respectively. Since the precoding vectors are now based on quantized CDIs, for the typical user $u\in\Phi_u^m$ served by the macro BS b_m, the desired channel power gain $\hat{\beta}_{b_m}\sim\mathrm{Gamma}(\Delta_m,\kappa_m)$, where $\Delta_m=K_m-\min(Q_m,K_m-T_{\min})$ and $\kappa_m=1-2^{B_m}\mathrm{Beta}(2^{B_m},\frac{K_m}{K_m-1})$ [31]. However, as the precoding vector of the interfering BS at $x_q\in\Psi_q\backslash b_m, q\in\{m,p\}$ is independent of the channel to the typical user u, the interference channel power gain $\hat{\zeta}_{x_q}$ is still distributed as Gamma(1, 1), i.e., Exp[1]. Similarly, for the typical user $u\in\Phi_u^p$ served by the pico BS b_p, the desired channel power gain $\hat{\beta}_{b_p}\sim\mathrm{Gamma}(\Delta_p,\kappa_p)$, where $\Delta_p=K_p$ and $\kappa_p=1-2^{B_p}\mathrm{Beta}(2^{B_p},\frac{K_p}{K_p-1})$. The interference channel power gain from each interfering BS other than υ_m is distributed as Exp[1]. If υ_m does not apply interference nulling, the interference channel power gain from υ_m, $\hat{\zeta}_{\upsilon_m}$ is also distributed as Exp[1]. However, if υ_m applies nulling, unlike the perfect CDI case, where the interference from υ_m is completely nulled, there will be residual interference due to the quantization error. The interference channel power gain in this case is approximated as an exponential RV with mean $\kappa_I=2^{-\frac{B_m}{K_m-1}}$ [31]. Thus, $\hat{\zeta}_{\upsilon_m}\sim\mathrm{Exp}[1/\kappa_I]$, if $u\in\chi$; otherwise $\hat{\zeta}_{\upsilon_m}\sim\mathrm{Exp}[1]$. The SIR of the typical user u can be expressed as

$$
\mathrm{SIR}_l=\frac{P_l\hat{\beta}_{b_l}D_l^{-\alpha}}{\hat{I}_{b_l,m}+\hat{I}_{b_l,p}},\quad \forall l\in\{m,p\},\tag{50}
$$

where

$$
\begin{aligned}
\hat{I}_{b_l,m}&=P_m\sum_{x_m\in\Psi_m\backslash b_l}\hat{\zeta}_{x_m}||x_m||^{-\alpha},\\
\hat{I}_{b_l,p}&=P_p\sum_{x_p\in\Psi_p\backslash b_l}\hat{\zeta}_{x_p}||x_p||^{-\alpha}.
\end{aligned}\tag{51}
$$

Corollary 2: With limited feedback, the coverage probability of a typical pico-user u in the interference-limited scenario is given by

$$
\mathrm{P}_{p,LF}(\gamma)=\mathrm{T}_{1,LF}(\gamma)\varphi+\mathrm{T}_{2,LF}(\gamma)(1-\varphi),\tag{52}
$$

where $\mathrm{T}_{1,LF}(\gamma)$ *is the coverage probability of* $u\in\chi$ *with limited feedback, and is given by* (53) *shown at the top of this page and* $\mathrm{T}_{2,LF}(\gamma)=\mathrm{T}_2\left(\gamma/\kappa_p\right)$ *is the coverage probability of* $u\notin\chi$*, expressed in terms of the corresponding probability for the perfect CSI,* $\mathrm{T}_2(\cdot)$*. Similarly, the coverage probability of a typical macro-user u with limited feedback is given by* $\mathrm{P}_{m,LF}(\gamma)=\mathrm{P}_m\left(\gamma/\kappa_m\right)$.

Proof: Due to the limited feedback, even when a typical pico-user u belongs to χ, it receives residual interference $Y=P_m\hat{\zeta}_mV_m^{-\alpha}$ from its nearest active macro BS, where $\hat{\zeta}_m\sim\mathrm{Exp}[1/\kappa_I]$. Thus, the LT of total macro tier interference when $u\in\chi$ is given by

$$
\begin{aligned}
\mathcal{L}_{\hat{I}_{b_p,m}}(s|u\in\chi)&=\mathcal{L}^1_{I_{b_p,m}}(s)\mathbb{E}[e^{-sY}]\\
&=\mathcal{L}^1_{I_{b_p,m}}(s)(1+sP_m\kappa_Ir_1^{-\alpha})^{-1},
\end{aligned}
$$

where $\mathcal{L}^1_{I_{b_p,m}}(s)$ is the LT of the total macro tier interference for the perfect CSI in (24). The LT of the total pico tier interference $\mathcal{L}_{\hat{I}_{b_p,p}}(s)$ is equal to $\mathcal{L}_{I_{b_p,p}}$ in (26). Since $\hat{\beta}_{b_p}\sim\mathrm{Gamma}(K_p,\kappa_p)$, $\mathrm{T}_{1,LF}(\gamma)$ can then be derived in the same way as $\mathrm{T}_1(\gamma)$ in Theorem 1 with γ replaced by γ/κ_p. For $\mathrm{T}_{2,LF}(\gamma)$ and $\mathrm{P}_{m,LF}(\gamma)$, since the LTs of interference powers are the same as those of the perfect CSI case, $\mathrm{T}_{2,LF}(\gamma)$ is given by (34) with γ replaced by γ/κ_p, and similarly $\mathrm{P}_{m,LF}(\gamma)$ by (37) with γ replaced by γ/κ_m. ■

Note that $\mathrm{T}_{2,LF}(\gamma)$ and $\mathrm{P}_{m,LF}(\gamma)$ reduce to $\mathrm{T}_2(\gamma)$ and $\mathrm{P}_m(\gamma)$, respectively, if $\kappa_m=\kappa_p=1$. Similarly, if $\kappa_p=1$ and $\kappa_I=0$, by using $0^0=1$, $\mathrm{T}_{1,LF}(\gamma)$ also reduces to $\mathrm{T}_1(\gamma)$. After deriving the coverage probabilities for limited feedback, the rate coverage and average rate can be obtained by using Theorem 2 and Theorem 3, respectively, with $P_l(\cdot)$ replaced by $P_{l,LF}(\cdot)$.

VI. Simulation and Numerical Results

In this section, we validate our analytical results via Monte Carlo simulations on a square window of 20×20 Km^2 and present numerical analysis to provide insights into important design parameters. Unless otherwise stated, we set $\delta=\frac{P_m}{P_p}=100$, $\lambda_m=1$BS/Km^2 and $W=1$ MHz.

The average data rate (Theorem 3) for perfect CSI, and the data rate distribution (Theorem 2) for both the perfect CSI and limited feedback scenarios are validated via Monte Carlo simulations for different system configurations in Figure 1.a and Figure 1.b, respectively. The analytical and simulation results match with each other quite well in these figures. The PPP based assumptions of the thinned processes Φ_u^m, Φ_u^p and Ψ_u^p obtained from the parent process Φ_u hardly impact the probability distributions of the number of users of corresponding sets in a typical cell. The small gaps between the simulations and analytical curves are thus mostly due to the approximation of cell area distribution by Gamma. Note that the validation of Theorem 3 for perfect CSI naturally validates the conditional SIR distributions

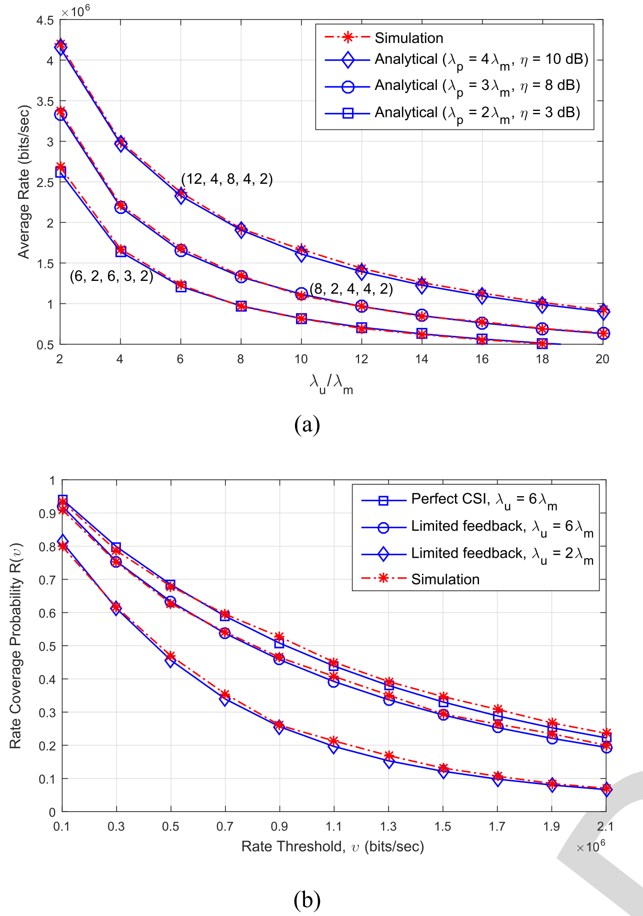


Fig. 1. (a) Validation of the average user data rate (Theorem 3) with perfect CSI for different values of λ_p, η and $(K_m, L^m_{\max}, T_{\min}, K_p, L^p_{\max})$; (b) Validation of the rate coverage probability (Theorem 2) for both the perfect CSI and limited feedback scenarios: $K_m = 12$, $K_p = 4$, $L^m_{\max} = L^p_{\max} = 1$, $T_{\min} = 2$, $\lambda_u = 10\lambda_m$, $\alpha = 3.5$, $\eta = 15$dB.

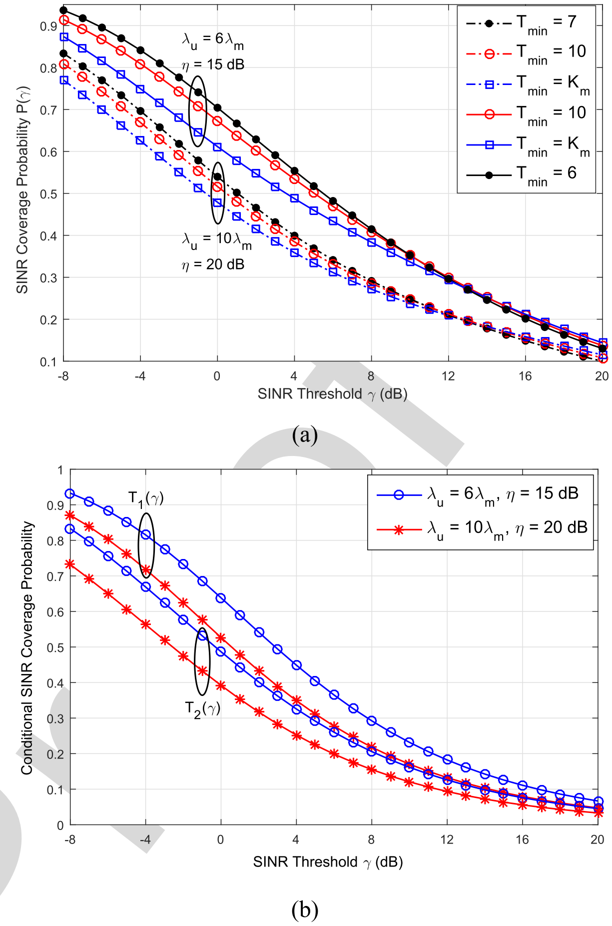


Fig. 2. Impact of interference nulling on the SIR coverage probability: $K_m = 14$, $L^m_{\max} = 4$, $K_p = 6$, $L^p_{\max} = 4$, $\lambda_p = 6\lambda_m$, $\alpha = 3.5$.

derived in Theorem 1 and Corollary 1, and the validation of Theorem 2 for limited feedback validates the SIR distribution in Corollary 2. In Figure 1.a, the average data rate decreases with an increase in user density λ_u because of the increase in interference and the decrease in users' share of resources. The interference power increases with an increase in user density because not just more BSs become active, but the average channel power gain from each interfering BS also increases until the number of users associated with the BS exceeds $L^l_{\max}$.

In Figure 2, we analyze the impact of interference nulling on the SIR coverage probability, where $T_{\min} = K_m$ implies no interference nulling employed. While the overall SIR coverage of a typical user is plotted in Figure 2.a, the coverage probability conditioned that the user belongs to pico tier and always gets the interference from its nearest active macro BS nulled, $T_1(\gamma)$ is compared against that its no-nulling counterpart, $T_2(\gamma)$ in Figure 2.b. Figure 2.a reveals that with properly chosen $T_{\min}$, the SIR coverage can be significantly improved with interference nulling. For example, if the required SIR level for a typical user to be under coverage is 0 dB, the average fraction of users under coverage improves from 61% to 70% with interference nulling for the $\lambda_u = 6\lambda_m$, $\eta = 15$ dB case. In both Figure 2.a and Figure 2.b, the performance gain decreases with an increasing threshold. At smaller values of thresholds, as interference nulling improves the SIRs of poor cell-edge pico-user lacking coverage due to strong interference from their corresponding nearest active macro BSs, the coverage probability of thepico users significantly improves. On the other hand, we know that the SIR of a typical macro-user degrades due to interference nulling as it costs the user its available DoF. At lower values of SIR thresholds, the degradation in SIR is, however, not significant enough to impact its coverage probability. Thus, the overall gain in coverage probability is high at smaller threshold levels. However, at larger threshold values, the users under coverage are basically those in the cell interior. Thus, interference nulling may not significantly improve the already high SIR of cell-interiorpico users, resulting in minimal improvement in pico coverage probability. The SIR degradation of macro-users due to interference nulling, which do not have any significance on macro coverage probability at lower thresholds eventually causes the coverage probability to degrade after certain level. This degradation further reduces the overall gain in coverage probability.

In Figure 2.a, the performance gain in the overall coverage probability for $\lambda_u = 10\lambda_m$, $\eta = 20$ dB is relatively low compared to the $\lambda_u = 6\lambda_m$, $\eta = 15$ dB case. However, in Figure 2.b, given that the nulling is performed for each pico-

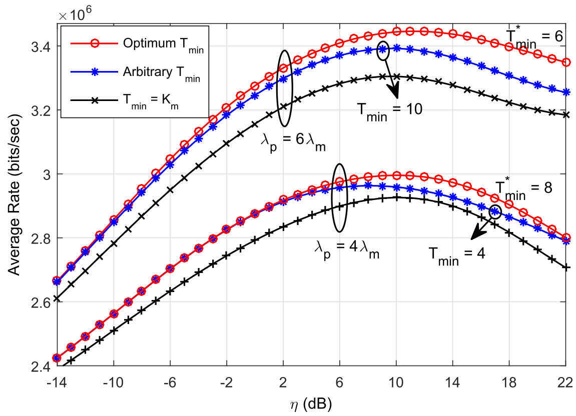


Fig. 3. Effect of pico cell density λ_p on the optimal choices of $T_{\min}$ and η: $\lambda_u = 6\lambda_m$, $K_m = 12$, $L^m_{\max} = 4$, $K_p = 4$, $L^p_{\max} = 4$, $\alpha = 4$.

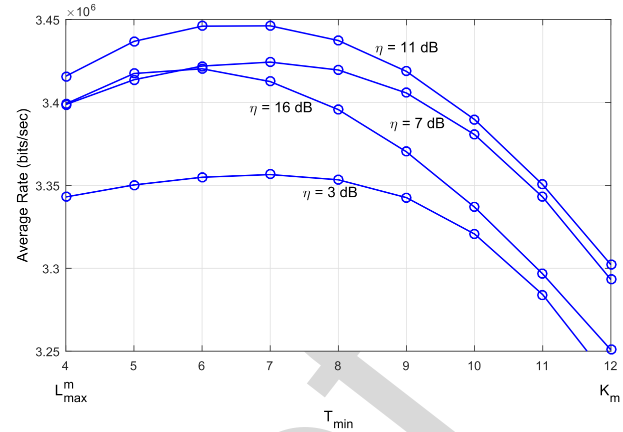


Fig. 4. Average rate vs. $T_{\min}$ for different values of η: $\lambda_u = 6\lambda_m$, $\lambda_p = 6\lambda_m$, $K_m = 12$, $L^m_{\max} = 4$, $K_p = 4$, $L^p_{\max} = 4$, $\alpha = 4$.

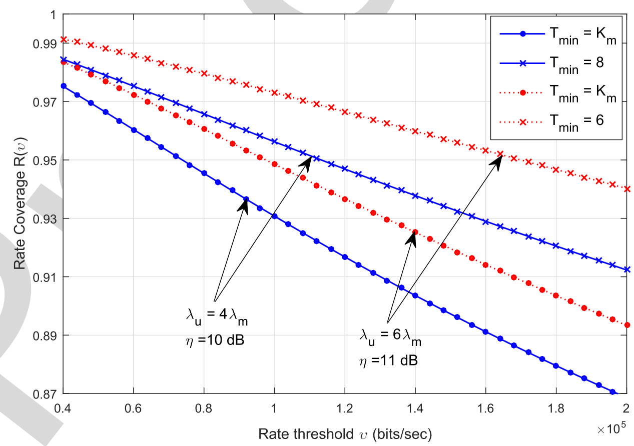


Fig. 5. Effect of interference nulling on cell-edge data rate: $\lambda_p = 6\lambda_m$, $K_m = 12$, $L^m_{\max} = 4$, $K_p = 4$, $L^p_{\max} = 4$, $\alpha = 4$.

user, both cases have similar gains in pico coverage probability due to nulling. Thus, the reason for the lower performance gain for higher user density λ_u and higher bias η is the lack of sufficient resources for interference nulling. For the $\lambda_u = 6\lambda_m$ and $\eta = 15$ dB case, with $T_{\min} = 6$, interference to 83% of thepico users from their corresponding nearest active macro BSs are nulled. The fraction of interference nulledpico users reduces to 53% for $\lambda_u = 10\lambda_m$ and $\eta = 20$ dB, with optimal $T_{\min}$ of 7.

Next, we investigate the optimal value of η to maximize the average user data rate. η controls the number of users offloaded from the macro to the pico tier to obtain a balanced distribution of the user load across tiers so that the radio resources are better utilized in each tier. Meanwhile, since $T_{\min}$ determines the spatial DoF available for serving the macro-users, as well as the number of interference-nulledpico users, $T_{\min}$ must be tuned according to user offloading. The joint tuning of $T_{\min}$ and η for optimal average data rate is investigated in Figure 3. The optimal pair $(\eta, T_{\min})$ is found to be (10 dB, 8) and (11 dB, 6) for pico density $\lambda_p = 4\lambda_m$ and $\lambda_p = 6\lambda_m$, respectively. For the given user density, the optimal $T_{\min}$ decreases with the increase in pico density because the number of interference-nulling requests received by a typical active macro BS increases with the increase in pico density. Thus, the allocated interference-nulling resources $(K_m - T_{\min})$ must be increased.

The variation in the average rate with $T_{\min}$ for the given value of η is plotted in Figure 4. The average rate of the macro-users increases with an increasing $T_{\min}$ due to the increase in the spatial DoF available at each macro BS for serving its own users. In contrast, the average pico rate decreases with an increasing $T_{\min}$ due to the decrease in the number of interference nulledpico users. The net result is the initial increase in the average rate with an increasing $T_{\min}$ and the subsequent decrease beyond a certain value of $T_{\min}$. The optimal $T_{\min}$ shifts towards the lower values as the value of η increases. For example, the optimal $T_{\min}$ of 7 for $\eta = 3$ dB decreases to 6 for $\eta = 11$ dB and to 5 for $\eta = 16$ dB. With an increasing η, more users are offloaded to the pico tier. Thus, allocating more antenna resources for interference nulling is desirable.

In Figure 5, the rate coverage corresponding to the optimal pair $(\eta, T_{\min})$ which maximized the average rate in Figure 3 for $\lambda_p = 4\lambda_m$ and $\lambda_p = 6\lambda_m$ is plotted. Let the 5th percentile rate R_{95}, which corresponds to the 5th percentile of the users with rate less than R_{95} (i.e., $\mathcal{R}(R_{95}) = 0.95$), be considered as the cell-edge data rate. For $\lambda_p = 4\lambda_m$ and $\eta = 10$ dB, $T_{\min} = 8$, which maximized the average rate is found to improve the cell-edge rate from 7.2×10^4 bits/sec to 1.12×10^5 bits/sec as compared to that without interference nulling. Similarly, for $\lambda_p = 6\lambda_m$, the cell-edge rate improves from 9.6×10^4 bits/sec to 1.68×10^5 bits/sec if interference nulling with $T_{\min} = 6$ is employed corresponding to $\eta = 11$ dB.

In Figure 6, the average data rate is assessed for different values of $L^m_{\max}$ and $L^p_{\max}$ with and without interference nulling. The curve corresponding to the interference nulling employed is plotted by computing the average rate with optimum $T_{\min}$ for each corresponding value of $L^m_{\max}$ and $L^p_{\max}$. As Figure 6 reveals, the average data rate can be significantly improved by selecting a proper value of $L^m_{\max}$ compared to either SU-BF or full-SDMA, and similarly a proper value of $L^p_{\max}$. For the case with no interference nulling employed, in which all the antennas at each macro BS are used for serving its own users, the variation of $L^m_{\max}$ has little or no impact on the average

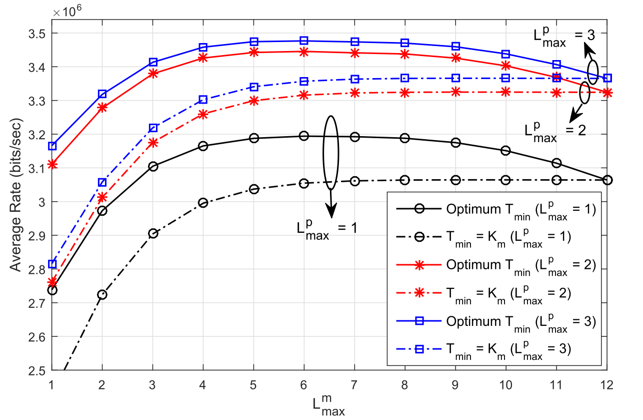


Fig. 6. Average rate vs. $L^m_{\max}$ for different values of $L^p_{\max}$ with optimum $T_{\min}$ and no interference nulling: $\lambda_p = 6\lambda_m$, $\lambda_u = 6\lambda_m$, $K_m = 12$, $K_p = 4$, $\eta = 12$ dB, $\alpha = 4$.

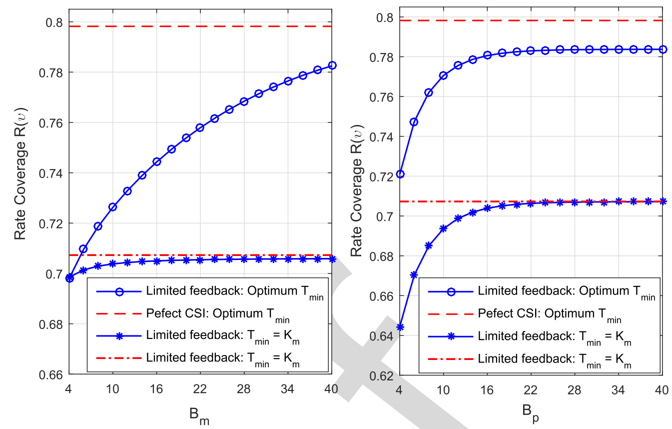


Fig. 7. Impact of number of feedback bits on the rate coverage performance: $\lambda_p = 6\lambda_m$, $\lambda_u = 10\lambda_m$, $K_m = 12$, $K_p = 4$, $L^m_{\max} = L^p_{\max} = 1$, $\eta = 15$ dB, $\alpha = 3.5$.

rate from $L^m_{\max} = 7$ to $L^m_{\max} = 12$. This result can be observed for each given value of $L^p_{\max}$ because beyond $L^m_{\max} = 7$, the number of users simultaneously served by a macro BS in each time slot is limited by the number of users in that cell, rather than $L^m_{\max}$. This explanation is further corroborated by the fact that with interference nulling employed, the optimal $T_{\min}$ beyond $L^m_{\max} = 7$ is found to be the corresponding $L^m_{\max}$ itself, which is the minimum possible value of $T_{\min}$. Since beyond $L^m_{\max} = 7$, the number of macro-users in a cell is typically less than $L^m_{\max}$, allocating more antenna resources than $L^m_{\max}$ would be wasting resources as those surplus resources can be utilized for performance improvement through interference nulling. For each possible value of $L^p_{\max}$, the optimal pair $(L^m_{\max}, T_{\min})$ which maximizes the average rate is found to be $(6, 7)$. The average rate slightly degrades for $L^p_{\max} = 4$ as compared to $L^p_{\max} = 3$ (not shown in the figure). Thus, the optimal values of $L^m_{\max}$, $T_{\min}$, and $L^p_{\max}$ for the given system configuration are 6, 7, and 3, respectively.

After numerically analyzing the proposed SDMA scheme with interference nulling for the perfect CSI, we now investigate the impact of limited feedback on the performance. As explained in Section V, each macro-user feeds back B_m CSI bits to its home BS. In contrast, each pico-user feeds back B_p CSI bits to its home BS and B_m CSI bits to its nearest active macro BS if the BS is performing interference nulling to the user. In Figure 7, the impact of the number of feedback bits B_m and B_p on the rate coverage with and without interference nulling is investigated. As the number of feedback bits increases, the performance approaches that of the perfect CSI. Clearly, the impact of limited feedback bits B_m on the performance is higher for the interference-nulling scenario than that without nulling. $B_m > 16$, which is more than sufficient for the non-coordination case, appears to be insufficient for interference nulling case to reap the full benefits of nulling. Nevertheless, nulling does improve performance even with limited feedback as compared to the non-coordination case. With no interference nulling employed, the feedback bits B_m are only required for signal power boosting to the single user being served in the cell and such processing is found to be less sensitive to CSI errors as compared to interference nulling. If we observe the rate coverage curve against B_p for the non-coordination case, $B_p > 20$ is near perfect. However, we can observe a performance gap for interference nulling case even beyond $B_p = 20$ because of the limitation in B_m, which is considered to be 40 in this case.

VII. Conclusion

We analyzed the downlink performance of multi-antenna HetNets with SDMA, in which the ZF precoding matrix at macro BS also considered interference nulling to certainpico users. Further, the number of users served with SDMA in each cell was a function of user distribution. Our results showed that the SIR and rate coverage of victimpico users (those suffering strong interference from macro BS) can be significantly improved with the proposed interference nulling scheme if $T_{\min}$ is carefully chosen. The optimal choice of $T_{\min}$ for maximum data rate was found to be coupled with association bias. The optimal values of $L^m_{\max}$ and $L^p_{\max}$ which maximize the average data rate was also investigated and were found to outperform both SU-BF and full-SDMA. The impact of CSI quantization error on the performance of interference nulling due to limited feedback was also analyzed. It was observed that interference nulling is highly sensitive to CSI errors as the residual interference due to CSI imperfection significantly degrades the performance. However, depending on the degree of CSI imperfection, the performance may still be better than that without interference nulling.

Appendix

A. Proof of Theorem 1

By substituting (23) into (30), followed by $\Delta_p = K_p - M'_p$, and then averaging over the joint PDF of D_p and V_m, expressed as $f_{V_m|D_p}(r_1) f_{D_p}(r)$, and the PMF of M'_p, we get (32), where

$$\mathrm{T}_1(\gamma) = \int_{r=0}^{\infty} \int_{r_1=\rho r}^{\infty} \sum_{k=0}^{L^p_{\max}-1} \mathbb{P}(M'_p = k) \sum_{l=0}^{K_p-k-1} \frac{(-s)^l}{l!} \times \frac{\mathrm{d}^l}{\mathrm{d}s^l} \left(\mathcal{L}^1_{I_{bp,m}}(s) \mathcal{L}_{I_{bp,p}}(s) \right) \Big|_{s=\frac{\gamma r^\alpha}{P_p}} f_{V_m|D_p}(r_1|r) f_{D_p}(r)\, dr_1\, dr, \quad (54)$$

and $\mathrm{T}_2(\gamma)$ is given by a similar expression with $\mathcal{L}^1_{I_{bp,m}}(s)$ replaced by $\mathcal{L}^2_{I_{bp,m}}(s)$. However, since the LT in $\mathrm{T}_2(\gamma)$ is not a function of r_1, averaging over the PDF of D_p only is required. We thus derive $\mathrm{T}_1(\gamma)$ first, as $\mathrm{T}_2(\gamma)$ then follows immediately.

Let $y(s) = e^{-\pi s}$, and $t(s) = p_m \lambda_m r_1^2 \, \Xi_0^m \left(1, 1, \frac{P_m}{r_1^\alpha} s\right) + p_p \lambda_p r^2 \, \Xi_0^p \left(1, 1, \frac{P_p}{r^\alpha} s\right)$. The LT in (53) can be expressed as $\mathcal{L}^1_{I_{bp,m}}(s) \mathcal{L}_{I_{bp,p}}(s) = e^{\pi \left(p_m \lambda_m r_1^2 + p_p \lambda_p r^2\right)} y(t(s))$, the lth derivative of which can be evaluated by applying Faà di Bruno's formula (31). While computing the lth derivative, we use $y_{t(s)}^{(\omega_o^l)}(t(s)) = (-\pi)^{\omega_o^l} \exp(-\pi t(s))$;

$$\frac{\mathrm{d}^q}{ds^q} \Xi_0^l \left(1, 1, \frac{P_l}{\varpi_l^\alpha} s\right) = \left(-\frac{P_l}{\varpi_l^\alpha}\right)^q \Xi_q^l \left(1, 1, \frac{P_l}{\varpi_l^\alpha} s\right), \tag{55}$$

which follows from the property of the Gauss Hypergeometric function; and the properties of integer partition $\sum_{q=1}^{l} q \mu_{oq}^l = l$ and $\sum_{q=1}^{l} \mu_{oq}^l = \omega_o^l$. The final expression for $\mathrm{T}_1(\gamma)$ in (33) is then obtained by changing the order of integration, followed by substituting $\frac{r}{r_1} \to \theta$, $r_1 \to r_1$, then integrating with respect to r_1.

References

[1] "LTE Advanced: Heterogeneous networks," Q. Incorp., White Paper, Jan. 2011.

[2] S. Alamouti, "A simple transmit diversity technique for wireless communications," *IEEE J. Sel. Areas Commun.*, vol. 16, no. 8, pp. 1451–1458, Oct. 1998.

[3] H. Jafarkhani, *Space-Time Coding: Theory and Practice*. Cambridge, U.K.: Cambridge Univ. Press, 2005.

[4] M. Kang and M. S. Alouini, "Largest eigenvalue of complex Wishart matrices and performance analysis of MIMO MRC systems," *IEEE J. Sel. Areas Commun.*, vol. 21, no. 3, pp. 418–426, Apr. 2003.

[5] G. J. Foschini, "Layered space-time architecture for wireless communication in a fading environment when using multi-element antennas," *Bell Labs Techn. J.*, vol. 5, no. 2, pp. 41–59, Autumn 1996.

[6] D. Gesbert, M. Kountouris, R. W. Heath, Jr., C.-B. Chae, and T. Sälzer, "From single user to multiuser communications: Shifting the MIMO paradigm," *IEEE Signal Process. Mag.*, vol. 24, no. 5, pp. 36–46, Sep. 2007.

[7] Q. H. Spencer, A. L. Swindlehurst, and M. Haardt, "Zero-forcing methods for downlink spatial multiplexing in multiuser MIMO channels," *IEEE Trans. Signal Process.*, vol. 52, no. 2, pp. 461–471, Feb. 2004.

[8] J. G. Andrews, W. Choi, and R. W. Heath, Jr., "Overcoming interference in spatial multiplexing MIMO cellular networks," *IEEE Wireless Commun.*, vol. 14, no. 6, pp. 95–104, Dec. 2007.

[9] J. Hoydis, S. ten Brink, and M. Debbah, "Massive MIMO in the UL/DL of cellular networks: How many antennas do we need?" *IEEE J. Sel. Areas Commun.*, vol. 31, no. 2, pp. 160–171, Feb. 2013.

[10] D. Ying, H. Yang, T. L. Marzetta, and D. J. Love, "Heterogeneous massive MIMO with small cells," in *Proc. IEEE Veh. Technol. Conf. (VTC-Spring)*, Nanjing, China, May 2016, pp. 1–6.

[11] Y. Kim *et al.*, "Full dimension MIMO (FD-MIMO): The next evolution of MIMO in LTE systems," *IEEE Wireless Commun. Mag.*, vol. 21, no. 3, pp. 92–100, Jun. 2014.

[12] Q. Ye, O. Y. Bursalioglu, H. C. Papadopoulos, C. Caramanis, and J. G. Andrews, "User association and interference management in massive MIMO HetNets," *IEEE Trans. Commun.*, vol. 64, no. 5, pp. 2049–2065, May 2016.

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

[14] M. D. Renzo and P. Guan, "Stochastic geometry modeling of coverage and rate of cellular networks using the Gil-Pelaez inversion theorem," *IEEE Commun. Lett.*, vol. 18, no. 9, pp. 1575–1578, Sep. 2014.

[15] L. Afify, H. ElSawy, T. Al-Naffouri, and M. Alouini, "A Unified stochastic geometry model for MIMO cellular networks with retransmissions," *IEEE Trans. Wireless Commun.*, vol. 15, no. 12, pp. 8595–8609, Dec. 2016.

[16] V. Chandrasekhar, M. Kountouris, and J. G. Andrews, "Coverage in multi-antenna two-tier networks," *IEEE Trans. Wireless Commun.*, vol. 10, no. 10, pp. 5314–5327, Oct. 2009.

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

[18] H. S. Dhillon, M. Kountouris, and J. G. Andrews, "Downlink MIMO HetNets: Modeling, ordering results and performance analysis," *IEEE Trans. Wireless Commun.*, vol. 12, no. 10, pp. 5208–5222, Oct. 2013.

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

[20] C. Li, J. Zhang, J. G. Andrews, and K. B. Letaief, "Success probability and area spectral efficiency in multiuser MIMO HetNets," *IEEE Trans. Commun.*, vol. 64, no. 4, pp. 1544–1556, Apr. 2016.

[21] S. T. Veetil, K. Kuchi, and R. K. Ganti, "Performance of PZF and MMSE receivers in cellular networks with multi-user spatial multiplexing," *IEEE Trans. Wireless Commun.*, vol. 14, no. 9, pp. 4867–4878, Sep. 2015.

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

[23] M. Cierny, H. Wang, R. Wichman, Z. Ding, and C. Wijting, "On number of almost blank subframes in heterogeneous cellular networks," *IEEE Trans. Wireless Commun.*, vol. 12, no. 10, pp. 5061–5073, Oct. 2013.

[24] Y. Dhungana and C. Tellambura, "Multi-channel analysis of cell range expansion and resource partitioning in two-tier heterogeneous cellular networks," *IEEE Trans. Wireless Commun.*, vol. 15, no. 3, pp. 2306–2394, Mar. 2016.

[25] T. D. Novlan, R. K. Ganti, A. Ghosh, and J. G. Andrews, "Analytical evaluation of fractional frequency reuse for heterogeneous cellular networks," *IEEE Trans. Commun.*, vol. 60, no. 7, pp. 2029–2039, Jul. 2012.

[26] A. Shojaeifard, K. Hamdi, E. Alsusa, D. So, and J. Tang, "Design, modeling, and performance analysis of multiantenna heterogeneous cellular networks," *IEEE Trans. Commun.*, vol. 64, no. 7, pp. 3104–3118, Jul. 2016.

[27] J. Zhang, R. Chen, J. G. Andrews, A. Ghosh, and R. W. Heath, Jr., "Networked MIMO with clustered linear precoding," *IEEE Trans. Wireless Commun.*, vol. 8, no. 4, pp. 1910–1921, Apr. 2009.

[28] M. Feng, X. She, L. Chen, and Y. Kishiyama, "Enhanced dynamic cell selection with muting scheme for DL CoMP in LTE-A," in *Proc. IEEE Veh. Technol. Conf. (VTC-Spring)*, Taipei, Taiwan, May 2010, pp. 1–5.

[29] Y. Wu, Y. Cui, and B. Clerckx, "Analysis and optimization of inter-tier interference coordination in downlink multi-antenna HetNets with offloading," *IEEE Trans. Wireless Commun.*, vol. 14, no. 12, pp. 6550–6564, Dec. 2015.

[30] 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.

[31] J. Zhang and J. G. Andrews, "Adaptive spatial intercell interference cancellation in multicell wireless networks," *IEEE J. Sel. Areas Commun.*, vol. 28, no. 9, pp. 1455–1468, Dec. 2010.

[32] 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. 1419–1430, Apr. 2015.

[33] 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.

[34] S. N. Chiu, D. Stoyan, W. S. Kendall, and J. Mecke, *Stochastic Geometry and its Applications*, 3rd ed. Hoboken, NJ, USA: Wiley, 2013.

[35] J.-S. Ferenc and Z. Neda, "On the size distribution of Poisson-Voronoi cells," *Phys. A, Statist. Mech. Appl.*, vol. 385, no. 2, pp. 518–526, 2007.

[36] N. Jindal, J. G. Andrews, and S. Weber, "Multi-antenna communication in ad hoc networks: Achieving MIMO gains with SIMO transmission," *Phys. A, Statist. Mech. Appl.*, vol. 59, no. 2, pp. 529–540, Feb. 2011.

[37] I. S. Gradshteyn and I. M. Ryzhik, *Table of Integrals, Series, and Products*, 7th ed. San Diego, CA, USA: Academic, 2007.

[38] N. Jindal, "MIMO broadcast channels with finite rate feedback," *IEEE Trans. Inf. Theory*, vol. 52, no. 11, pp. 5045–5060, Nov. 2006.

[39] C. K. Au-yeung and D. J. Love, "On the performance of random vector quantization limited feedback beamforming in a MISO system," *IEEE Trans. Wireless Commun.*, vol. 6, no. 2, pp. 458–462, Feb. 2007.

Yamuna Dhungana (S'15) received the B.Eng. degree in electronics and communications engineering from the Institute of Engineering, Tribhuvan University, Nepal, in 2009, the M.Eng. degree in telecommunications from the Asian Institute of Technology, Thailand, in 2011, and the Ph.D. degree in electrical and computer engineering from the University of Alberta, Edmonton, AB, Canada in 2016. She is currently a Research and Development Engineer with Ericsson Canada Inc., Ottawa, ON, Canada. Her research interests include wireless communications theory, heterogeneous cellular networks, and MIMO systems.

Chintha Tellambura (F'11) received the B.Sc. degree (Hons.) from the University of Moratuwa, Moratuwa, Sri Lanka, in 1986, the M.Sc. degree in electronics from the King's College, University of London, London, U.K., in 1988, and the Ph.D. degree in electrical engineering from the University of Victoria, Victoria, BC, Canada, in 1993.

He was a Post-Doctoral Research Fellow with the University of Victoria from 1993 to 1994, and the University of Bradford from 1995 to 1996. He was with Monash University, Melbourne, Australia, from 1997 to 2002. He is currently a Professor with the Department of Electrical and Computer Engineering, University of Alberta. His current research interests include the design, modelling and analysis of cognitive radio, heterogeneous cellular networks and 5G wireless networks.

He served as an editor for IEEE TRANSACTIONS ON COMMUNCIATIONS from 1999 to 2011 and for TRANSACTIONS ON WIRELESS COMMUNICATIONS from 2001 to 2007. He was the Area Editor for Wireless Communications Systems and Theory in the IEEE TRANSACTIONS ON WIRELESS COMMUNICATIONS from 2007 to 2012. He and the co-authors received the Communication Theory Symposium Best Paper Award in the 2012 IEEE International Conference on Communications, Ottawa, Canada. He is the winner of the prestigious McCalla Professorship and the Killam Annual Professorship from the University of Alberta. He has authored/coauthored over 500 journal and conference publications with total citations more than 12,000, and has an h-index of 59 in Google Scholar.

Performance Analysis of SDMA with Inter-tier Interference Nulling in HetNets

Yamuna Dhungana, *Student Member, IEEE*, and Chintha Tellambura, *Fellow, IEEE*

***Abstract*— The downlink performance of two-tier (macro/pico) multi-antenna cellular heterogeneous networks employing space division multiple access (SDMA) technique with zero-forcing precoding is analyzed in this paper. The number of users simultaneously served with SDMA by a base-station (BS) depends on the number of active users in its cell, with the maximum served users limited to $L_{\max}$. To protect the pico users from severe macro-interference, part of the antennas at each macro BS is proposed to be utilized toward interference nulling to pico users. The partitioning of macro antenna resources to serve macro-users and to null interference to pico users for optimal performance is investigated in this paper. Biased-nearest-distance-based user association scheme is proposed, where the bias value accounts for the natural bias due to the differences in multi-antenna transmission schemes across tiers, as well as the artificial bias for load balancing. The signal-to-interference-ratio coverage probability, rate distribution, and average rate of a typical user are then derived. Our results demonstrate that the proposed interference nulling scheme has strong potential for improving performance if the macro antennas partitioning is carefully done. The optimal $L^*_{\max}$ for both macro and pico-tier, which maximize the average data rate, is also investigated and it is found to outperform both single-user beamforming and full-SDMA. Finally, the impact of imperfect channel state information due to limited feedback is analyzed.**

***Index Terms*— Heterogeneous networks (HetNets), interference nulling, limited feedback, Poisson point process (PPP), space division multiple access (SDMA), stochastic geometry.**

I. Introduction

NETWORK densification (dense deployment of base-stations (BSs)) and multi-antenna techniques are well-known for their tremendous potential to increase spectral efficiency of wireless networks. In a conventional macro only cellular network, where the locations of high-power macro BSs are strictly planned, adding more BSs can be very challenging for dense urban areas due to extremely high site acquisition cost. Thus, the cost-effective way of network densification is to deploy a diverse set of low-power BSs within the areas covered by macro cellular infrastructure [1]. The resulting network of mixed types of BSs is known as heterogeneous network (HetNet). If the BSs are equipped with multiple antennas, the additional degrees of freedom (DoF) in the spatial dimension can be utilized in a number of ways, for example, to improve the spectral efficiency, and to enhance the link reliability. The diversity and spatial multiplexing gains have been extensively studied in general for point-to-point links without interference. Some examples of diversity techniques are space-time coding [2], [3] and coherent processing known as beamforming [4]. The spatial multiplexing which utilizes the multiple antennas to transmit independent data streams simultaneously over spatial sub-channels, has been explored in [5]. Space division multiple access (SDMA) which allows multiple users to be served simultaneously on the same time-frequency resource has also been analyzed [6], [7]. However, in interference-prone cellular networks, for example, a dense deployed HetNet, where complex interference scenarios may arise due to power disparities between the BSs, the effectiveness of spatial multiplexing may diminish [8]. Nevertheless, if the available spatial DoF are intelligently utilized to suppress/mitigate interference as well as to harvest diversity and multiplexing gain, the performance of cellular networks can be improved. In this paper, we develop a tractable framework to analyze the downlink performance of zero-forcing (ZF) precoding based joint SDMA and inter-tier interference-nulling scheme in HetNets.

Manuscript received June 12, 2016; revised November 12, 2016; accepted January 4, 2017. This work was supported by the Alberta Innovates - Technology Futures (AITF) Graduate Student Scholarship. This paper was presented in part at IEEE ICC, Kuala Lampur, Malaysia, May 2016. The associate editor coordinating the review of this paper and approving it for publication was P. Salvo Rossi.

Y. Dhungana is with Ericsson Canada Inc., Ottawa, ON K2K 2V6, Canada (e-mail: dhungana@ualberta.ca).

C. Tellambura is with the Department of Electrical and Computer Engineering, University of Alberta, Edmonton, AB T6G 1H9, Canada (e-mail: chintha@ece.ualberta.ca).

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TWC.2017.2656083

A. Related Work and Contributions of the Paper

Although multiple antenna in wireless communications is a mature technology, its incorporation into cellular networks, traditional single tier, as well as HetNets, has received much momentum both in academic research and standardization efforts only recently with the introduction of massive-MIMO concept [9]–[12]. By utilizing the stochastic geometry framework which enables systematic modeling of interference, several studies on the modeling and analysis of downlink single-tier multi-antenna cellular networks have been reported in the literature. For example, error probability analysis by using the equivalent-in-distribution approach in [13], coverage and rate analyses using the Gil-Pelaez inversion theorem in [14], and a unified approach to error probability, outage and rate analyses for different multi-antenna configurations with retransmissions in [15]. Apart from single-tier networks, stochastic geometric modeling of downlink multi-antenna HetNets have been significantly explored as well. Reference [16] compared the signal-to-interference-and-noise ratio (SINR) coverage of SU-BF with that of ZF SDMA for a two-tier multi-antenna HetNet by considering a single fixed-radius circular macro cell with multiple femto cells of fixed radii, distributed according to a Poisson point process (PPP) within the macro cell. However, since BS-user association and macro-tier interference are ignored, the insights in [16] may not be accurate for practical HetNets. The coverage probability and average link spectral efficiency of ZF precoding in multi-antenna HetNet, spatially averaged over a given cell of known radius and guard region are derived in [17]. Unlike the spatial averaging over a given cell in [17], system-wide spatial averaging is considered in [18] and the upper bounds on coverage probability of ZF SDMA and SU-BF are derived. The ordering results for the coverage probability and rate per user performance of SDMA, SU-BF and single-antenna transmission are also derived in [18] by using tools from stochastic orders. While the analysis in [18] is based on maximum instantaneous SINR based BS-user association, association rules intended to maximize the average receive SINR (and thus, the SINR coverage), and biased association for optimal rate coverage are proposed for multi-antenna HetNets in [19]. Closed form expressions for the signal-to-interference ratio (SIR) of ZF SDMA and SU-BF are derived in [20] for user association based on the received power of the reference signal transmitted from a single-antenna with total power. In all of these downlink multi-antenna HetNet analyses [16]–[20], each cell of a tier is assumed to be spatially multiplexing to the same number of users, say L, and it can be any arbitrary integer in the interval $[1, K_i]$, where K_i is the number of antennas at a BS of the ith tier. This assumption, however, is not suitable for cellular networks because the number of users, which depends on user distribution, is generally different from one cell to another. An open-loop SDMA with each antenna serving an independent data stream to its user with the limiting requirement that the number of users in each cell must be at least equal to the number of transmit antennas is analyzed in [21] for single-tier cellular networks with ZF and MMSE receivers. In this paper, we consider user-distribution dependent SDMA scheme, i.e., the number of users simultaneously served with SDMA in each cell depends on the total number of users in that cell. If the number of users in a cell is less than the maximum number of users served per resource block (RB), say $L_{\max}$, all the users are simultaneously served; otherwise only $L_{\max}$ users chosen randomly are served.

One of the key challenges in downlink cellular HetNets is inter-tier interference management. Due to large transmit power disparities between macro and small-cell nodes such as picos and femtos, and proactive user offloading from macro to small cells, interference management between the macro and pico/femto tiers is very important because the performance of small-cell cell-edge users could be severely degraded. While almost blank subframes (ABSF) [22], [23] and frequency-domain resource partitioning [24], [25] can be used, inter-tier interference can be more efficiently managed without compromising time/frequency resources by using multiple antennas. Inter-tier interference mitigation by using multiple receive antennas at the user devices is analyzed in [26]. In this paper, we analyze ZF-precoding based *interference-nulling* method by using BS antennas to suppress the interference from the macro tier to small-cell users. Compared to other potential techniques such as joint transmission [27] and transmission point selection [28], which require both user data and channel state information (CSI) to be shared between the coordinating BSs, interference nulling requires only CSI to be shared. Joint transmission with local precoding, which requires no CSI exchange between the coordinating BSs, is studied in [12]. However, it stills requires user data sharing, which could be very challenging due to backhaul overhead. In [29], interference nulling to U offloadedpico users by each macro BS is analyzed, where the optimal U for maximum rate coverage is also investigated. However, unlike [29] which considers a single served user per RB in each cell, we consider a user-distribution dependent SDMA scheme. SU-BF with interference nulling to a fixed number of neighboring-cells users at each BS of any tier for general multi-tier HetNets is analyzed in [30], without specifying how these users are selected. SU-BF with interference nulling in single-tier cellular networks is studied in [31] and [32]. Although SU-BF with interference nulling has been relatively well analyzed, to the best of our knowledge, this paper is the first work to analyze a user-distribution dependent SDMA scheme with inter-tier interference nulling in cellular HetNets. The main contributions of this paper are summarized as follows.

1) We develop a tractable framework to analyze a user-distribution dependent SDMA scheme in a two-tier (macro/pico) multi-antenna HetNet with ZF precoding, in which the number of users simultaneously served by a BS in an RB depends on the number of active users in its cell. The framework also allows the analysis of SU-BF and full-SDMA by setting the limit on the number of users served per RB to one, and the total number of transmit antennas, respectively.
2) To suppress the detrimental macro-to-pico interference, interference-nulling precoding, jointly with user-distribution dependent SDMA, is proposed. That is, the precoding matrix at each macro BS is designed to null interference to a set of activepico users while spatially multiplexing the macro-users in the cell. In the proposed interference-nulling scheme, the candidatepico users for interference nulling from a macro BS, say b, are the ones which have b as their nearest interfering macro BS.
3) Considering the complexity of BS-user association in multi-antenna HetNets, a simple biased-nearest-distance based association rule is introduced, in which the bias value accounts for the natural bias required for SINR maximization in multi-antenna HetNets, as well as the artificial bias for load balancing.
4) By considering interference limited scenario, we derive analytical expressions for the SIR and rate distributions, as well as the average rate of a typical user. We then perform comprehensive analysis to investigate the optimal association bias, and the inherent trade-off between interference cancellation, signal power boosting and spatial multiplexing. The following useful network design insights are obtained from these analyses:
   a) By optimizing the maximum number of users simultaneously served per RB, SDMA can achieve

significantly higher average data rate than both SU-BF and full SDMA.

b) If the number of users in a typical cell is less than the maximum number of users served per RB, say L_{max}, the optimal number of antennas towards spatial multiplexing and signal power boosting of local users is found to be L_{max}.Thus, rather than allocating additional antennas to these users, the average data rate can be significantly increased if the surplus antennas are used towards interference nulling topico users.

c) The optimal number of antennas towards interference nulling topico users increases with the increase in pico cell density, as well as association bias.

5) Finally, the impact of the CSI quantization error due to limited feedback on interference nulling is also investigated.

The paper is organized as follows. The system model and the proposed multi-antenna technique are presented in Section II. Section III derives the SIR distribution. The rate coverage and the average rate are derived in Section IV. In Section V, the impact of limited feedback is analyzed. The numerical results are presented in Section VI, and the concluding remarks in Section VII.

II. System Model

We consider the downlink of a two-tier multi-antenna HetNet comprising macro and pico BSs spatially distributed on $\mathbb{R}^2$ plane as independent homogeneous PPPs Φ_m with density λ_m and Φ_p with density λ_p, respectively. The macro BSs are equipped with K_m transmit antennas, and the pico BSs with K_p antennas. Similarly, users are assumed to be distributed according to an independent PPP Φ_u with density λ_u, and each has a single receive antenna. The two network tiers share the same spectrum with the universal frequency reuse.

The transmission scheme is SDMA with ZF precoding applied at each BS to serve multiple users simultaneously in each RB. We assume only one RB per time slot. As the BSs and users are independently distributed on the $\mathbb{R}^2$ plane, the number of users varies across cells. Thus, in our proposed SDMA scheme, a typical active macro cell with $N_m \geq 1$ users serves $M_m = \min(N_m, L^M_{\max})$ users simultaneously in a given time slot, where $L^M_{\max}$ is the maximum number of users it can serve. If $N_m > L^M_{\max}$, the BS choses $L^M_{\max}$ users for service randomly, else, all N_m users are served. Similarly, $M_p = \min(N_p, L^P_{\max})$ users are simultaneously served by a typical active pico cell in a given time slot, which has $N_p \geq 1$ users, and $L^P_{\max}$ is the maximum number the pico cell can serve. The macro and pico BSs transmit to each of their users with power P_m and P_p, respectively.

A. User Association

According to the user association rule introduced in [19] for average SINR maximization, a typical user at the origin is associated with the nearest pico BS if $P_p\sqrt{\Delta_p \tau_p} X_p^{-\alpha} \geq P_m\sqrt{\Delta_m \tau_m} X_m^{-\alpha}$, and otherwise, is associated with the nearest macro BS, where $X_m = \min_{x_m \in \Phi_m} \|x_m\|$ and $X_p = \min_{x_m \in \Phi_p} \|x_p\|$ are the distances from the origin to the nearest macro and pico BSs, respectively. If associated with the macro tier, Δ_m is the average desired channel gain from the nearest macro BS, and τ_p is the average interference channel gain from the nearest pico BS. Similarly, Δ_p and τ_m are the corresponding values, if associated with the pico tier. These channel gains depend on the number of users served with SDMA. This association rule is thus not suitable for our proposed SDMA scheme, where the number of users served with SDMA in each cell is a function of the number of users in that cell. The number of users, on the other hand, is determined by the association rule. The above rule however can be equivalently expressed as follows: a user is associated with the pico tier only if

$$X_m \geq \left(\frac{P_m}{P_p}\right)^{\frac{1}{\alpha}} \left(\frac{1}{\varrho}\right)^{\frac{1}{\alpha}} X_p, \tag{1}$$

where $\varrho = \sqrt{\frac{\Delta_p \tau_p}{\Delta_m \tau_m}}$. If we compare (1) with the popular received power based association in HetNets [24], [33], ϱ can be interpreted as the natural bias required for average SINR maximization in multi-antenna HetNets due to the differences in transmission schemes. This coverage maximization bias, however, may not always achieve optimum load balancing for maximum rate. Thus, by further introducing an artificial bias B for load balancing, the resultant condition for pico tier association becomes $X_m \geq \rho X_p$, which can be perceived as biased nearest distance association with bias value $\rho = (\frac{P_m}{P_p}\frac{1}{\eta})^{\frac{1}{\alpha}}$, where $\eta = B\varrho$. We investigate the optimal value of η for the average data rate in Section VI, which determines the optimal ρ.

As X_m and X_p follow Rayleigh distributions with mean $(2\sqrt{\lambda_m})^{-1}$ and $(2\sqrt{\lambda_p})^{-1}$, respectively [34], the probability that a typical user at the origin is associated with the pico tier is

$$A_p = \mathbb{P}(X_m \geq \rho X_p) = \frac{\lambda_p}{\lambda_p + \lambda_m \rho^2}, \tag{2}$$

and the probability that this user is associated with the macro tier is $A_m = 1 - A_p$. These tier association probabilities are also valid for any randomly selected user. Thus, the total set of users in the network, Φ_u can be divided into two disjoint subsets: Φ_u^m and Φ_u^p, the set of macro- and pico-users, respectively. A_m and A_p can be interpreted as the average fraction of users belonging to Φ_u^m and Φ_u^p, respectively. As we are interested in the number of users in a typical cell, rather than the actual locations of the users, Φ_u^m and Φ_u^p can be equivalently modeled as independent PPPs with density $A_m\lambda_u$ and $A_p\lambda_u$, respectively. Since each macro-user is always associated with the nearest macro BS and each pico-userr with the nearest pico BS, the network can be viewed as a superposition of two independent Voronoi tessellations of the macro and pico tiers. Let the number of users in a randomly chosen macro and pico cell be denoted by U_m and U_p, respectively. Their approximate[1] probability mass

[1] The PDF of the normalized Poisson-Voronoi cell area is approximated as Gamma(3.5, 3.5) [35] while deriving the PMFs.

function (PMFs) are given by [24, Lemma 2]

$$\mathbb{P}(U_l = n) = \frac{3.5^{3.5}\Gamma(3.5+n)(A_l\lambda_u/\lambda_l)^n}{\Gamma(3.5)n!(A_l\lambda_u/\lambda_l+3.5)^{n+3.5}}, n \geq 0, \quad \forall l \in \{m, p\}. \tag{3}$$

A BS without any user associated does not transmit at all and is inactive. The PMFs of the number of users in a randomly chosen active cell of the macro and pico tiers are given by

$$\mathbb{P}(N_l = n) = \frac{\mathbb{P}(U_l = n)\mathbf{1}(n \geq 1)}{p_l}, \ \forall l \in \{m, p\}, \tag{4}$$

where p_m and p_p are the probabilities that a typical BS of the macro and pico tiers, respectively, is active, and are given by

$$p_l = 1 - \mathbb{P}(U_l = 0) = 1 - \left(1 + 3.5^{-1}\frac{A_l\lambda_u}{\lambda_l}\right)^{-3.5}, \quad \forall l \in \{m, p\}. \tag{5}$$

Let the sets of active macro and active pico BSs be denoted by Ψ_m and Ψ_p, respectively. Ψ_m and Ψ_p are thinned versions of the original PPPs Φ_m and Φ_p, respectively, and hence are independent PPPs with densities $p_m\lambda_m$ and $p_p\lambda_p$, respectively.

By using the PMFs in (4), the PMFs of the number of users simultaneously served by a typical active BS of macro and pico tiers in a given time slot for $L^l_{\max} > 1$ can be obtained as

$$\mathbb{P}(M_l = n) = \begin{cases} \mathbb{P}(N_l = n), & 1 \leq n < L^l_{\max} \\ 1 - \sum_{k=1}^{L^l_{\max}-1} \mathbb{P}(N_l = k), & n = L^l_{\max}, \end{cases} \quad \forall l \in \{m, p\}. \tag{6}$$

For the special case of $L^l_{\max} = 1, \mathbb{P}(M_l = 1) = 1, \forall l \in \{m, p\}$.

B. Interference Nulling

We assume K_m to be typically much larger than K_p. By using the interference nulling strategy, the additional spatial DoF of macro BSs can be utilized to suppress the strong macro interference topico users. Thus, we propose that each served pico-user requests its nearest active macro BS to perform interference nulling. However, as nulling costs macro BSs their available DoF for their own users, we assume that each macro BS can handle at most $K_m - T_{\min}$ requests only. This limit ensures that each macro BS has at least $T_{\min} \geq L^M_{\max}$ antennas dedicated for serving its own users. Hence, if Q_m requests are received by a typical active macro BS, it will perform interference nulling to $o = \min(Q_m, K_m - T_{\min})$pico users. For $Q_m > (K_m - T_{\min})$, the BS will randomly choose $K_m - T_{\min}$pico users.

The number of interference-nulling requests Q_m received by a typical active macro BS is equal to the number of servedpico users within a typical Voronoi cell Υ of the tessellation formed by Ψ_m. Although the number ofpico users served by a typical active pico BS cannot exceed $L^p_{\max}$, Q_m is unbounded because the number of active pico BSs within Υ is Poisson distributed with mean $p_p\lambda_p/(p_m\lambda_m)$. To derive the PMF of Q_m, we first derive $\mathbb{E}[M_p] = A_p\vartheta_p\lambda_u/(p_p\lambda_p)$, where

$$\vartheta_p = \frac{L^p_{\max}p_p\lambda_p}{A_p\lambda_u} - \frac{3.5^{3.5}}{\Gamma(3.5)}\sum_{k=1}^{L^p_{\max}-1} \times \left[\frac{\Gamma(3.5+n)}{n!}\frac{(A_p\lambda_u/\lambda_p)^{n-1}(L^p_{\max}-k)}{(A_p\lambda_u/\lambda_p+3.5)^{n+3.5}}\right]. \tag{7}$$

Note that for $L^p_{\max} = 1$, $\vartheta_p = \frac{p_p\lambda_p}{A_p\lambda_u}$. Next, let us denote the set ofpico users requesting interference nulling by Ψ^p_u. Because we are only interested in the number of such users in a typical Voronoi cell Υ, and not their actual locations, and we know that $\mathbb{E}[Q_m] = A_p\vartheta_p\lambda_u/(p_m\lambda_m)$, Ψ^p_u can be assumed to be a PPP with density $A_p\vartheta_p\lambda_u$. The PMF of Q_m can then be obtained as

$$\mathbb{P}(Q_m = n) = \frac{3.5^{3.5}\Gamma(3.5+n)\left(\frac{A_p\vartheta_p\lambda_u}{p_m\lambda_m}\right)^n}{\Gamma(3.5)n!\left(\frac{A_p\vartheta_p\lambda_u}{p_m\lambda_m}+3.5\right)^{n+3.5}}, n \geq 0. \tag{8}$$

Due to the limited resources as discussed earlier, not all interference-nulling requests received by an active macro BS are satisfied. Let χ denotes the set ofpico users whose interference-nulling requests to their corresponding nearest active macro BSs are satisfied. In the following lemma, we derive the probability that a randomly chosen pico-user in service belongs to χ.

Lemma 1: The probability φ that the interference-nulling request made by a randomly chosen pico-user to its nearest active macro BS is fulfilled is given by

$$\varphi = \frac{(K_m - T_{min})p_m\lambda_m}{A_p\vartheta_p\lambda_u}\left(1 - \left(1 + 3.5^{-1}\frac{A_p\vartheta_p\lambda_u}{p_m\lambda_m}\right)^{-3.5}\right) - \frac{3.5^{3.5}}{\Gamma(3.5)}\sum_{n=1}^{K_m-T_{min}}\frac{\Gamma(3.5+n)\left(\frac{A_p\vartheta_p\lambda_u}{p_m\lambda_m}\right)^{n-1}(K_m-T_{min}-n)}{n!\left(\frac{A_p\vartheta_p\lambda_u}{p_m\lambda_m}+3.5\right)^{n+3.5}}. \tag{9}$$

Proof: Let Q'_m denotes the number of other requests received by the macro BS, which received nulling request from a randomly chosen pico-user. Then, conditioned on Q'_m, $\varphi = 1$ if $Q'_m + 1 \leq K_m - T_{\min}$; otherwise, $\varphi = (K_m - T_{\min})/(Q'_m + 1)$. Thus, φ can be expressed as

$$\begin{aligned}\varphi &= \sum_{n=0}^{K_m-T_{\min}-1}\mathbb{P}(Q'_m = n) + \sum_{n=K_m-T_{\min}}^{\infty}\frac{K_m - T_{\min}}{n+1}\mathbb{P}(Q'_m = n) \\ &= \sum_{n=1}^{\infty}\frac{K_m - T_{\min}}{n}\mathbb{P}(Q'_m = n-1) \\ &\quad - \sum_{n=1}^{K_m-T_{\min}}\left(\frac{K_m - T_{\min}}{n} - 1\right)\mathbb{P}(Q'_m = n-1). \end{aligned} \tag{10}$$

By using the fact that the conditional probability density function (PDF) $f'_Y(y)$ of the area of a Voronoi cell given that a randomly chosen user belongs to it is equal to $cyf_Y(y)$, where $f_Y(y)$ is the unconditional PDF and c is a constant such that $\int_o^\infty f'_Y(y)dy = 1$ [22], the PMF of Q'_m can be derived as $\mathbb{P}(Q'_m = n) = (n+1)\mathbb{P}(Q_m = n+1)/\mathbb{E}[Q_m]$, $n \geq 0$. Theorem 1 is then obtained by substituting the PMF of Q'_m in (10), and then using $\sum_{n=1}^{\infty} \mathbb{P}(Q_m = n) = 1 - \mathbb{P}(Q_m = 0)$. ■

C. Channel Model and Precoding Matrices

Assuming standard power law path-loss with exponent α, linear precoding and frequency-flat fading, the received signal z_m at a typical user u located at the origin if $u \in \Phi_u^m$ is given by

$$\begin{aligned} z_m &= \sqrt{P_m} D_m^{-\frac{\alpha}{2}} \mathbf{h}_{b_m,1}^* \mathbf{W}_{b_m} \mathbf{s}_{b_m} \\ &\quad + \sum_{q \in \{m,p\}} \sqrt{P_q} \sum_{x_q \in \Psi_q \setminus b_m} ||x_q||^{-\frac{\alpha}{2}} \mathbf{g}_{x_q,1}^* \mathbf{W}_{x_q} \mathbf{s}_{x_q} + n_m, \end{aligned} \tag{11}$$

where b_m is the serving macro BS at a distance D_m, which is serving M'_m other users simultaneously; $\mathbf{h}_{b_m,1} \sim \mathcal{CN}(\mathbf{0}_{K_m \times 1}, \mathbf{I}_{K_m})$ and $\mathbf{g}_{x_q,1} \sim \mathcal{CN}(\mathbf{0}_{K_q \times 1}, \mathbf{I}_{K_q})$ are the desired and interference complex Gaussian channel vectors from the tagged BS b_m and the interfering BS at x_q, respectively, with independent and identically distributed (i.i.d.) unit variance components; $n_m \sim \mathcal{CN}(0, \sigma^2)$ is complex Gaussian noise with variance σ^2; $\mathbf{s}_{b_m} = [s_{b_m,i}]_{1 \leq i \leq M'_m + 1} \in \mathbb{C}^{(M'_m+1) \times 1}$ is the complex-valued signal vector transmitted from b_m to its $M'_m + 1$ served users with the symbol $s_{b_m,1}$ intended for u and $\mathbf{W}_{b_m} = [\mathbf{w}_{b_m,i}]_{1 \leq i \leq (M'_m+1)} \in \mathbb{C}^{K_m \times (M'_m+1)}$ is the corresponding ZF precoding matrix.

Let the channel vectors from the tagged BS b_m to its M'_m users other than u be represented by $[\, \mathbf{h}_{b_m,i} \,]_{2 \leq i \leq M'_m + 1}$, and the interference channel vector from the tagged BS to $O = \min(Q_m, K_m - T_{\min})$pico users chosen for interference nulling by $\mathbf{F} = [\, \mathbf{f}_i \,]_{1 \leq i \leq O} \in \mathbb{C}^{K_m \times O}$. Under the perfect CSI assumption, the ZF precoding matrix $\mathbf{W}_{b_m} = [\mathbf{w}_{b_m,i}]_{1 \leq i \leq (M'_m+1)}$ is designed such that $|\mathbf{h}_{b_m,j}^* \mathbf{w}_{b_m,j}|^2$ is maximized for each $j = 1, 2, \ldots, M'_m + 1$, while satisfying the orthogonality conditions $\mathbf{h}_{b_m,j}^* \mathbf{w}_{b_m,i} = 0$ for $\forall i \neq j$ and $\mathbf{f}_i^* \mathbf{w}_{b_m,j} = 0, \forall i = 1, 2, \ldots, O, \forall j = 1, 2, \ldots, M'_m + 1$. It can be achieved by choosing $\mathbf{w}_{b_m,i}$ in the direction of the projection of $\mathbf{h}_{b_m,i}$ on $\mathrm{Null}\big([\, \mathbf{h}_{b_m,j} \,]_{1 \leq j \leq (M'_m+1), j \neq i}, [\, \mathbf{f}_i \,]_{1 \leq i \leq O}\big)$. The nullspace is $K_m - M'_m - O$ dimensional and thus, the desired channel power gain $\beta_{b_m} = |\mathbf{h}_{b_m,1}^* \mathbf{w}_{b_m,1}|^2 \sim \mathrm{Gamma}(\Delta_m, 1)$, where $\Delta_m = K_m - M'_m - O$ [36]. Given that an interfering macro BS at x_m is serving M_m users simultaneously, $\mathbf{W}_{x_m} = [\mathbf{w}_{x_m,i}]_{1 \leq i \leq M_m} \in \mathbb{C}^{K_m \times M_m}$, which is designed independent of $\mathbf{g}_{x_m,1}$. Assuming that the precoding matrix has linearly independent unit norm columns, $\mathbf{g}_{x_m,1}^* \mathbf{w}_{x_m,1}, \mathbf{g}_{x_m,1}^* \mathbf{w}_{x_m,2}, \ldots, \mathbf{g}_{x_m,1}^* \mathbf{w}_{x_m,M_m}$ are i.i.d. complex Gaussian random variables (RVs), and their squared norms are i.i.d. exponential RVs. Thus, the interference channel power gain $\zeta_{x_m} = ||\mathbf{g}_{x_m,1}^* \mathbf{W}_{x_m}||^2 \sim \mathrm{Gamma}(M_m, 1)$, as it is a sum of M_m i.i.d. exponential RVs [18].

A feasible choice of the precoding matrix $\mathbf{W}_{b_m} = [\mathbf{w}_{b_m,i}]_{1 \leq i \leq (M'_m+1)}$ is the pseudo inverse[2] of $\tilde{\mathbf{H}}_{b_m}^*$, i.e., $\mathbf{W}_{b_m} = \tilde{\mathbf{H}}_{b_m} (\tilde{\mathbf{H}}_{b_m}^* \tilde{\mathbf{H}}_{b_m})^{-1}$ with normalized columns, where $\tilde{\mathbf{H}}_{b_m} = [\tilde{\mathbf{h}}_{b_m,i}]_{1 \leq i \leq (M'_m+1)} \in \mathbb{C}^{K_m \times (M'_m+1)}$, $\tilde{\mathbf{h}}_{b_m,i} = (\mathbf{I}_{K_m} - \mathbf{F}(\mathbf{F}^* \mathbf{F})^{-1} \mathbf{F}^*) \mathbf{h}_{b_m,i}$ being the projection of $\mathbf{h}_{b_m,i}$ on the nullspace of $\mathbf{F} = [\, \mathbf{f}_i \,]_{1 \leq i \leq O}$ [31], [36].

[2]Pseudo inversion of the channel matrix is an easy choice of ZF precoding [7].

Similarly, the received signal z_p at u when $u \in \Phi_u^p$ is

$$\begin{aligned} z_p &= \sqrt{P_p} D_p^{-\frac{\alpha}{2}} \mathbf{h}_{b_p,1}^* \mathbf{W}_{b_p} \mathbf{s}_{b_p} + \xi \\ &\quad + \sum_{q \in \{m,p\}} \sqrt{P_q} \sum_{x_q \in \Psi_q \setminus \{v_m, b_p\}} ||x_q||^{-\frac{\alpha}{2}} \mathbf{g}_{x_q,1}^* \mathbf{W}_{x_q} \mathbf{s}_{x_q} + n_p, \end{aligned} \tag{12}$$

where

$$\xi = \begin{cases} 0, & \text{if } u \in \chi \\ \sqrt{P_m} V_m^{-\frac{\alpha}{2}} \mathbf{g}_{v_m,1}^* \mathbf{W}_{v_m} \mathbf{s}_{v_m}, & \text{if } u \notin \chi; \end{cases} \tag{13}$$

b_p is the serving pico BS at a distance D_p, which is serving M'_p other users simultaneously; $n_p \sim \mathcal{CN}(0, \sigma^2)$ is complex Gaussian noise, v_m is the nearest active macro BS to u at a distance V_m, which receives an interference-nulling request from u. The ZF precoding matrix $\mathbf{W}_{b_p} = [\mathbf{w}_{b_p,i}]_{1 \leq i \leq (M'_p+1)}$ is given by $\mathbf{H}_{b_p} (\mathbf{H}_{b_p}^* \mathbf{H}_{b_p})^{-1}$ with normalized columns, where $\mathbf{H}_{b_p} = [\mathbf{h}_{b_p,i}]_{1 \leq i \leq (M'_p+1)} \in \mathbb{C}^{K_p \times (M'_p+1)}$ is the channel matrix from the tagged BS b_p to its $M'_p + 1$ servedpico users. The desired channel power gain $\beta_{b_p} = ||\mathbf{h}_{b_p,1}^* \mathbf{W}_{b_p}||^2 = |\mathbf{h}_{b_p,1}^* \mathbf{w}_{b_p,1}|^2 \sim \mathrm{Gamma}(\Delta_p, 1)$, where $\Delta_p = K_m - M'_p$, and the interference channel power gain $\zeta_{x_p} = ||\mathbf{g}_{x_p,1}^* \mathbf{W}_{x_p}||^2 \sim \mathrm{Gamma}(M_p, 1)$ given that the interfering pico BS at x_p is serving M_p users simultaneously.

D. Distance to the Serving BS and the BS Receiving Interference Nulling Request

The distance D_l to the serving BS from a typical user $u \in \Phi_u^l$ is a RV, and the corresponding PDFs for each $l \in \{m, p\}$ are derived in the following lemma.

Lemma 2: The PDF $f_{D_m}(r)$ of the distance D_m between the serving macro BS and a typical user u when $u \in \Phi_u^m$ is given by

$$f_{D_m}(r) = \frac{2\pi \lambda_m}{A_m} r \exp(-\pi(\lambda_m + \lambda_p/\rho^2) r^2), \tag{14}$$

and the PDF $f_{D_p}(r)$ of the distance D_p between the serving pico BS and a typical user u when $u \in \Phi_u^p$ is given by

$$f_{D_p}(r) = \frac{2\pi \lambda_p}{A_p} r \exp(-\pi(\lambda_m \rho^2 + \lambda_p) r^2). \tag{15}$$

Proof: Given that $u \in \Phi_u^m$, D_m is the distance to the nearest macro BS from u. The cumulative distribution function (CDF) $F_{D_m}(r) = \mathbb{P}(D_m \leq r)$ is thus given by

$$\begin{aligned} F_{D_m}(r) &= \mathbb{P}(X_m \leq r | u \in \Phi_u^m) = \frac{\mathbb{P}(X_m \leq r, u \in \Phi_m^m)}{\mathbb{P}(u \in \Phi_u^m)} \\ &= \frac{1}{A_m} \int_0^r \mathbb{P}\Big(X_p > \frac{y}{\rho}\Big) f_{X_m}(y) dy. \end{aligned} \tag{16}$$

The PDF $f_{D_m}(r)$ in (14) is obtained by differentiating (16) with respect to r and then applying the probability distributions of Rayleigh RVs X_m and X_p. The PDF $f_{D_p}(r)$ is similarly derived. ■

Another quantity of interest is the distance V_m between a typical pico-user in service and its nearest active macro BS to which it requests interference nulling.

Lemma 3: The conditional PDF of the distance V_m between a typical user $u \in \Phi_u^p$ and the macro BS to which it request interference nulling, given that its distance to the serving pico BS is $D_p = r$, is given by

$$f_{V_m|D_p}(r_1|r) = 2\pi p_m \lambda_m r_1 \exp\left(-\pi p_m \lambda_m (r_1^2 - \rho^2 r^2)\right), \quad r_1 > \rho r. \tag{17}$$

Proof: Given that $u \in \Phi_u^p$, V_m is the distance to the nearest active macro BS. The conditional complementary cumulative distribution function (CCDF) of V_m is thus given by

$$\begin{aligned}\bar{F}_{V_m|D_p}(r_1|r) &= \mathbb{P}(X'_m \geq r_1 | u \in \Phi_u^p, D_p = r)\\ &= \mathbb{P}(X'_m \geq r_1 | X_m > \rho r),\end{aligned} \tag{18}$$

where $X'_m = \min_{x_m \in \Psi_m} \|x_m\|$ is the distance from the origin to the nearest active macro BS. The condition $X_m > \rho r$ implies that no points of Φ_m are within a circle of radius ρr. Thus, no points of Ψ_m as well are within ρr because Ψ_m is the thinned version of Φ_m. Thus, given that no active macro BS is closer than ρr, the probability of no active macro BS closer than r_1 is equal to the probability that no points of Ψ_m are within an annulus centered at the origin with inner radius ρr and outer radius r_1. The conditional CCDF $\bar{F}_{V_m|Dp}(r_1|r)$ is thus given by

$$\bar{F}_{V_m|D_p}(r_1|r) = \exp\left(-\pi p_m \lambda_m (r_1^2 - \rho^2 r^2)\right). \tag{19}$$

The conditional PDF of V_m in (17) is obtained by differentiating (19) with respect to r_1. ■

III. SIR Coverage Analysis

We consider interference-limited scenario, and thus derive the SIR coverage probability in this section. The SIR coverage, i.e., the probability that the SIR of a typical user is greater than a given threshold γ is defined as $\mathrm{P}(\gamma) = \mathbb{P}(\mathrm{SIR} > \gamma)$, where $\mathrm{SIR} = \sum_{l \in \{m,p\}} \mathbf{1}(u \in \Phi_u^l)\,\mathrm{SIR}_l$. From (11) and (12) and the discussion that follows, the SIR of a typical user u at the origin when it belongs to Φ_u^l can be expressed as

$$\mathrm{SIR}_l = \frac{P_l \beta_{b_l} D_l^{-\alpha}}{I_{b_l,m} + I_{b_l,p}}, \quad \forall l \in \{m, p\}, \tag{20}$$

where $I_{b_l,m}$ and $I_{b_l,p}$ are the interference powers from the macro and pico tiers, respectively when $u \in \Phi_u^l$, $l \in \{m, p\}$, and are given by

$$\begin{aligned}
I_{b_p,p} &= P_p \sum_{x_p \in \Psi_p \backslash b_p} \zeta_{x_p} ||x_p||^{-\alpha}\\
I_{b_p,m} &= \begin{cases} P_m \displaystyle\sum_{x_m \in \Psi_m \backslash \upsilon_m} \zeta_{x_m} ||x_m||^{-\alpha} & \text{if } u \in \chi\\ P_m \displaystyle\sum_{x_m \in \Psi_m} \zeta_{x_m} ||x_m||^{-\alpha} & \text{if } u \notin \chi, \end{cases}\\
I_{b_m,p} &= P_p \sum_{x_p \in \Psi_p} \zeta_{x_p} ||x_p||^{-\alpha}\\
I_{b_m,m} &= P_m \sum_{x_m \in \Psi_m \backslash b_m} \zeta_{x_m} ||x_m||^{-\alpha}.
\end{aligned} \tag{21}$$

By using the law of total probability, the SIR coverage probability of a typical user u is

$$\mathrm{P}(\gamma) = \mathrm{P}_m(\gamma) A_m + \mathrm{P}_p(\gamma) A_p, \tag{22}$$

where $A_l = \mathbb{P}(u \in \Phi_u^l)$, $l \in \{m, p\}$ is the tier association probability, and $\mathrm{P}_m(\gamma) = \mathbb{P}(\mathrm{SIR}_m > \gamma | u \in \Phi_u^m)$, and $\mathrm{P}_p(\gamma) = \mathbb{P}(\mathrm{SIR}_p > \gamma | u \in \Phi_u^p)$ are the conditional coverage probabilities of the user u when associated with the macro and pico tiers, respectively. To evaluate (22), we first derive the Laplace transform (LT) of the total interference power received by u.

Lemma 4: The LT $\mathcal{L}_{I_{b_p}}(s)$ of the total interference power $I_{b_p} = I_{b_p,m} + I_{b_p,p}$ received by u when $u \in \Phi_u^p$ conditional on $D_p = r$ and $V_m = r_1$ is given by

$$\mathcal{L}_{I_{b_p}}(s) = \left(\varphi \mathcal{L}^1_{I_{b_p,m}}(s) + (1-\varphi)\mathcal{L}^2_{I_{b_p,m}}(s)\right) \mathcal{L}_{I_{b_p,p}}(s), \tag{23}$$

where $\mathcal{L}_{I_{b_p,p}}(s)$ is the LT of $I_{b_p,p}$; $\mathcal{L}^1_{I_{b_p,m}}(s) = \mathcal{L}_{I_{b_p,m}}(s|u \in \chi)$, and $\mathcal{L}^2_{I_{b_p,m}}(s) = \mathcal{L}_{I_{b_p,m}}(s|u \notin \chi)$ are the LTs of $I_{b_p,m}$ conditional on $u \in \chi$ and $u \notin \chi$, respectively. These LTs are given by (24)-(26)*, as shown at the top of the next page, where ${}_2F_1(a, b, c, z)$ is the Gauss Hypergeometric function [37].*

Proof: The LT $\mathcal{L}_{I_{b_p,q}}(s) = \mathbb{E}[e^{-sI_{b_p,q}}]$, $\forall q \in \{m, p\}$ can be derived as

$$\mathcal{L}_{I_{b_p,q}}(s) = \mathbb{E}_{\hat{\Psi}_q} \prod_{x_l \in \hat{\Psi}_q} \mathbb{E}_{\zeta_{x_q}}\left[\exp(-sP_q \zeta_{x_q} ||x_q||^{-\alpha})\right], \tag{28}$$

where $\hat{\Psi}_p = \Psi_p \backslash b_p$, and $\hat{\Psi}_m = \Psi_m \backslash \upsilon_m$ if $u \in \chi$, else $\hat{\Psi}_m = \Psi_m$. By performing the expectation over the distribution of $\zeta_{x_q} \sim \mathrm{Gamma}(M_q, 1)$ conditioned on M_q, and then applying the probability generating functional of PPP with density $p_q \lambda_q$ [34], and finally taking the expectation over the PMF of M_q, we have

$$\begin{aligned}\mathcal{L}_{I_{b_p,q}}(s) = \exp\Bigg\{ &-\pi p_q \lambda_q \varpi_{p,q}^2 \Bigg(\sum_{i=1}^{L^q_{\max}} \mathbb{P}(M_q = i)\\ &\times {}_2F_1\Big[i, -\frac{2}{\alpha}, \frac{\alpha-2}{\alpha}, -\frac{P_q}{\varpi_{p,q}^\alpha} s\Big] - 1\Bigg)\Bigg\}, \end{aligned} \tag{29}$$

where $\varpi_{p,q}$ is the lower bound on the distance to the closest interferer from u in the tier $q \in \{m, p\}$. Thus, $\varpi_{p,p} = r$, and $\varpi_{p,m} = r_1$ if $u \in \chi$; otherwise, $\varpi_{p,m} = \rho r$. ■

Similarly, the LT of $I_{b_m} = I_{b_m,m} + I_{b_m,p}$ conditional on $D_m = r$ can be derived as $\mathcal{L}_{I_{b_m}}(s) = \mathcal{L}_{I_{b_m,m}}(s)\mathcal{L}_{I_{b_m,p}}(s)$, where $\mathcal{L}_{I_{b_m,q}}(s)$, $q \in \{m, p\}$ is given by (27) shown at the top of the next page, with $\varpi_{m,m} = r$ and $\varpi_{m,p} = r/\rho$.

Having derived the LTs, we now evaluate $\mathrm{P}_l(\gamma) = \mathbb{P}\left(P_l \beta_{b_l} D_l^{-\alpha} > \gamma I_{b_l} | u \in \Phi_u^l\right)$, $\forall l \in \{m, p\}$. Conditional on $D_l = r$, $V_m = r_1$ and $\Delta_l = n$, we have

$$\mathrm{P}_l(\gamma | r, r_1, \Delta_l = n) = \sum_{l=0}^{n-1} \frac{(-s)^l}{l!} \frac{\mathrm{d}^l}{\mathrm{d}s^l}\left(\mathcal{L}_{I_{b_l}}(s)\right)\Big|_{s=\frac{\gamma r^\alpha}{P_l}}, \tag{30}$$

which follows from the distribution $\mathrm{Gamma}(n, 1)$ of β_{b_l} for a given $\Delta_l = n$, and the differentiation property of LT. Since the LTs in (24)-(27) are composite functions, (30) requires evaluating lth derivatives of composite functions. These derivatives

$$\mathcal{L}^1_{I_{bp,m}}(s) = \exp\left\{-\pi p_m \lambda_m r_1^2 \left(\sum_{i=1}^{L^m_{\max}} \mathbb{P}(M_m = i)\, {}_2F_1\left[i, -\frac{2}{\alpha}, \frac{\alpha-2}{\alpha}, -\frac{P_m s}{r_1^\alpha}\right] - 1\right)\right\} \tag{24}$$

$$\mathcal{L}^2_{I_{bp,m}}(s) = \exp\left\{-\pi p_m \lambda_m \rho^2 r^2 \left(\sum_{i=1}^{L^m_{\max}} \mathbb{P}(M_m = i)\, {}_2F_1\left[i, -\frac{2}{\alpha}, \frac{\alpha-2}{\alpha}, -\frac{P_m s}{\rho^\alpha r^\alpha}\right] - 1\right)\right\} \tag{25}$$

$$\mathcal{L}_{I_{bp,p}}(s) = \exp\left\{-\pi p_p \lambda_p r^2 \left(\sum_{i=1}^{L^p_{\max}} \mathbb{P}(M_p = i)\, {}_2F_1\left[i, -\frac{2}{\alpha}, \frac{\alpha-2}{\alpha}, -\frac{P_p s}{r^\alpha}\right] - 1\right)\right\} \tag{26}$$

$$\mathcal{L}_{I_{bm,q}}(s) = \exp\left\{-\pi p_q \lambda_q \varpi_{m,q}^2 \left(\sum_{i=1}^{L^q_{\max}} \mathbb{P}(M_q = i)\, {}_2F_1\left[i, -\frac{2}{\alpha}, \frac{\alpha-2}{\alpha}, -\frac{P_q}{\varpi_{m,q}^\alpha} s\right] - 1\right)\right\} \tag{27}$$

are computed by using Faà di Bruno's formula expressed in terms of integer partition, which is introduced in the following section.

A. Integer Partition and Faà di Bruno's Formula

Integer partition is a partition of a positive integer n as a sum of positive integers. The set of all possible partitions of n is represented by $\mathbf{\Omega}_n$ with the number of partitions denoted by $\mathcal{P}(n)$. The integer 4, for example, can be partitioned as $\mathbf{\Omega}_4 = \{\{4\}, \{3, 1\}, \{2, 2\}, \{2, 1, 1\}, \{1, 1, 1\}\}$. Thus, $\mathcal{P}(4) = 5$. Let ω_i^n denotes the number of elements in the ith partition p_i^n of n. Also, let μ_{ij}^n denotes the number of positive integer $j \in \{1, 2, \ldots, n\}$ in that partition, and a_{ik}^n denotes the kth element ($k \in \{1, 2, \ldots, \omega_i^n\}$). *Example*: for the second partition of integer 4 in $\mathbf{\Omega}_4$, i.e., $p_2^4 = \{3, 1\}$, we have $\omega_2^4 = 2$, $\mu_{21}^4 = 1$, $\mu_{22}^4 = 0$, $\mu_{23}^4 = 1$, $\mu_{24}^4 = 0$, $a_{21}^4 = 3$, $a_{22}^4 = 1$. For any partition p_i^n, we have the properties $\sum_{j=1}^n j\mu_{ij}^n = n$ and $\sum_{j=1}^n \mu_{ij}^n = \omega_i^n$.

Faà di Bruno's formula for the lth derivative of the composite function $y(t(s))$ in terms of integer partition can be expressed as

$$y_s^{(l)}(t(s)) = \sum_{o=1}^{\mathcal{P}(l)} c_o^l \, y_{t(s)}^{(\omega_o^l)}(t(s)) \prod_{q=1}^{l} \left(t_s^{(q)}(s)\right)^{\mu_{oq}^l}, \tag{31}$$

where

$$c_o^l = \frac{l!}{\prod_{k=1}^{\omega_o^l} a_{ok}^l! \prod_{q=1}^{l} \mu_{oq}^l!},$$

and $y_{t(s)}^{(k)}(t(s))$ is the kth derivative of the function $y(t(s))$ with respect to $t(s)$. Note that the integer partition version has much lesser number of summations as compared to the set partition version used in [21]. The complexity of the numerical computation is thus significantly reduced.

Theorem 1: The SIR coverage probability of a typical pico-user u is given by

$$\mathrm{P}_p(\gamma) = \varphi \mathrm{T}_1(\gamma) + (1 - \varphi)\mathrm{T}_2(\gamma), \tag{32}$$

where $\mathrm{T}_1(\gamma) = \mathbb{P}(SIR_p > \gamma | u \in \Phi_u^p, u \in \chi)$ *and* $\mathrm{T}_2(\gamma) = \mathbb{P}(SIR_p > \gamma | u \in \Phi_u^p, u \notin \chi)$ *are the conditional coverage probabilities of a typical pico-user u when $u \notin \chi$ and $u \in \chi$, respectively. These conditional probabilities can be computed by using* (33) *and* (34)*, as shown at the top of the next page, where $\delta = P_m/P_p$, and the function $\Xi_q^l(\varsigma, \kappa, \varepsilon)$ is defined as*

$$\Xi_q^l(\varsigma, \kappa, \varepsilon) = \sum_{i=1}^{L^l_{max}} \left(\frac{(i)_q (-\frac{2}{\alpha})_q}{(\frac{\alpha-2}{\alpha})_q} \mathbb{P}(M_l = i) \right.$$
$$\left. \times {}_2F_1\left[i + q, -\frac{2}{\alpha} + q, \frac{\alpha-2}{\alpha} + q, -\varsigma\kappa^\alpha\varepsilon\right]\right), \tag{35}$$

where $(a)_q$ is a Pochhammer symbol.

Proof: The proof is given in Appendix A. ■

Remark 1: The number of other users served by the BS which is serving the typical user $u \in \Phi_u^l$ is given by $M_l' = \min(U_l', L^l_{max} - 1)$, where U_l' is the number of other users in the Voronoi cell to which the user u belongs. The PMF of U_l' can be derived as $\mathbb{P}(U_l' = n) = (n+1)\mathbb{P}(U_l = n+1)/\mathbb{E}[U_l]$. The PMF of M_l' for $L^l_{max} > 1$ is thus given by

$$\mathbb{P}(M_l' = n) = \begin{cases} \mathbb{P}(U_l' = n), & 0 \le n < L^l_{max} - 1 \\ 1 - \sum_{k=1}^{L^l_{max}-2} \mathbb{P}(U_l' = k), & n = L^l_{max} - 1, \end{cases}$$
$$\forall l \in \{m, p\}. \tag{36}$$

For $L^l_{max} = 1$, $\mathbb{P}(M_l' = 0) = 1, \forall l \in \{m, p\}$.

Corollary 1: The coverage probability of a typical macro-user $\mathrm{P}_m(\gamma)$ is given by (37)*, as shown at the top of the next page, where the PMF of Δ_m conditional on $M_m' = k$ for $T_{min} < K_m$ is given by*

$$\mathbb{P}(\Delta_m = n | M_m' = k)$$
$$= \begin{cases} 1 - \sum_{\upsilon=0}^{K_m - T_{min} - 1} \mathbb{P}(Q_m = \upsilon), & n = T_{min} - k \\ \mathbb{P}(Q_m = K_m - k - n), & T_{min} - k + 1 \le n \le K_m - k. \end{cases} \tag{38}$$

For the special case of $T_{min} = K_m$ which implies no interference nulling, $\Delta_m = K_m - M_m'$, thus $\mathbb{P}(\Delta_m = K_m - k | M_m' = k) = 1$.

Proof: $\mathrm{P}_m(\gamma)$ is derived in the same way as $\mathrm{T}_2(\gamma)$. However, since $\Delta_m = K_m - M_m' - \min(Q_m, K_m - T_{\min})$ is a function of the two RVs M_m' and Q_m, deconditioning with respect to Δ_m is achieved in two steps, first averaging over the

$$
\begin{aligned}
\mathrm{T}_1(\gamma) = {} & 2p_m\lambda_m\frac{\lambda_p}{A_p}\int_{\theta=0}^{\frac{1}{\rho}}\Bigg[\sum_{k=0}^{L_{\max}^p-1}\mathbb{P}(M_p^{'}=k)\sum_{l=0}^{K_p-k-1}\frac{\gamma^l}{l!}\theta^{\alpha l+1}\sum_{o=1}^{\mathcal{P}(l)}c_o^l(-1)^{\omega_o^l}\Big(p_m\lambda_m\,\Xi_0^m\,(\delta,\theta,\gamma)+p_p\lambda_p\theta^2\,\Xi_0^p\,(1,1,\gamma) \\
& +(1-p_m)\lambda_m\rho^2\theta^2+(1-p_p)\lambda_p\theta^2\Big)^{-(\omega_o^l+2)}\Gamma(\omega_o^l+2)\prod_{q=1}^{l}\Big(p_m\lambda_m\delta^q\,\Xi_q^m\,(\delta,\theta,\gamma)+\frac{p_p\lambda_p}{\theta^{\alpha q-2}}\,\Xi_q^p\,(1,1,\gamma)\Big)^{\mu_{oq}^l}\Bigg]\mathrm{d}\theta \qquad (33)
\end{aligned}
$$

$$
\begin{aligned}
\mathrm{T}_2(\gamma) = {} & \frac{\lambda_p}{A_p}\sum_{k=0}^{L_{\max}^p-1}\mathbb{P}(M_p^{'}=k)\sum_{l=0}^{K_p-k-1}\frac{\gamma^l}{l!}\sum_{o=1}^{P(l)}c_o^l(-1)^{\omega_o^l}\Gamma(\omega_o^l+1)\prod_{q=1}^{l}\left(\frac{p_m\lambda_m\delta^q}{\rho^{\alpha q-2}}\,\Xi_q^m\left(\delta,\frac{1}{\rho},\gamma\right)+p_p\lambda_p\,\Xi_q^p\,(1,1,\gamma)\right)^{\mu_{oq}^l} \\
& \times\left(p_p\lambda_p\,\Xi_0^p\,(1,1,\gamma)+(1-p_m)\lambda_m\rho^2+(1-p_p)\lambda_p+p_m\lambda_m\rho^2\,\Xi_0^m\left(\delta,\frac{1}{\rho},\gamma\right)\right)^{-(\omega_o^l+1)} \qquad (34)
\end{aligned}
$$

$$
\begin{aligned}
\mathrm{P}_m(\gamma) = {} & \frac{\lambda_m}{A_m}\sum_{k=0}^{L_{\max}^m-1}\mathbb{P}(M_m'=k)\sum_{n=T_{\min}-k}^{K_m-k}\mathbb{P}(\Delta_m=n|M_m'=k)\sum_{l=0}^{n-1}\frac{\gamma^l}{l!}\sum_{o=1}^{\mathcal{P}(l)}c_o^l(-1)^{\omega_o^l}\Gamma(\omega_o^l+1)\Bigg((1-p_m)\lambda_m+(1-p_p)\frac{\lambda_p}{\rho^2} \\
& +p_m\lambda_m\,\Xi_0^m\,(1,1,\gamma)+\frac{p_p\lambda_p}{\rho^2}\,\Xi_0^p\left(\frac{1}{\delta},\rho,\gamma\right)\Bigg)^{-(\omega_o^l+1)}\prod_{q=1}^{l}\left(p_m\lambda_m\,\Xi_q^m\,(1,1,\gamma)+p_p\lambda_p\frac{\rho^{\alpha q-2}}{\delta^q}\,\Xi_q^p\left(\frac{1}{\delta},\rho,\gamma\right)\right)^{\mu_{oq}^l} \qquad (37)
\end{aligned}
$$

---

conditional PMF of Δ_m for the given M_m', and then averaging over the PMF of M_m'. ■

Remark 2: For the special case of $L_{max}^m = L_{max}^p = 1$,

$$
\begin{aligned}
\mathrm{P}_p(\gamma) &= \mathrm{P}_p(\gamma\,|M_p'=0) \\
&= \varphi\mathrm{T}_1(\gamma\,|M_p'=0)+(1-\varphi)\mathrm{T}_2(\gamma\,|M_p'=0), \qquad (39) \\
\mathrm{P}_m(\gamma) &= \mathrm{P}_m(\gamma\,|M_m'=0), \qquad (40)
\end{aligned}
$$

where for each $l \in \{m, p\}$,

$$
\begin{aligned}
\Xi_q^l(\varsigma,\kappa,\varepsilon) = \Xi_q(\varsigma,\kappa,\varepsilon) = {} & \frac{(1)_q(-\frac{2}{\alpha})_q}{(\frac{\alpha-2}{\alpha})_q} \\
& \times{}_2F_1\left(1+q,-\frac{2}{\alpha}+q,\frac{\alpha-2}{\alpha}+q,-\varsigma\kappa^\alpha\varepsilon\right). \qquad (41)
\end{aligned}
$$

IV. Rate Analysis

In this section, we analyze the achievable downlink rate of a typical user. We derive the CCDF of downlink rate, also defined as the rate coverage, and the average rate of a typical user.

Assuming adaptive transmission scheme such that the Shannon limit is achieved, and treating the interference as noise, the data rate of a typical user u is given by

$$
R = \sum_{l\in\{m,p\}} S_l W \log_2(1+\mathrm{SIR}_l)\mathbf{1}(u\in\Phi_u^l), \qquad (42)
$$

where S_l is the fraction of resources received by u when $u \in \Phi_u^l$. For each $l \in \{m, p\}$, given that U_l' is the number of other users in the cell to which the user u belongs, the total users in the tagged cell are $U_l'+1$. We assume one RB per time slot with total bandwidth W, and at most $L_{\max}^l$ users served simultaneously in each RB through spatial multiplexing. Thus, if the total number of users in the tagged cell is less than $L_{\max}^l$ (i.e., $U_l' + 1 < L_{\max}^l$), each user can utilize the entire bandwidth W without sharing; thus, $S_l = 1$. However, if $U_l'+1$ is no less than $L_{\max}^l$ (i.e., $U_l' + 1 \geq L_{\max}^l$), we assume that the time-frequency resources are shared equally among the total users; thus, $S_l = L_{\max}^l/(U_l' + 1)$. Hence, the fraction of resources received by $u \in \Phi_u^l$ can be expressed as

$$
S_l = \min\left(\frac{L_{\max}^l}{U_l'+1}, 1\right).
$$

Theorem 2: The CCDF of the downlink rate of a typical user u, $\mathcal{R}(\upsilon) = \mathbb{P}(R > \upsilon)$ can be expressed as $\mathcal{R}(\upsilon) = A_m\mathcal{R}_m(\upsilon) + A_p\mathcal{R}_p$, where $A_l = \mathbb{P}(u \in \Phi_u^l)$ and $\mathcal{R}_l(\upsilon) = \mathbb{P}(S_l W \log_2(1+SIR_l) > \upsilon)$ is the rate distribution of $u \in \Phi_u^l$. $\mathcal{R}_l(\upsilon)$ for each $l \in \{m, p\}$ is given by (43), as shown at the top of the next page, where $\mathrm{P}_l(\gamma\,|M_l' = k)$ is the conditional SIR coverage probability of $u \in \Phi_u^l$ for given $M_l' = k$.

Proof: From (42),

$$
\mathbb{P}(R>\upsilon) = \sum_{l\in\{m,p\}}\mathbb{P}(u\in\Phi_u^l)\underbrace{\mathbb{P}(S_l W\log_2(1+\mathrm{SIR}_l)>\upsilon)}_{\mathcal{R}_l(\upsilon)}
$$

where

$$
\begin{aligned}
\mathcal{R}_l(\upsilon) &= \mathbb{P}(W\log_2(1+\mathrm{SIR}_l)>\upsilon, U_l'\leq L_{\max}^l-2) \\
&\quad+\mathbb{P}\Big(\frac{L_{\max}^l}{U_l^{'}+1}W\log_2(1+\mathrm{SIR}_l)>\upsilon, U_l'\geq L_{\max}^l-1\Big) \\
&= \sum_{k=0}^{L_{\max}^l-2}\mathbb{P}(\mathrm{SIR}_l>2^{\upsilon/W}-1|U_l'=k)\mathbb{P}(U_l'=k) \\
&\quad+\sum_{k\geq L_{\max}^l-1}\mathbb{P}(\mathrm{SIR}_l>2^{\frac{\upsilon}{W}\frac{(k+1)}{L_{\max}^l}}-1|U_l'=k)\mathbb{P}(U_l'=k). \qquad (44)
\end{aligned}
$$

The conditional SIR coverage probabilities in (44) can be conditioned on the given value of M_l' by using $M_l' = \min(U_l', L_{\max}^l - 1)$. ■

$$\mathcal{R}_l(v) = \sum_{k=0}^{L_{\max}^l-2} \mathrm{P}_l\left(2^{v/W}-1 \,\middle|\, M_l'=k\right)\mathbb{P}(U_l'=k) + \sum_{k\geq L_{\max}^l-1} \mathrm{P}_l\left(2^{\frac{v}{W}\frac{(k+1)}{L_{\max}^l}}-1 \,\middle|\, M_l'=L_{\max}^l-1\right)\mathbb{P}(U_l'=k) \tag{43}$$

$$\bar{R}_l = \frac{W}{\ln 2}\int_0^\infty \frac{1}{1+y}\left[\sum_{k=0}^{L_{\max}^l-2} \mathrm{P}_l\left(y \,\middle|\, M_l'=k\right)\mathbb{P}(U_l'=k) + O_l\,\mathrm{P}_l\left(y \,\middle|\, M_l'=L_{\max}^l-1\right)\right]dy \tag{46}$$

$$O_l = \frac{L_{\max}^l\lambda_l}{A_l\lambda_u}\left(1-\left(1+3.5^{-1}A_l\lambda_u/\lambda_l\right)^{-3.5}\right) - \frac{3.5^{3.5}}{\Gamma(3.5)}\sum_{k=1}^{L_{\max}^l-1}\frac{\Gamma(3.5+k)\left(\frac{A_l\lambda_u}{\lambda_l}\right)^{k-1}L_{\max}^l}{k!(\frac{A_l\lambda_u}{\lambda_l}+3.5)^{3.5+k}} \tag{47}$$

For the special case of $L_{\max}^l = 1$, the rate distribution of $u \in \Phi_u^l$ further simplifies to

$$\mathcal{R}_l(v) = \sum_{k\geq 0}\mathrm{P}_l\left(2^{\frac{v}{W}(k+1)}-1\right)\mathbb{P}(U_l'=k). \tag{45}$$

After the rate coverage, we next derive the average data rate of any randomly chosen user.

Theorem 3: The average rate $\bar{R} = \mathbb{E}[R]$ of a typical user u is given by $\bar{R} = A_m\bar{R}_m + A_p\bar{R}_p$, where $\bar{R}_l = \mathbb{E}[S_l W\log_2(1+\mathrm{SIR}_l)]$ is the average rate of $u \in \Phi_u^l$, $l \in \{m, p\}$. $\bar{R}_l$ is given by (46), as shown at the top of this page, where O_l is computed according to (47), shown at the top of this page.

Proof: From (42),

$$\mathbb{E}[R] = \sum_{l\in\{m,p\}}\mathbb{P}(u\in\Phi_u^l)\underbrace{\mathbb{E}[S_l W\log_2(1+\mathrm{SIR}_l)]}_{\bar{R}_l},$$

where

$$\begin{aligned}\bar{R}_l &= W\sum_{k=0}^{L_{\max}^l-2}\mathbb{E}\left[\log_2(1+\mathrm{SIR}_l)|U_l'=k\right]\mathbb{P}(U_l'=k)\\ &\quad+W\sum_{k\geq L_{\max}^l-1}\frac{L_{\max}^l}{k+1}\mathbb{E}\left[\log_2(1+\mathrm{SIR}_l)|U_l'=k\right]\mathbb{P}(U_l'=k).\end{aligned} \tag{48}$$

The computation of $\mathbb{E}[\log_2(1+\mathrm{SIR}_l)]$ requires integrating $\log_2(1+\mathrm{SIR}_l)$ with respect to the PDF of SIR_l. However, the integral can be transformed into $1/(\ln 2)\int_0^\infty \mathrm{P}_l(y)(1+y)^{-1}\,dy$ by applying integration by parts, along with the fact that PDF is the negative differentiation of CCDF. Also, we have $M_l' = \min(U_l', L_{\max}^l-1)$. Equation (48) thus can be simplified to (46), where

$$\begin{aligned}O_l &= \sum_{k\geq L_{\max}^l-1}\frac{L_{\max}^l}{k+1}\mathbb{P}(U_l'=k)\\ &= \sum_{k=1}^{\infty}\frac{L_{\max}^l}{k}\mathbb{P}(U_l'=k-1) - \sum_{k=1}^{L_{\max}^l-1}\frac{L_{\max}^l}{k}\mathbb{P}(U_l'=k-1).\end{aligned}$$

Equation (47) is obtained by substituting $\mathbb{P}(U_l^{'}=k) = (k+1)\mathbb{P}(U_l=k+1)/\mathbb{E}[U_l]$, $k \geq 0$ and further simplifying by using $\sum_{k=1}^\infty \mathbb{P}(U_l=k) = 1-\mathbb{P}(U_l=0)$. ■

For the special case of $L_{\max}^l = 1$, the average data rate of $u \in \Phi_u^l$ simplifies to

$$\bar{R}_l = O_l\frac{W}{\ln 2}\int_0^\infty\frac{\mathrm{P}_l(y)}{1+y}\,dy. \tag{49}$$

V. Impact of Limited Feedback on Interference Nulling

The results so far have been derived based on the perfect CSI assumption. However, in practical systems, the CSI is never perfectly accurate. In frequency division duplex systems, the downlink CSI is fed back by the users to serving BSs. Due to the limited feedback, the BSs receive quantized CSI. In this section, we analyze the impact of the quantization error due to limited feedback on the performance of interference nulling. As the focus is on interference-nulling performance, we consider $L_{\max}^m = L_{\max}^p = 1$.

The feedback model is similar to the one used in [31] and [32]. The quantized channel direction information (CDI) is fed back by using a quantization codebook of 2^B unit norm vectors, where B is the number of feedback bits. The codebook is known at both the transmitter and the receiver. Each user feeds back the index of the codeword closest to its channel direction, measured by the inner product. For example, a typical user, when it belongs to the macro tier, uses the codebook $\mathcal{C}_m = \{\mathbf{c}_{m,j} : j = 1, 2, \ldots, 2^{B_m}\}$ of size 2^{B_m} to quantize the channel direction $\tilde{\mathbf{h}}_{b_m,1} = \frac{\mathbf{h}_{b_m,1}}{||\mathbf{h}_{b_m,1}||}$ from its serving maco BS b_m. The quantized channel direction is

$$\hat{\mathbf{h}}_{b_m,1} = \arg\max_{\mathbf{c}_{m,j}\in\mathcal{C}_m}\left|\tilde{\mathbf{h}}_{b_m,1}^*\mathbf{c}_{m,j}\right|.$$

Similarly, the typical user, when it belongs to the pico tier, uses the codebook $\mathcal{C}_p = \{\mathbf{c}_{p,j} : j = 1, 2, \ldots, 2^{B_p}\}$ of size 2^{B_p} to quantize the channel direction from its serving pico BS b_p, and the codebook $\mathcal{C}_m = \{\mathbf{c}_{m,j} : j = 1, 2, \ldots, 2^{B_m}\}$ to quantize the channel direction from its nearest active macro BS v_m. Otherpico users which request v_m for interference nulling, as well as the user served by v_m, also employ codebooks of size 2^{B_m}, but the codebooks differ from user to user to avoid the possibility of receiving the same quantization vector index from different users. The codebooks are generated by using random vector quantization [38], [39], where each vector $\mathbf{c}_{m,j}$ of $\mathcal{C}_m$ and $\mathbf{c}_{p,j}$ of $\mathcal{C}_p$ are independently chosen from the isotropic distribution on the K_m- dimensional and

$$
\begin{aligned}
\mathrm{T}_{1,LF}(\gamma) = 2p_m\lambda_m \frac{\lambda_p}{A_p}\int_{\theta=0}^{\frac{1}{\rho}} \sum_{l=0}^{K_p-1}\left(\frac{\gamma}{\kappa_p}\right)^l \theta^{\alpha l+1}\sum_{\upsilon=0}^{l}\frac{(\delta\kappa_I)^{l-\upsilon}}{\upsilon!(1+\delta\kappa_I\gamma/\kappa_p\theta^{\alpha})^{l-\upsilon+1}}\sum_{o=1}^{\mathcal{P}(\upsilon)} c_o^{\upsilon}(-1)^{\omega_o^{\upsilon}}\Gamma(\omega_o^{\upsilon}+2) \\
\times\left(p_m\lambda_m\,\Xi_0\left(\delta,\theta,\frac{\gamma}{\kappa_p}\right)+p_p\lambda_p\theta^2\,\Xi_0\left(1,1,\frac{\gamma}{\kappa_p}\right)+(1-p_m)\lambda_m\rho^2\theta^2+(1-p_p)\lambda_p\theta^2\right)^{-(\omega_o^{\upsilon}+2)} \\
\times\prod_{q=1}^{l}\left(p_m\lambda_m\delta^q\,\Xi_q\left(\delta,\theta,\frac{\gamma}{\kappa_p}\right)+\frac{p_p\lambda_p}{\theta^{\alpha q-2}}\,\Xi_q\left(1,1,\frac{\gamma}{\kappa_p}\right)\right)^{\mu_{oq}^{\upsilon}}\mathrm{d}\theta \qquad (53)
\end{aligned}
$$

K_p- dimensional unit spheres, respectively. Since the precoding vectors are now based on quantized CDIs, for the typical user $u \in \Phi_u^m$ served by the macro BS b_m, the desired channel power gain $\hat{\beta}_{b_m} \sim \text{Gamma}(\Delta_m, \kappa_m)$, where $\Delta_m = K_m - \min(Q_m, K_m - T_{\min})$ and $\kappa_m = 1 - 2^{B_m}\text{Beta}(2^{B_m}, \frac{K_m}{K_m-1})$ [31]. However, as the precoding vector of the interfering BS at $x_q \in \Psi_q \backslash b_m, q \in \{m, p\}$ is independent of the channel to the typical user u, the interference channel power gain $\hat{\zeta}_{x_q}$ is still distributed as Gamma(1, 1), i.e., Exp[1]. Similarly, for the typical user $u \in \Phi_u^p$ served by the pico BS b_p, the desired channel power gain $\hat{\beta}_{b_p} \sim \text{Gamma}(\Delta_p, \kappa_p)$, where $\Delta_p = K_p$ and $\kappa_p = 1 - 2^{B_p}\text{Beta}(2^{B_p}, \frac{K_p}{K_p-1})$. The interference channel power gain from each interfering BS other than υ_m is distributed as Exp[1]. If υ_m does not apply interference nulling, the interference channel power gain from υ_m, $\hat{\zeta}_{\upsilon_m}$ is also distributed as Exp[1]. However, if υ_m applies nulling, unlike the perfect CDI case, where the interference from υ_m is completely nulled, there will be residual interference due to the quantization error. The interference channel power gain in this case is approximated as an exponential RV with mean $\kappa_I = 2^{-\frac{B_m}{K_m-1}}$ [31]. Thus, $\hat{\zeta}_{\upsilon_m} \sim \text{Exp}[1/\kappa_I]$, if $u \in \chi$; otherwise $\hat{\zeta}_{\upsilon_m} \sim \text{Exp}[1]$. The SIR of the typical user u can be expressed as

$$\text{SIR}_l = \frac{P_l\hat{\beta}_{b_l}D_l^{-\alpha}}{\hat{I}_{b_l,m} + \hat{I}_{b_l,p}}, \quad \forall l \in \{m, p\}, \qquad (50)$$

where

$$\hat{I}_{b_l,m} = P_m \sum_{x_m \in \Psi_m \backslash b_l} \hat{\zeta}_{x_m}||x_m||^{-\alpha},$$

$$\hat{I}_{b_l,p} = P_p \sum_{x_p \in \Psi_p \backslash b_l} \hat{\zeta}_{x_p}||x_p||^{-\alpha}. \qquad (51)$$

Corollary 2: With limited feedback, the coverage probability of a typical pico-user u in the interference-limited scenario is given by

$$\mathrm{P}_{p,LF}(\gamma) = \mathrm{T}_{1,LF}(\gamma)\varphi + \mathrm{T}_{2,LF}(\gamma)(1-\varphi), \qquad (52)$$

where $\mathrm{T}_{1,LF}(\gamma)$ *is the coverage probability of* $u \in \chi$ *with limited feedback, and is given by* (53) *shown at the top of this page and* $\mathrm{T}_{2,LF}(\gamma) = \mathrm{T}_2\left(\gamma/\kappa_p\right)$ *is the coverage probability of* $u \notin \chi$*, expressed in terms of the corresponding probability for the perfect CSI,* $\mathrm{T}_2(\cdot)$*. Similarly, the coverage probability of a typical macro-user u with limited feedback is given by* $\mathrm{P}_{m,LF}(\gamma) = \mathrm{P}_m(\gamma/\kappa_m)$.

Proof: Due to the limited feedback, even when a typical pico-user u belongs to χ, it receives residual interference $Y = P_m\hat{\zeta}_m V_m^{-\alpha}$ from its nearest active macro BS, where $\hat{\zeta}_m \sim \text{Exp}[1/\kappa_I]$. Thus, the LT of total macro tier interference when $u \in \chi$ is given by

$$\begin{aligned}\mathcal{L}_{\hat{I}_{b_p,m}}(s|u \in \chi) &= \mathcal{L}^1_{I_{b_p,m}}(s)\mathbb{E}[e^{-sY}] \\ &= \mathcal{L}^1_{I_{b_p,m}}(s)(1 + sP_m\kappa_I r_1^{-\alpha})^{-1},\end{aligned}$$

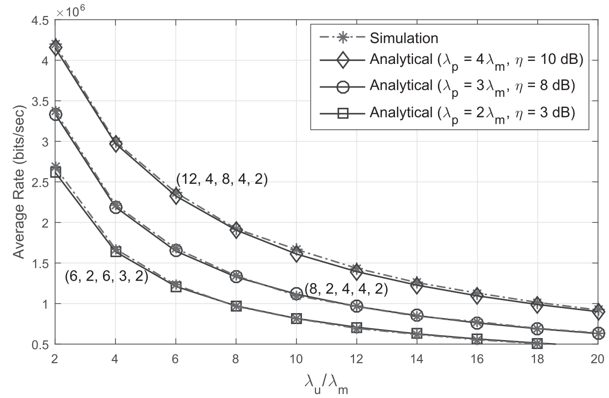
where $\mathcal{L}^1_{I_{b_p,m}}(s)$ is the LT of the total macro tier interference for the perfect CSI in (24). The LT of the total pico tier interference $\mathcal{L}_{\hat{I}_{b_p,p}}(s)$ is equal to $\mathcal{L}_{I_{b_p,p}}$ in (26). Since $\hat{\beta}_{b_p} \sim \text{Gamma}(K_p, \kappa_p)$, $\mathrm{T}_{1,LF}(\gamma)$ can then be derived in the same way as $\mathrm{T}_1(\gamma)$ in Theorem 1 with γ replaced by γ/κ_p. For $\mathrm{T}_{2,LF}(\gamma)$ and $\mathrm{P}_{m,LF}(\gamma)$, since the LTs of interference powers are the same as those of the perfect CSI case, $\mathrm{T}_{2,LF}(\gamma)$ is given by (34) with γ replaced by γ/κ_p, and similarly $\mathrm{P}_{m,LF}(\gamma)$ by (37) with γ replaced by γ/κ_m. ■

Note that $\mathrm{T}_{2,LF}(\gamma)$ and $\mathrm{P}_{m,LF}(\gamma)$ reduce to $\mathrm{T}_2(\gamma)$ and $\mathrm{P}_m(\gamma)$, respectively, if $\kappa_m = \kappa_p = 1$. Similarly, if $\kappa_p = 1$ and $\kappa_I = 0$, by using $0^0 = 1$, $\mathrm{T}_{1,LF}(\gamma)$ also reduces to $\mathrm{T}_1(\gamma)$. After deriving the coverage probabilities for limited feedback, the rate coverage and average rate can be obtained by using Theorem 2 and Theorem 3, respectively, with $P_l(\cdot)$ replaced by $P_{l,LF}(\cdot)$.

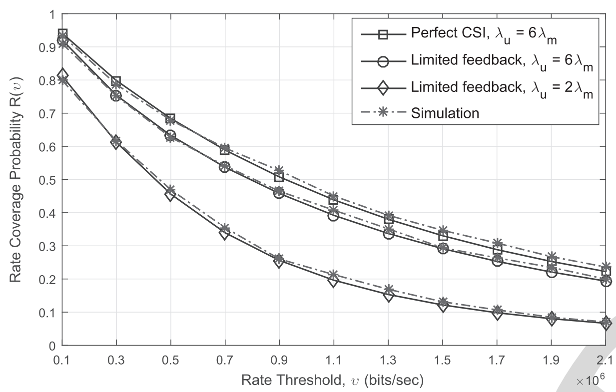
VI. Simulation and Numerical Results

In this section, we validate our analytical results via Monte Carlo simulations on a square window of 20×20 Km^2 and present numerical analysis to provide insights into important design parameters. Unless otherwise stated, we set $\delta = \frac{P_m}{P_p} = 100$, $\lambda_m = 1$BS/Km^2 and $W = 1$ MHz.

The average data rate (Theorem 3) for perfect CSI, and the data rate distribution (Theorem 2) for both the perfect CSI and limited feedback scenarios are validated via Monte Carlo simulations for different system configurations in Figure 1.a and Figure 1.b, respectively. The analytical and simulation results match with each other quite well in these figures. The PPP based assumptions of the thinned processes Φ_u^m, Φ_u^p and Ψ_u^p obtained from the parent process Φ_u hardly impact the probability distributions of the number of users of corresponding sets in a typical cell. The small gaps between the simulations and analytical curves are thus mostly due to the approximation of cell area distribution by Gamma. Note that the validation of Theorem 3 for perfect CSI naturally validates the conditional SIR distributions

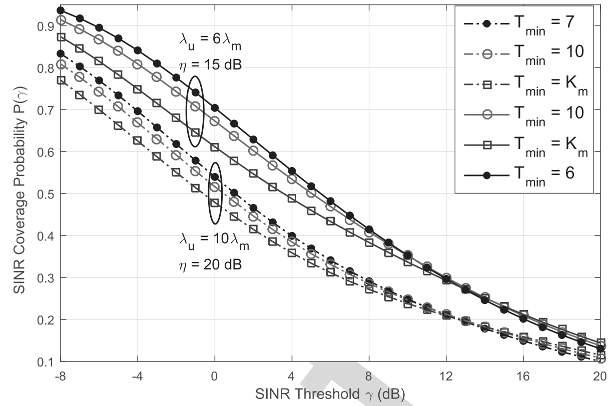


(a)

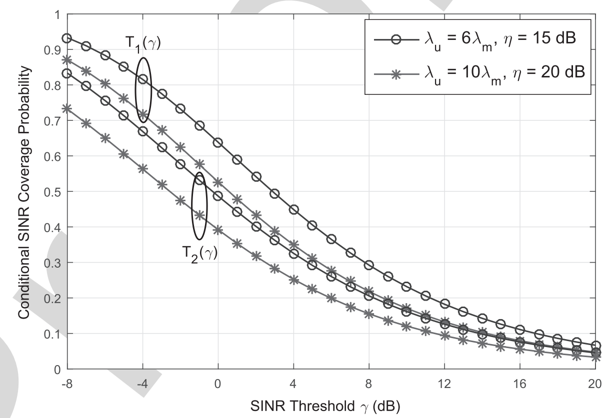


(b)

Fig. 1. (a) Validation of the average user data rate (Theorem 3) with perfect CSI for different values of λ_p, η and $(K_m, L_{\max}^m, T_{\min}, K_p, L_{\max}^p)$; (b) Validation of the rate coverage probability (Theorem 2) for both the perfect CSI and limited feedback scenarios: $K_m = 12$, $K_p = 4$, $L_{\max}^m = L_{\max}^p = 1$, $T_{\min} = 2$, $\lambda_u = 10\lambda_m$, $\alpha = 3.5$, $\eta = 15$dB.



(a)



(b)

Fig. 2. Impact of interference nulling on the SIR coverage probability: $K_m = 14$, $L_{\max}^m = 4$, $K_p = 6$, $L_{\max}^p = 4$, $\lambda_p = 6\lambda_m$, $\alpha = 3.5$.

derived in Theorem 1 and Corollary 1, and the validation of Theorem 2 for limited feedback validates the SIR distribution in Corollary 2. In Figure 1.a, the average data rate decreases with an increase in user density λ_u because of the increase in interference and the decrease in users' share of resources. The interference power increases with an increase in user density because not just more BSs become active, but the average channel power gain from each interfering BS also increases until the number of users associated with the BS exceeds $L_{\max}^l$.

In Figure 2, we analyze the impact of interference nulling on the SIR coverage probability, where $T_{\min} = K_m$ implies no interference nulling employed. While the overall SIR coverage of a typical user is plotted in Figure 2.a, the coverage probability conditioned that the user belongs to pico tier and always gets the interference from its nearest active macro BS nulled, $T_1(\gamma)$ is compared against that its no-nulling counterpart, $T_2(\gamma)$ in Figure 2.b. Figure 2.a reveals that with properly chosen $T_{\min}$, the SIR coverage can be significantly improved with interference nulling. For example, if the required SIR level for a typical user to be under coverage is 0 dB, the average fraction of users under coverage improves from 61% to 70% with interference nulling for the $\lambda_u = 6\lambda_m$, $\eta = 15$ dB case. In both Figure 2.a and Figure 2.b, the performance gain decreases with an increasing threshold. At smaller values of thresholds, as interference nulling improves the SIRs of poor cell-edge pico-user lacking coverage due to strong interference from their corresponding nearest active macro BSs, the coverage probability of thepico users significantly improves. On the other hand, we know that the SIR of a typical macro-user degrades due to interference nulling as it costs the user its available DoF. At lower values of SIR thresholds, the degradation in SIR is, however, not significant enough to impact its coverage probability. Thus, the overall gain in coverage probability is high at smaller threshold levels. However, at larger threshold values, the users under coverage are basically those in the cell interior. Thus, interference nulling may not significantly improve the already high SIR of cell-interiorpico users, resulting in minimal improvement in pico coverage probability. The SIR degradation of macro-users due to interference nulling, which do not have any significance on macro coverage probability at lower thresholds eventually causes the coverage probability to degrade after certain level. This degradation further reduces the overall gain in coverage probability.

In Figure 2.a, the performance gain in the overall coverage probability for $\lambda_u = 10\lambda_m$, $\eta = 20$ dB is relatively low compared to the $\lambda_u = 6\lambda_m$, $\eta = 15$ dB case. However, in Figure 2.b, given that the nulling is performed for each pico-

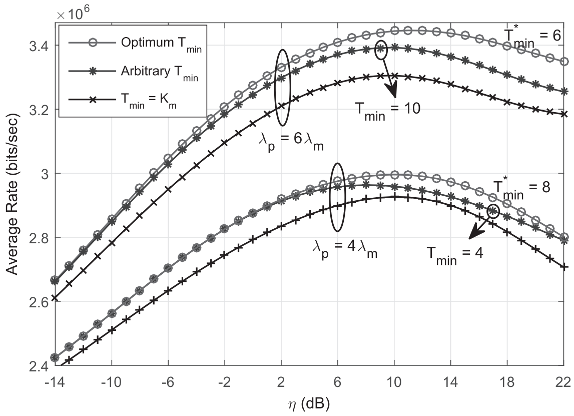


Fig. 3. Effect of pico cell density λ_p on the optimal choices of $T_{\min}$ and η: $\lambda_u = 6\lambda_m$, $K_m = 12$, $L_{\max}^m = 4$, $K_p = 4$, $L_{\max}^p = 4$, $\alpha = 4$.

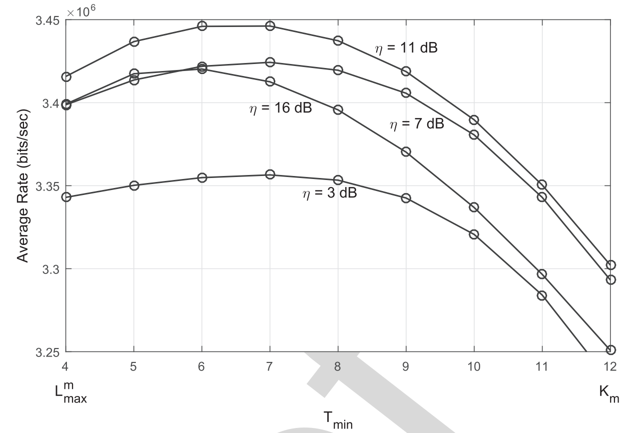


Fig. 4. Average rate vs. $T_{\min}$ for different values of η: $\lambda_u = 6\lambda_m$, $\lambda_p = 6\lambda_m$, $K_m = 12$, $L_{\max}^m = 4$, $K_p = 4$, $L_{\max}^p = 4$, $\alpha = 4$.

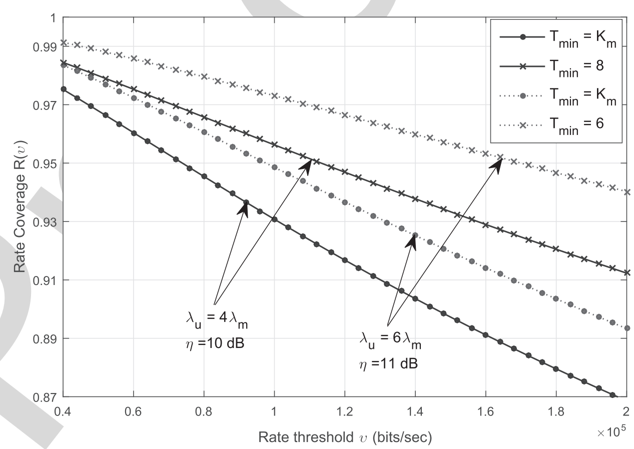


Fig. 5. Effect of interference nulling on cell-edge data rate: $\lambda_p = 6\lambda_m$, $K_m = 12$, $L_{\max}^m = 4$, $K_p = 4$, $L_{\max}^p = 4$, $\alpha = 4$.

user, both cases have similar gains in pico coverage probability due to nulling. Thus, the reason for the lower performance gain for higher user density λ_u and higher bias η is the lack of sufficient resources for interference nulling. For the $\lambda_u = 6\lambda_m$ and $\eta = 15$ dB case, with $T_{\min} = 6$, interference to 83% of thepico users from their corresponding nearest active macro BSs are nulled. The fraction of interference nulledpico users reduces to 53% for $\lambda_u = 10\lambda_m$ and $\eta = 20$ dB, with optimal $T_{\min}$ of 7.

Next, we investigate the optimal value of η to maximize the average user data rate. η controls the number of users offloaded from the macro to the pico tier to obtain a balanced distribution of the user load across tiers so that the radio resources are better utilized in each tier. Meanwhile, since $T_{\min}$ determines the spatial DoF available for serving the macro-users, as well as the number of interference-nulledpico users, $T_{\min}$ must be tuned according to user offloading. The joint tuning of $T_{\min}$ and η for optimal average data rate is investigated in Figure 3. The optimal pair $(\eta, T_{\min})$ is found to be (10 dB, 8) and (11 dB, 6) for pico density $\lambda_p = 4\lambda_m$ and $\lambda_p = 6\lambda_m$, respectively. For the given user density, the optimal $T_{\min}$ decreases with the increase in pico density because the number of interference-nulling requests received by a typical active macro BS increases with the increase in pico density. Thus, the allocated interference-nulling resources $(K_m - T_{\min})$ must be increased.

The variation in the average rate with $T_{\min}$ for the given value of η is plotted in Figure 4. The average rate of the macro-users increases with an increasing $T_{\min}$ due to the increase in the spatial DoF available at each macro BS for serving its own users. In contrast, the average pico rate decreases with an increasing $T_{\min}$ due to the decrease in the number of interference nulledpico users. The net result is the initial increase in the average rate with an increasing $T_{\min}$ and the subsequent decrease beyond a certain value of $T_{\min}$. The optimal $T_{\min}$ shifts towards the lower values as the value of η increases. For example, the optimal $T_{\min}$ of 7 for $\eta = 3$ dB decreases to 6 for $\eta = 11$ dB and to 5 for $\eta = 16$ dB. With an increasing η, more users are offloaded to the pico tier. Thus, allocating more antenna resources for interference nulling is desirable.

In Figure 5, the rate coverage corresponding to the optimal pair $(\eta, T_{\min})$ which maximized the average rate in Figure 3 for $\lambda_p = 4\lambda_m$ and $\lambda_p = 6\lambda_m$ is plotted. Let the 5th percentile rate R_{95}, which corresponds to the 5th percentile of the users with rate less than R_{95} (i.e., $\mathcal{R}(R_{95}) = 0.95$), be considered as the cell-edge data rate. For $\lambda_p = 4\lambda_m$ and $\eta = 10$ dB, $T_{\min} = 8$, which maximized the average rate is found to improve the cell-edge rate from 7.2×10^4 bits/sec to 1.12×10^5 bits/sec as compared to that without interference nulling. Similarly, for $\lambda_p = 6\lambda_m$, the cell-edge rate improves from 9.6×10^4 bits/sec to 1.68×10^5 bits/sec if interference nulling with $T_{\min} = 6$ is employed corresponding to $\eta = 11$ dB.

In Figure 6, the average data rate is assessed for different values of $L_{\max}^m$ and $L_{\max}^p$ with and without interference nulling. The curve corresponding to the interference nulling employed is plotted by computing the average rate with optimum $T_{\min}$ for each corresponding value of $L_{\max}^m$ and $L_{\max}^p$. As Figure 6 reveals, the average data rate can be significantly improved by selecting a proper value of $L_{\max}^m$ compared to either SU-BF or full-SDMA, and similarly a proper value of $L_{\max}^p$. For the case with no interference nulling employed, in which all the antennas at each macro BS are used for serving its own users, the variation of $L_{\max}^m$ has little or no impact on the average

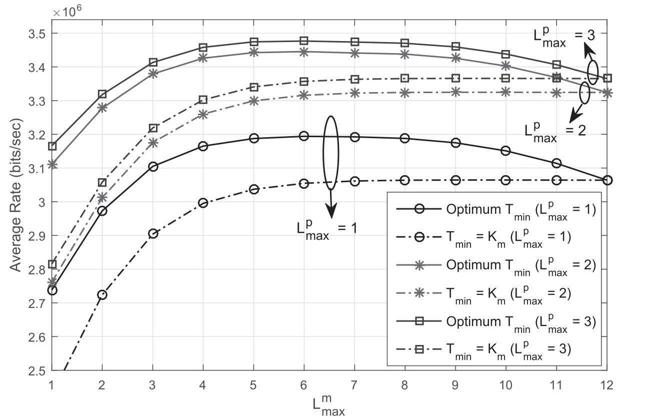


Fig. 6. Average rate vs. $L_{\max}^m$ for different values of $L_{\max}^p$ with optimum $T_{\min}$ and no interference nulling: $\lambda_p = 6\lambda_m$, $\lambda_u = 6\lambda_m$, $K_m = 12$, $K_p = 4$, $\eta = 12\,\text{dB}$, $\alpha = 4$.

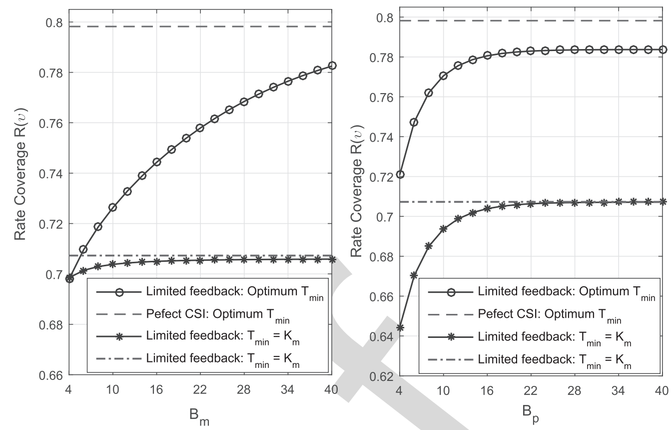


Fig. 7. Impact of number of feedback bits on the rate coverage performance: $\lambda_p = 6\lambda_m$, $\lambda_u = 10\lambda_m$, $K_m = 12$, $K_p = 4$, $L_{\max}^m = L_{\max}^p = 1$, $\eta = 15\,\text{dB}$, $\alpha = 3.5$.

rate from $L_{\max}^m = 7$ to $L_{\max}^m = 12$. This result can be observed for each given value of $L_{\max}^p$ because beyond $L_{\max}^m = 7$, the number of users simultaneously served by a macro BS in each time slot is limited by the number of users in that cell, rather than $L_{\max}^m$. This explanation is further corroborated by the fact that with interference nulling employed, the optimal $T_{\min}$ beyond $L_{\max}^m = 7$ is found to be the corresponding $L_{\max}^m$ itself, which is the minimum possible value of $T_{\min}$. Since beyond $L_{\max}^m = 7$, the number of macro-users in a cell is typically less than $L_{\max}^m$, allocating more antenna resources than $L_{\max}^m$ would be wasting resources as those surplus resources can be utilized for performance improvement through interference nulling. For each possible value of $L_{\max}^p$, the optimal pair $(L_{\max}^m, T_{\min})$ which maximizes the average rate is found to be $(6, 7)$. The average rate slightly degrades for $L_{\max}^p = 4$ as compared to $L_{\max}^p = 3$ (not shown in the figure). Thus, the optimal values of $L_{\max}^m$, $T_{\min}$, and $L_{\max}^p$ for the given system configuration are 6, 7, and 3, respectively.

After numerically analyzing the proposed SDMA scheme with interference nulling for the perfect CSI, we now investigate the impact of limited feedback on the performance. As explained in Section V, each macro-user feeds back B_m CSI bits to its home BS. In contrast, each pico-user feeds back B_p CSI bits to its home BS and B_m CSI bits to its nearest active macro BS if the BS is performing interference nulling to the user. In Figure 7, the impact of the number of feedback bits B_m and B_p on the rate coverage with and without interference nulling is investigated. As the number of feedback bits increases, the performance approaches that of the perfect CSI. Clearly, the impact of limited feedback bits B_m on the performance is higher for the interference-nulling scenario than that without nulling. $B_m > 16$, which is more than sufficient for the non-coordination case, appears to be insufficient for interference nulling case to reap the full benefits of nulling. Nevertheless, nulling does improve performance even with limited feedback as compared to the non-coordination case. With no interference nulling employed, the feedback bits B_m are only required for signal power boosting to the single user being served in the cell and such processing is found to be less sensitive to CSI errors as compared to interference nulling. If we observe the rate coverage curve against B_p for the non-coordination case, $B_p > 20$ is near perfect. However, we can observe a performance gap for interference nulling case even beyond $B_p = 20$ because of the limitation in B_m, which is considered to be 40 in this case.

VII. Conclusion

We analyzed the downlink performance of multi-antenna HetNets with SDMA, in which the ZF precoding matrix at macro BS also considered interference nulling to certainpico users. Further, the number of users served with SDMA in each cell was a function of user distribution. Our results showed that the SIR and rate coverage of victimpico users (those suffering strong interference from macro BS) can be significantly improved with the proposed interference nulling scheme if $T_{\min}$ is carefully chosen. The optimal choice of $T_{\min}$ for maximum data rate was found to be coupled with association bias. The optimal values of $L_{\max}^m$ and $L_{\max}^p$ which maximize the average data rate was also investigated and were found to outperform both SU-BF and full-SDMA. The impact of CSI quantization error on the performance of interference nulling due to limited feedback was also analyzed. It was observed that interference nulling is highly sensitive to CSI errors as the residual interference due to CSI imperfection significantly degrades the performance. However, depending on the degree of CSI imperfection, the performance may still be better than that without interference nulling.

Appendix

A. Proof of Theorem 1

By substituting (23) into (30), followed by $\Delta_p = K_p - M'_p$, and then averaging over the joint PDF of D_p and V_m, expressed as $f_{V_m|D_p}(r_1) f_{D_p}(r)$, and the PMF of M'_p, we get (32), where

$$\mathrm{T}_1(\gamma) = \int_{r=0}^{\infty}\int_{r_1=\rho r}^{\infty} \sum_{k=0}^{L_{\max}^p - 1} \mathbb{P}(M'_p = k) \sum_{l=0}^{K_p-k-1} \frac{(-s)^l}{l!} \times \frac{\mathrm{d}^l}{\mathrm{d}s^l}\Big(\mathcal{L}^1_{I_{b_p,m}}(s)\mathcal{L}_{I_{b_p,p}}(s)\Big)\Big|_{s=\frac{\gamma r^\alpha}{P_p}} f_{V_m|D_p}(r_1|r) f_{D_p}(r)\, dr_1\, dr, \tag{54}$$

and $\mathrm{T}_2(\gamma)$ is given by a similar expression with $\mathcal{L}^1_{I_{b_p,m}}(s)$ replaced by $\mathcal{L}^2_{I_{b_p,m}}(s)$. However, since the LT in $\mathrm{T}_2(\gamma)$ is not a function of r_1, averaging over the PDF of D_p only is required. We thus derive $\mathrm{T}_1(\gamma)$ first, as $\mathrm{T}_2(\gamma)$ then follows immediately.

Let $y(s) = e^{-\pi s}$, and $t(s) = p_m \lambda_m r_1^2\, \Xi_0^m \left(1, 1, \frac{P_m}{r_1^\alpha} s\right) + p_p \lambda_p r^2\, \Xi_0^p \left(1, 1, \frac{P_p}{r^\alpha} s\right)$. The LT in (53) can be expressed as $\mathcal{L}^1_{I_{b_p,m}}(s)\mathcal{L}_{I_{b_p,p}}(s) = e^{\pi\left(p_m \lambda_m r_1^2 + p_p \lambda_p r^2\right)} y(t(s))$, the lth derivative of which can be evaluated by applying Faà di Bruno's formula (31). While computing the lth derivative, we use $y_{t(s)}^{(\omega_o^l)}(t(s)) = (-\pi)^{\omega_o^l} \exp(-\pi t(s))$;

$$\frac{d^q}{ds^q} \Xi_0^l \left(1, 1, \frac{P_l}{\varpi_l^\alpha} s\right) = \left(-\frac{P_l}{\varpi_l^\alpha}\right)^q \Xi_q^l \left(1, 1, \frac{P_l}{\varpi_l^\alpha} s\right), \tag{55}$$

which follows from the property of the Gauss Hypergeometric function; and the properties of integer partition $\sum_{q=1}^{l} q \mu_{oq}^l = l$ and $\sum_{q=1}^{l} \mu_{oq}^l = \omega_o^l$. The final expression for $\mathrm{T}_1(\gamma)$ in (33) is then obtained by changing the order of integration, followed by substituting $\frac{r}{r_1} \to \theta$, $r_1 \to r_1$, then integrating with respect to r_1.

References

[1] "LTE Advanced: Heterogeneous networks," Q. Incorp., White Paper, Jan. 2011.

[2] S. Alamouti, "A simple transmit diversity technique for wireless communications," *IEEE J. Sel. Areas Commun.*, vol. 16, no. 8, pp. 1451–1458, Oct. 1998.

[3] H. Jafarkhani, *Space-Time Coding: Theory and Practice*. Cambridge, U.K.: Cambridge Univ. Press, 2005.

[4] M. Kang and M. S. Alouini, "Largest eigenvalue of complex Wishart matrices and performance analysis of MIMO MRC systems," *IEEE J. Sel. Areas Commun.*, vol. 21, no. 3, pp. 418–426, Apr. 2003.

[5] G. J. Foschini, "Layered space-time architecture for wireless communication in a fading environment when using multi-element antennas," *Bell Labs Techn. J.*, vol. 5, no. 2, pp. 41–59, Autumn 1996.

[6] D. Gesbert, M. Kountouris, R. W. Heath, Jr., C.-B. Chae, and T. Sälzer, "From single user to multiuser communications: Shifting the MIMO paradigm," *IEEE Signal Process. Mag.*, vol. 24, no. 5, pp. 36–46, Sep. 2007.

[7] Q. H. Spencer, A. L. Swindlehurst, and M. Haardt, "Zero-forcing methods for downlink spatial multiplexing in multiuser MIMO channels," *IEEE Trans. Signal Process.*, vol. 52, no. 2, pp. 461–471, Feb. 2004.

[8] J. G. Andrews, W. Choi, and R. W. Heath, Jr., "Overcoming interference in spatial multiplexing MIMO cellular networks," *IEEE Wireless Commun.*, vol. 14, no. 6, pp. 95–104, Dec. 2007.

[9] J. Hoydis, S. ten Brink, and M. Debbah, "Massive MIMO in the UL/DL of cellular networks: How many antennas do we need?" *IEEE J. Sel. Areas Commun.*, vol. 31, no. 2, pp. 160–171, Feb. 2013.

[10] D. Ying, H. Yang, T. L. Marzetta, and D. J. Love, "Heterogeneous massive MIMO with small cells," in *Proc. IEEE Veh. Technol. Conf. (VTC-Spring)*, Nanjing, China, May 2016, pp. 1–6.

[11] Y. Kim *et al.*, "Full dimension MIMO (FD-MIMO): The next evolution of MIMO in LTE systems," *IEEE Wireless Commun. Mag.*, vol. 21, no. 3, pp. 92–100, Jun. 2014.

[12] Q. Ye, O. Y. Bursalioglu, H. C. Papadopoulos, C. Caramanis, and J. G. Andrews, "User association and interference management in massive MIMO HetNets," *IEEE Trans. Commun.*, vol. 64, no. 5, pp. 2049–2065, May 2016.

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

[14] M. D. Renzo and P. Guan, "Stochastic geometry modeling of coverage and rate of cellular networks using the Gil-Pelaez inversion theorem," *IEEE Commun. Lett.*, vol. 18, no. 9, pp. 1575–1578, Sep. 2014.

[15] L. Afify, H. ElSawy, T. Al-Naffouri, and M. Alouini, "A Unified stochastic geometry model for MIMO cellular networks with retransmissions," *IEEE Trans. Wireless Commun.*, vol. 15, no. 12, pp. 8595–8609, Dec. 2016.

[16] V. Chandrasekhar, M. Kountouris, and J. G. Andrews, "Coverage in multi-antenna two-tier networks," *IEEE Trans. Wireless Commun.*, vol. 10, no. 10, pp. 5314–5327, Oct. 2009.

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

[18] H. S. Dhillon, M. Kountouris, and J. G. Andrews, "Downlink MIMO HetNets: Modeling, ordering results and performance analysis," *IEEE Trans. Wireless Commun.*, vol. 12, no. 10, pp. 5208–5222, Oct. 2013.

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

[20] C. Li, J. Zhang, J. G. Andrews, and K. B. Letaief, "Success probability and area spectral efficiency in multiuser MIMO HetNets," *IEEE Trans. Commun.*, vol. 64, no. 4, pp. 1544–1556, Apr. 2016.

[21] S. T. Veetil, K. Kuchi, and R. K. Ganti, "Performance of PZF and MMSE receivers in cellular networks with multi-user spatial multiplexing," *IEEE Trans. Wireless Commun.*, vol. 14, no. 9, pp. 4867–4878, Sep. 2015.

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

[23] M. Cierny, H. Wang, R. Wichman, Z. Ding, and C. Wijting, "On number of almost blank subframes in heterogeneous cellular networks," *IEEE Trans. Wireless Commun.*, vol. 12, no. 10, pp. 5061–5073, Oct. 2013.

[24] Y. Dhungana and C. Tellambura, "Multi-channel analysis of cell range expansion and resource partitioning in two-tier heterogeneous cellular networks," *IEEE Trans. Wireless Commun.*, vol. 15, no. 3, pp. 2306–2394, Mar. 2016.

[25] T. D. Novlan, R. K. Ganti, A. Ghosh, and J. G. Andrews, "Analytical evaluation of fractional frequency reuse for heterogeneous cellular networks," *IEEE Trans. Commun.*, vol. 60, no. 7, pp. 2029–2039, Jul. 2012.

[26] A. Shojaeifard, K. Hamdi, E. Alsusa, D. So, and J. Tang, "Design, modeling, and performance analysis of multiantenna heterogeneous cellular networks," *IEEE Trans. Commun.*, vol. 64, no. 7, pp. 3104–3118, Jul. 2016.

[27] J. Zhang, R. Chen, J. G. Andrews, A. Ghosh, and R. W. Heath, Jr., "Networked MIMO with clustered linear precoding," *IEEE Trans. Wireless Commun.*, vol. 8, no. 4, pp. 1910–1921, Apr. 2009.

[28] M. Feng, X. She, L. Chen, and Y. Kishiyama, "Enhanced dynamic cell selection with muting scheme for DL CoMP in LTE-A," in *Proc. IEEE Veh. Technol. Conf. (VTC-Spring)*, Taipei, Taiwan, May 2010, pp. 1–5.

[29] Y. Wu, Y. Cui, and B. Clerckx, "Analysis and optimization of inter-tier interference coordination in downlink multi-antenna HetNets with offloading," *IEEE Trans. Wireless Commun.*, vol. 14, no. 12, pp. 6550–6564, Dec. 2015.

[30] 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.

[31] J. Zhang and J. G. Andrews, "Adaptive spatial intercell interference cancellation in multicell wireless networks," *IEEE J. Sel. Areas Commun.*, vol. 28, no. 9, pp. 1455–1468, Dec. 2010.

[32] 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. 1419–1430, Apr. 2015.

[33] 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.

[34] S. N. Chiu, D. Stoyan, W. S. Kendall, and J. Mecke, *Stochastic Geometry and its Applications*, 3rd ed. Hoboken, NJ, USA: Wiley, 2013.

[35] J.-S. Ferenc and Z. Neda, "On the size distribution of Poisson-Voronoi cells," *Phys. A, Statist. Mech. Appl.*, vol. 385, no. 2, pp. 518–526, 2007.

[36] N. Jindal, J. G. Andrews, and S. Weber, "Multi-antenna communication in ad hoc networks: Achieving MIMO gains with SIMO transmission," *Phys. A, Statist. Mech. Appl.*, vol. 59, no. 2, pp. 529–540, Feb. 2011.

[37] I. S. Gradshteyn and I. M. Ryzhik, *Table of Integrals, Series, and Products*, 7th ed. San Diego, CA, USA: Academic, 2007.

[38] N. Jindal, "MIMO broadcast channels with finite rate feedback," *IEEE Trans. Inf. Theory*, vol. 52, no. 11, pp. 5045–5060, Nov. 2006.

[39] C. K. Au-yeung and D. J. Love, "On the performance of random vector quantization limited feedback beamforming in a MISO system," *IEEE Trans. Wireless Commun.*, vol. 6, no. 2, pp. 458–462, Feb. 2007.

Yamuna Dhungana (S'15) received the B.Eng. degree in electronics and communications engineering from the Institute of Engineering, Tribhuvan University, Nepal, in 2009, the M.Eng. degree in telecommunications from the Asian Institute of Technology, Thailand, in 2011, and the Ph.D. degree in electrical and computer engineering from the University of Alberta, Edmonton, AB, Canada in 2016. She is currently a Research and Development Engineer with Ericsson Canada Inc., Ottawa, ON, Canada. Her research interests include wireless communications theory, heterogeneous cellular networks, and MIMO systems.

Chintha Tellambura (F'11) received the B.Sc. degree (Hons.) from the University of Moratuwa, Moratuwa, Sri Lanka, in 1986, the M.Sc. degree in electronics from the King's College, University of London, London, U.K., in 1988, and the Ph.D. degree in electrical engineering from the University of Victoria, Victoria, BC, Canada, in 1993.

He was a Post-Doctoral Research Fellow with the University of Victoria from 1993 to 1994, and the University of Bradford from 1995 to 1996. He was with Monash University, Melbourne, Australia, from 1997 to 2002. He is currently a Professor with the Department of Electrical and Computer Engineering, University of Alberta. His current research interests include the design, modelling and analysis of cognitive radio, heterogeneous cellular networks and 5G wireless networks.

He served as an editor for IEEE TRANSACTIONS ON COMMUNCIATIONS from 1999 to 2011 and for TRANSACTIONS ON WIRELESS COMMUNICATIONS from 2001 to 2007. He was the Area Editor for Wireless Communications Systems and Theory in the IEEE TRANSACTIONS ON WIRELESS COMMUNICATIONS from 2007 to 2012. He and the co-authors received the Communication Theory Symposium Best Paper Award in the 2012 IEEE International Conference on Communications, Ottawa, Canada. He is the winner of the prestigious McCalla Professorship and the Killam Annual Professorship from the University of Alberta. He has authored/coauthored over 500 journal and conference publications with total citations more than 12,000, and has an h-index of 59 in Google Scholar.