Non-orthogonal Multiple Access in Hybrid Heterogeneous Networks

Yuanwei Liu, Zhijin Qin, Maged Elkashlan, Arumugam Nallanathan, and Julie A. McCann

***Abstract*—In this paper, the application of non-orthogonal multiple access (NOMA) into K-tier heterogeneous networks (HetNets) is investigated. A new promising transmission framework is proposed, in which massive multiple-input multiple-output (MIMO) is employed in macro cells and NOMA is adopted in small cells. For maximizing the biased average received power at mobile users, a massive MIMO and NOMA based user association scheme is developed. In an effort to evaluate the performance of the proposed framework, analytical expressions for the spectrum efficiency of each tier are derived using stochastic geometry. Simulation results are presented to verify the accuracy of the proposed analytical derivations and confirm that NOMA is capable of enhancing the spectrum efficiency of the network compared to the orthogonal multiple access (OMA) based HetNets.**

***Index Terms*—HetNets, massive MIMO, NOMA, stochastic geometry**

I. Introduction

Non-orthogonal multiple access (NOMA), as a promising technology in 5G networks, has attracted much attention in both industry and academia [1–4] for its potential ability to enhance spectrum efficiency. The system-level performance of a two user NOMA system in terms of downlink transmission was demonstrated in [1]. In [2], the performance of a general NOMA transmission was evaluated in which one BS is able to communicate with several spatial randomly deployed users. By examining appropriate power allocation policies among the NOMA users, the fairness issue of NOMA was addressed in [3]. On the standpoint of tackling spectrum efficiency and energy efficiency, an incentive user cooperation NOMA protocol was proposed in [4], by regarding near users as energy harvesting relays for improving reliability of far users.

Heterogeneous networks (HetNets) and massive multiple-input multiple-output (MIMO), as two of the "big three" technologies [5], laid the fundamental structure for emerging 5G communication systems. The massive MIMO regime enables to equip tens of hundreds/thousands antennas at a BS, and hence is capable of offering an unprecedented level of freedom to serve multiple mobile users. The core idea of HetNets is to establish closer BS-user link by densely overlaying small cells. By doing so, the promising benefits such as lower power consumption, higher throughput and enhanced spatial reuse of spectrum can be experienced. Aiming to fully take advantages of both massive MIMO and HetNets, several research contributions have been made [6–9]. In [6], the interference coordination issue of massive MIMO enabled HetNets was addressed by utilizing the spatial blanking of macro cells. In [7], the authors investigated a joint user association and interference management optimization problem in massive MIMO HetNets. By applying stochastic geometry model, the spectrum efficiency of uplink and downlink massive MIMO aided HetNets were evaluated in [8] and [9], respectively.

Among the recent research contributions towards 5G, NOMA based HetNets has not been well investigated yet and is still in its infancy. We believe that the novel structure design in this work—by introducing NOMA based small cells in massive MIMO enabled HetNets—can be a new highly rewarding candidate, which will contribute to the design of a more promising 5G system due to the following key advantages:

- High spectrum efficiency: NOMA improves the spectrum efficiency with multiplexing users in power domain and invoking succussive interference cancelation (SIC) technique for canceling interference. In NOMA based HetNets, with employing higher BS densities, BSs are capable of accessing the served users closer, which can increase the signal-to-interference-plus-noise ratio (SINR) by intelligently tracking the multi-category interference, such as inter/intra-tier interference and intra-BS interference.
- High compatibility and low complexity: NOMA is regarded as a promising "add-on" technology for the existing multiple access systems due to the gradually mature of superposition coding (SC) and SIC technologies [10], and will not bring much implementation complexity. Additionally, with applying NOMA in the single-antenna based small cells, the complex precoding/cluster design for MIMO-NOMA systems can be avoided.
- Fairness/throughput tradeoff: NOMA is capable of dealing with the fairness issue by allocating more power to weak users, which is of great significance for HetNets when investigating efficient resource allocation in the sophisticated multi-tier networks.

and will not bring much implementation complexity or modification for the existing networks. Additionally, with applying NOMA in the single-antenna based small cells, the complex precoding/cluster design for the multi-antenna NOMA can be avoided.

Motivated by the aforementioned potential benefits, we

Y. Liu and M. Elkashlan are with Queen Mary University of London, London, UK (email: {yuanwei.liu, maged.elkashlan}@qmul.ac.uk).

Z. Qin and J. McCann are with Imperial College London, London, UK (email: {z.qin, j.mccann}@imperial.ac.uk).

A. Nallanathan is with King's College London, London, UK (e-mail: arumugam.nallanathan@kcl.ac.uk).

propose a novel hybrid HetNets framework with NOMA based small cells and massive MIMO aided macro cells to further enhance the performance of existing HetNets design. In this framework, we consider a downlink K-tier HetNets, where macro BSs are equipped with large antenna arrays with linear zero-forcing beamforming (ZFBF) capability to serve multiple single-antenna users simultaneously, and small cells BSs are equipped with single antenna each to serve two single-antenna users simultaneously with NOMA transmission. A stochastic geometry approach is adopted to model the considered K-tier HetNets. Based on the proposed framework, the primary contributions can be summarized as follows: 1) We consider the flexible biased association to address the impact of NOMA and massive MIMO on the maximum biased received power; 2) We derive the exact analytical expressions of the NOMA based small cells in term of spectrum efficiency; and 3) We show that NOMA based small cells are capable of achieving higher spectrum efficiency compared to conventional orthogonal multiple access (OMA) based small cells, which demonstrates the benefits of the proposed framework.

II. Network Model

A. Network Description

In this paper, we focus on the downlink transmission scenarios. We consider a K-tier HetNets model, where the first tier represents the macro cells and the other tiers represent the small cells such as pico cells and femto cells. The positions of macro BSs and all the k-th tier ($k \in \{1, \cdots, K\}$) BSs are modeled as homogeneous poisson point processes (HPPPs) Φ_k and with density λ_k, respectively. α_k is the path loss exponent of the k-th tier cells. All channel are assumed to undergo quasi-static Rayleigh fading, where the channel coefficients are constant for each transmission block but independent between different blocks.

Motivated by the fact that it is common to overlay a high-power macro cell with successively denser and lower power small cell, we consider to apply massive MIMO technologies to macro cells and NOMA transmission to small cells in this work. As shown in Fig. 1, in macro cells, macro BSs are considered to be equipped with M antennas, each macro BS transmit signals to N users over the same resource block (e.g., time/frequency/code)[1]. We assume $M \gg N > 1$ and the linear ZFBF technique is applied at each macro BS with assigning equal power to N data streams [11]. In small cells, each small cell BS is considered to be equipped with single antenna. In other words, in this scenario, macro cells are OMA based and small cells are NOMA based. All users are considered to be equipped with single antenna each as well. We consider to adopt user pairing in each tier of small cells to implement NOMA for lowering the system complexity [4]. It is worth pointing out that in Long term evolution advanced (LTE-A), NOMA is also in a form of two-user case [12]. All users are considered to be equipped with single antenna each.

[1]The aim is to avoid sophisticated MIMO-NOMA design in macro cells.

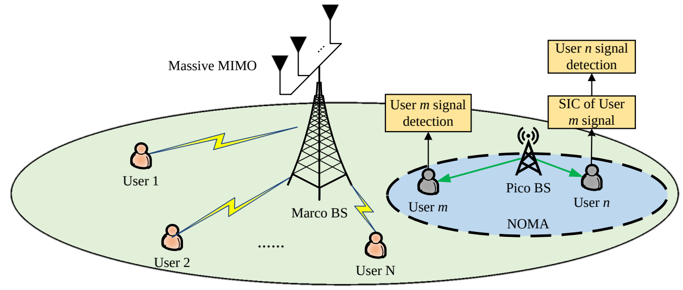

Fig. 1. Illustration of NOMA and massive MIMO based hybrid HetNets.

B. NOMA and Massive MIMO Based User Association

We assume that each tier of BS has already served one user in a random distance within a ring, the radius of rings for the k-th tier BSs is set as R_k. Then we squeeze one more user to each BS, and each BS is supposed to be able to communicate with its connected two users by applying NOMA protocol. As a result, in this work, we only focus on how to associate one more user to each BS. A user is allowed to access any tier BS, which can provide the best coverage [13]. We consider that the flexible user association is based on the maximum average received power of each tier.

1) Average received power in massive MIMO aided macro cells: In macro cells, since the macro BS is equipped with multiple antennas, users in macro cells can experience large array gains. Adopting ZFBF transmission scheme, the array gain obtained at macro users is given by $G_M = M - N + 1$ [11, 14]. As a result, the average received power that users connect with macro BS ℓ (where $\ell \in \Phi_M$) is given by

$$P_{r,M} = G_M P_M L\left(d_{\ell,1}\right)/N, \tag{1}$$

where P_M is the transmit power of the massive MIMO aided macro BSs, $L\left(d_{\ell,1}\right) = \eta d_{\ell,1}^{-\alpha_1}$ is large-scale path loss, $d_{\ell,1}$ is the distance between users and macro BSs, η is the frequency dependent factor, B_i is the identical bias factor. It is noted that the biasing factor B_i is useful for offloading data traffic in HetNets [13].

2) Average received power in NOMA based small cells: Different from the convectional user association of OMA based small cells, NOMA exploits the power sparsity for multiple access by allocating different power to different users. Due to the random spatial topology of our stochastic geometry model, the space information of users are not pre-determined. Our user association policy for the NOMA based small cells is based on assuming the typical user as far user first. As such, in the i-th tier small cell, the averaged received power that users connect with the i-th tier BS j (where $j \in \Phi_i$) is given by

$$P_{r,i} = a_{n,i} P_i L\left(d_{j,i}\right) B_i, \tag{2}$$

where P_i is the transmit power of i-th tier BS, $L\left(d_{j,i}\right) = \eta d_{j,i}^{-\alpha_i}$ is large-scale path loss, $d_{j,i}$ is the distance between the user and the i-th tier BS, α_i is the path loss exponent of the i-th tier small cells.

C. Channel Model

1) Massive MIMO aided macro cell transmission: Without loss of generality, we assume a typical user is located at the origin of an infinite two-dimension plane. Based on (1) and (2), the received SINR that a typical user connects with a macro BS at a random distance $d_{o,1}$ can be expressed as

$$\gamma_{r,1} = \frac{\frac{P_1}{N} h_{o,1} L\left(d_{o,1}\right)}{I_{M,1} + I_{S,1} + \sigma^2}, \tag{3}$$

where $I_{M,1} = \sum_{\ell \in \Phi_1 \backslash B_{o,1}} \frac{P_1}{N} h_{\ell,1} L\left(d_{\ell,1}\right)$ is the interference from macro cells, $I_{S,1} = \sum_{i=2}^{K} \sum_{j \in \Phi_i} P_i h_{j,i} L\left(d_{j,i}\right)$ is the interference from small cells, σ^2 is the additive white Gaussian noise (AWGN) power, $h_{o,1}$ is the small-scale fading coefficient between the typical user and the connected macro BS, $h_{\ell,1}$ and $d_{\ell,1}$ are the small-scale fading coefficients and distance between a typical user and connected macro BS ℓ except the serving macro BS $B_{o,1}$, respectively, $h_{j,i}$ and $d_{j,i}$ are the small-scale fading coefficients and distance between a typical user and connected i-th tier small cell BS j, respectively. Here, $h_{o,1}$ follows Gamma distribution with parameters $(M-N+1,1)$, $h_{\ell,1}$ follows Gamma distribution with parameters $(N,1)$, and $h_{j,i}$ follows exponential distribution with unit mean.

2) NOMA based small cell transmission: In small cells, without loss of generality, we consider that each small cell BS has already associated one user in the previous round of user association process. With applying NOMA protocol, we aim to squeeze a typical user into the same small cell to improve the spectral efficiency, which is one key feature of NOMA [10]. For simplicity, we assume that the distances between the existing users and the connected small cell BSs are the same as R_k, the distance between the typical user and the connected small cell BS is a random value (denoted as x). Since it is not pre-determined that the typical user is a near user n or a far user m. We denote d_{o,k_m} and d_{o,k_n} are the distance between the k-th tier small cell BS and user m and user n, respectively. As such, two possible cases can happen in the following.

Near user case: In the first case, we consider that the typical user is a near user n ($x \leq R_k$), then we have $d_{o,k_m} = R_k$. User n will first decode the information of User m with the SINR $\gamma_{k_{n \to m}} = \frac{a_{m,k} P_k g_{o,k_n} L(d_{o,k_n})}{a_{n,k} P_k g_{o,k_n} L(d_{o,k_n}) + I_K + \sigma^2}$. If this procedure is successful, User n then decode its own message. As such, the interference from the existing user can be canceled. As such, the received SINR that a typical user n connects with the k-th tier small cell can be expressed as

$$\gamma_{k_n} = \frac{a_{n,k} P_k g_{o,k} L\left(d_{o,k_n}\right)}{I_{M,k} + I_{S,k} + \sigma^2}, \tag{4}$$

where $L\left(d_{o,k_n}\right) = \eta d_{o,k_n}^{-\alpha_i}$, $I_{M,k} = \sum_{\ell \in \Phi_1} \frac{P_1}{N} g_{\ell,1} L\left(d_{\ell,1}\right)$ is the interference from macro cells, $I_{S,k} = \sum_{i=2}^{K} \sum_{j \in \Phi_i \backslash B_{o,k}} P_i g_{j,i} L\left(d_{j,i}\right)$ is the interference from small cells, $g_{o,k}$ and d_{o,k_n} is the small-scale fading coefficients and distance between the typical user and the connected k-th tier BS, $g_{\ell,1}$ and $d_{\ell,1}$ are the small-scale fading coefficients and distance between a typical user and connected macro BS ℓ, respectively, $g_{j,i}$ and $d_{j,i}$ are the small-scale fading coefficients and distance between a typical user and connected i-th tier small cell BS j except the serving BS $B_{o,k}$, respectively. Here, $g_{o,k}$ and $g_{j,i}$ follow exponential distributions with unit mean. $g_{\ell,1}$ follows Gamma distribution with parameters $(N,1)$.

For the existing far user m^*, it will directly decode its own message by treating the message of user n as interference. Therefore, the received SINR that for the existing user m^* in the k-th tier small cell can be expressed as

$$\gamma_{k_{m^*}} = \frac{a_{m,k} P_k g_{o,k} L\left(R_k\right)}{I_{k,n} + I_{M,k} + I_{S,k} + \sigma^2}, \tag{5}$$

where $I_{k,n} = a_{n,k} P_k g_{o,k} L\left(R_k\right)$ is the intra-BS interference from the connected k-th tier BS with superposition information of user n, and $L\left(R_k\right) = \eta {R_k}^{-\alpha_k}$.

Far user case: In the second case, we consider that the typical user is a far user m ($x > R_k$), then we have $d_{o,k_n} = R_k$. As such, received SINR that user m connects with the k-th tier small cell can be expressed as

$$\gamma_{k_m} = \frac{a_{m,k} P_k g_{o,k} L\left(d_{o,k_m}\right)}{I_{k,n^*} + I_{M,k} + I_{S,k} + \sigma^2}, \tag{6}$$

where $I_{k,n^*} = a_{n,k} P_k g_{o,k} L\left(d_{o,k_m}\right)$ is the interference from the BS with superposition the information of existing user n^* of the k-th tier small cell with $L\left(d_{o,k_m}\right) = \eta d_{o,k_m}^{-\alpha_k}$, d_{o,k_n} is the distance between the typical user m and the connected k-th tier BS.

Regarding the existing near user n^*, it is capable of cancelling interference from the typical user m by applying SIC technique. Therefore, the received SINR of user n^* is given by

$$\gamma_{k_{n^*}} = \frac{a_{n,k} P_k g_{o,k} L\left(R_k\right)}{I_{M,k} + I_{S,k} + \sigma^2}. \tag{7}$$

III. Coverage Probability of Non-orthogonal Multiple Access Aided Small Cells

A. New Statistics

As described in Section II-B, the user association of this proposed framework is based on maximizing the biased average received power at users. As such, based on (1) and (2), the user association of macro cells and small cells are given in the following.

Using the similar method as Lemma 1 of [13], the user association probability that a typical user connects with macro BSs can be calculated as

$$A_1 = 2\pi\lambda_1 \int_0^\infty r \exp\left[-\pi \sum_{i=2}^{K} \lambda_i \left(\frac{a_{n,i} P_i B_i N}{P_1 G_M r^{-\alpha_1}}\right)^{\delta_i} - \pi\lambda_1 r^2\right] dr, \tag{8}$$

where $\delta_i = \frac{2}{\alpha_i}$. The user association probability that a typical user connects with small cell BSs in the k-th tier can be calculated as

$$\begin{aligned} A_k =& 2\pi\lambda_k \int_0^\infty r \exp\left[-\pi \sum_{i=2}^{K} \lambda_i \left(\frac{P_i B_i r^{\alpha_k}}{P_k B_k}\right)^{\delta_i}\right. \\ & \left. -\pi\lambda_1 \left(\frac{P_1 G_M r^{\alpha_k}}{N a_{n,k} P_k B_k}\right)^{\delta_1}\right] dr, \end{aligned} \tag{9}$$

where $\delta_1 = \frac{2}{\alpha_1}$. We consider the probability density function (PDF) of the distance between a typical user and the connected macro BS. Based on (8), we obtain

$$f_{d_{o,1}}(x) = \frac{2\pi\lambda_1 x}{A_1} \exp\left[-\pi\sum_{i=2}^{K}\lambda_i\left(\frac{a_{n,i}P_iB_iN}{P_1G_Mx^{-\alpha_1}}\right)^{\delta_i} - \pi\lambda_1 x^2\right] \tag{10}$$

We then calculate the PDF of the distance between a typical user and the connected k-th tier small cell BS. Based on (9), we obtain

$$f_{d_{o,k}}(x) = \frac{2\pi\lambda_k x}{A_k}\exp\left[-\pi\sum_{i=2}^{K}\lambda_i\left(\frac{P_iB_ix^{\alpha_k}}{P_kB_k}\right)^{\delta_i} - \pi\lambda_1\left(\frac{P_1G_Mx^{\alpha_k}}{Na_{n,k}P_kB_k}\right)^{\delta_1}\right]. \tag{11}$$

We first derive the Laplace transform of a typical user.

Lemma 1. *The Laplace transform of interference to a typical user can be expressed as*

$$\mathcal{L}_{I_K,t}(s) = \exp\left\{-s\sum_{i=1}^{K}\frac{\lambda_i 2\pi P_i\eta(\omega_{i,k}(x_0))^{2-\alpha_i}}{\alpha_i(1-\delta_i)}\times {}_2F_1\left(1,1-\delta_i;2-\delta_i;-sP_i\eta(\omega_{i,k}(x_0))^{-\alpha_i}\right)\right\}. \tag{12}$$

Proof:

$$\begin{aligned}\mathcal{L}_{I_K,t}(s) &= \mathbb{E}_{I_K,t}\left[e^{-sI_{K,t}}\right]\\ &\overset{(a)}{=} \mathbb{E}_{\Phi_i}\left[\sum_{i=1}^{K}\prod_{j\in\Phi_i\backslash B_{o,k}}\mathbb{E}_{g_{j,i}}\left[e^{-sP_ig_{j,i}\eta d_{j,i}^{-\alpha_i}}\right]\right]\\ &\overset{(b)}{=} \exp\left(-\sum_{i=1}^{K}\lambda_i 2\pi\int_{\omega_{i,k}(x_0)}^{\infty}\left(1-\mathbb{E}_{g_{j,i}}\left[e^{-\frac{g_{j,i}sP_i\eta}{r^{\alpha_i}}}\right]\right)rdr\right)\\ &= \exp\left(-\sum_{i=1}^{K}\lambda_i 2\pi\int_{\omega_{i,k}(x_0)}^{\infty}\left(1-\mathcal{L}_{g_{j,i}}\left(sP_i\eta r^{-\alpha_i}\right)\right)rdr\right)\\ &\overset{(c)}{=} \exp\left(-\sum_{i=1}^{K}\lambda_i 2\pi\int_{\omega_{i,k}(x_0)}^{\infty}\left(1-\left(1+sP_i\eta r^{-\alpha_i}\right)^{-1}\right)rdr\right),\end{aligned} \tag{13}$$

where $\omega_{i,k}(x_0) = \left(\tilde{P}_{ik}\tilde{B}_{ik}\right)^{\frac{1}{\alpha_i}} x_0^{\frac{1}{\tilde{\alpha}_{ik}}}$, *(a) is resulted from .... (b) is obtained by using the generating functional of PPP, (c) is obtained by* $g_{j,i}$ *follows exponential distribution with unit mean. Then applying [15, Eq. (3.194.2)], we can obtain the laplace transform in an more elegant form in* (12)*. The proof is completed.* ∎

Then we turn our attention to calculate the laplace transform of interference for the connected user, which is given by

B. Coverage Probability

The successful probability is defined as the selected typical pair of users can successfully transmit with targeted data rate R_t and R_c, for the typical user and existing connected user, respectively. According to the distances, two cases are considered in the following.

Near user case: For the near user case $x_0 < r_k$, the success decoding will happen on the condition that the following two events are both satisfied. The first one is that the typical user can decode the message of the connected user. The second one is that after the SIC process, the typical user can decode the message of its own message. The successful probability of the typical user on the condition of the distance x_0 in the k-th tier is:

$$P_t(\tau_c,\tau_t,x_0)|_{x_0\le r_k} = \Pr\left\{\gamma_{k_{n\to m}} > \tau_c, \gamma_{k_n} > \tau_t\right\}, \tag{14}$$

where $\tau_t = 2^{R_t}-1$ and $\tau_c = 2^{R_c}-1$.

Lemma 2. *Conditioned on* $x_0 \le r_k$*, the successful probability of a typical user for the near user case is expressed as*

$$P_t(\tau_c,\tau_t,x_0)|_{x_0\le r_k} = \exp\left\{-\frac{\varepsilon^*(\tau_c,\tau_t)x_0^{\alpha_k}\sigma^2}{P_k\eta} - \sum_{i=1}^{K}\frac{\lambda_i\delta_i\pi\left(\tilde{B}_{ik}\right)^{\frac{2}{\alpha_i}-1}\left(\tilde{P}_{ik}\right)^{\frac{2}{\alpha_i}}x_0^{\frac{2}{\tilde{\alpha}_{ik}}}}{1-\delta_i}Q_{i,t}^n(\tau_c,\tau_t)\right\}, \tag{15}$$

where $Q_{i,t}^n(\tau_c,\tau_t) = \varepsilon^*(\tau_c,\tau_t)\,{}_2F_1\left(1,1-\delta_i;2-\delta_i;-\frac{\varepsilon^*(\tau_c,\tau_t)}{\tilde{B}_{ik}}\right)$, $\varepsilon_c^f = \frac{\tau_c}{a_{m,k}-\tau_c a_{n,k}}$, $\varepsilon_t^n = \frac{\tau_t}{a_{n,k}}$, *and* $\varepsilon^*(\tau_c,\tau_t) = \max\left\{\varepsilon_c^f,\varepsilon_t^n\right\}$.

Proof: Based on (**??**) *and* (14)*, assuming the condition* $a_{m,k}-\tau_c a_{n,k} \ge 0$ *[2] holds, we obtain*

$$\begin{aligned}P_t(\tau_c,\tau_t,x_0)|_{x_0\le r_k} &= \Pr\left\{\frac{g_{o,k_n}P_k\eta}{x_0^{\alpha_i}(I_K+\sigma^2)} > \varepsilon^*(\tau_c,\tau_t)\right\}\\ &= e^{-\frac{\varepsilon^*(\tau_c,\tau_t)x_0^{\alpha_k}\sigma^2}{P_k\eta}}\mathbb{E}_{I_K}\left\{e^{-\frac{\varepsilon^*(\tau_c,\tau_t)x_0^{\alpha_k}y}{P_k\eta}}\right\}\\ &= e^{-\frac{\varepsilon^*(\tau_c,\tau_t)x_0^{\alpha_k}\sigma^2}{P_k\eta}}\mathcal{L}_{I_K,t}\left(\frac{\varepsilon^*(\tau_c,\tau_t)}{P_k\eta}x_0^{\alpha_k}\right)\\ &= e^{-\frac{\varepsilon^*(\tau_c,\tau_t)x_0^{\alpha_k}\sigma^2}{P_k\eta}}\prod_{i=1}^{K}\mathcal{L}_{I_i,t}\left(\frac{\varepsilon^*(\tau_c,\tau_t)}{P_k\eta}x_0^{\alpha_k}\right).\end{aligned} \tag{16}$$

Substituting (12) *into* (16)*, we obtain the successful probability for the near user case on the condition of the distance* x_0 *in the* k*-th tier. The proof is completed.* ∎

Lemma 3. *Conditioned on* $x_0 \le r_k$*, the successful probability of a connect user for the near user case is expressed as*

$$P_t(\tau_c,\tau_t,x_0)|_{x_0\le r_k} = \exp\left\{-\frac{\varepsilon^*(\tau_c,\tau_t)x_0^{\alpha_k}\sigma^2}{P_k\eta} - \sum_{i=1}^{K}\frac{\lambda_i\delta_i\pi\left(\tilde{B}_{ik}\right)^{\frac{2}{\alpha_i}-1}\left(\tilde{P}_{ik}\right)^{\frac{2}{\alpha_i}}x_0^{\frac{2}{\tilde{\alpha}_{ik}}}}{1-\delta_i}Q_{i,t}^n(\tau_c,\tau_t)\right\}, \tag{17}$$

Proof: ∎

Far user case: For the far user case $x_0 > r_k$, the success decoding will happen on the condition that the typical user can decode the message of itself by treating the connected user as noise.

Lemma 4. *Conditioned on $x_0 > r_k$, the successful probability of a typical user for the far user case is expressed as*

$$P_t\left(\tau_t, x_0\right)|_{x_0>r_k} = \exp\left\{ -\frac{\varepsilon_t^f x_0^{\alpha_k}\sigma^2}{P_k\eta} - \sum_{i=1}^{K} \frac{\lambda_i\delta_i\pi\left(\tilde{B}_{ik}\right)^{\frac{2}{\alpha_i}-1}\left(\tilde{P}_{ik}\right)^{\frac{2}{\alpha_i}} x_0^{\frac{2}{\tilde{\alpha}_{ik}}}}{1-\delta_i} Q_{i,t}^f\left(\tau_t\right) \right\}, \quad (18)$$

where $\varepsilon_t^f = \frac{\tau_t}{a_{m,k}-\tau_t a_{n,k}}$ *and* $Q_{i,t}^f\left(\tau_t\right) = \varepsilon_t^f\, {}_2F_1\left(1, 1-\delta_i; 2-\delta_i; -\frac{\varepsilon_t^f}{\tilde{B}_{ik}}\right)$.

Proof: Based on (**??**) *and assuming the condition $a_{m,k} - \tau_t a_{n,k} \geq 0$ holds, we have*

$$P_t\left(\tau_t, x_0\right)|_{x_0>r_k} = \Pr\left\{ g_{o,k_m} > \frac{\varepsilon_t^f x_0^{\alpha_i}\left(I_K+\sigma^2\right)}{P_k\eta} \right\}. \quad (19)$$

Following the similar procedure to obtain (15)*, with interchanging $\varepsilon^*\left(\tau_c, \tau_t\right)$ with ε_t^f, we obtain the desired results in* (18)*. The proof is completed.* ■

Then we calculate the successful probability of the typical user, which is given by

Theorem 1.

$$P_t\left(\tau_c, \tau_t\right) = \int_0^{r_k} P_t\left(\tau_c, \tau_t, x_0\right)|_{x_0\leq r_k} f_{d_{o,k}}^c\left(x_0\right)dx_0 + \int_{r_k}^{\infty} P_t\left(\tau_t, x_0\right)|_{x_0>r_k} f_{d_{o,k}}^c\left(x_0\right)dx_0, \quad (20)$$

where $P_t\left(\tau_c, \tau_t, x_0\right)|_{x_0\leq r_k}$ *is given in* (15) *and* $P_t\left(\tau_t, x_0\right)|_{x_0>r_k}$ *is given in* (18)*. Here,* $f_{d_{o,k}}^c\left(x_0\right)$ *can be obtained by following the similar approach of [13].* $f_{d_{o,k}}^c\left(x_0\right) = \frac{2\pi\lambda_k x_0}{A_k^c}\exp\left[-\pi\sum\limits_{i=1}^{K}\lambda_i\left(\tilde{P}_{ik}\tilde{B}_{ik}\right)^{\delta_i} x_0^{\frac{2}{\tilde{\alpha}_{ik}}}\right]$ *with* $A_k^c = 2\pi\lambda_k\int_0^\infty r\exp\left[-\pi\sum\limits_{i=1}^{K}\lambda_i\left(\tilde{P}_{ik}\tilde{B}_{ik}\right)^{\delta_i} r^{\frac{2}{\tilde{\alpha}_{ik}}}\right]dr$.

Proof:

■

IV. Spectrum Efficiency

In an effort to evaluate the spectrum efficiency of the proposed NOMA and massive MIMO based HetNets framework, we calculate the achievable ergodic rate of each tier in the following subsections.

A. Achievable Ergodic Rate of Small Cells

The aim of this work is to apply NOMA transmission in small cells to further improve the spectrum efficiency. Recall that the distance order between the connected BS and the two users are not predetermined (as aforementioned in Section II), as such, in this subsection, we calculate the achievable ergodic rate of small cells both for the near user case and far user case in *Lemma 5* and *Lemma 6* in the following respectively.

Lemma 5. *Conditioned on the HPPPs, the achievable ergodic rate of the k-th tier small cell for the near user case can be expressed as follows:*

$$\tau_k^n = \frac{1}{\ln 2}\int_0^{\frac{a_{m,k}}{a_{n,k}}} \frac{1-F_{k_{m^*}}\left(z\right)}{1+z}dz + \frac{1}{\ln 2}\int_0^{\infty}\frac{1-F_{k_n}\left(z\right)}{1+z}dz, \quad (21)$$

where $F_{k_n}\left(z\right)$ and $F_{k_{m^*}}\left(z\right)$ are given in the following equations:

$$F_{k_n}\left(z\right) = 1 - \frac{2\pi\lambda_k}{A_k}\times \int_0^{R_k} x\exp\left[\Lambda\left(x\right) - \frac{\sigma^2 z x^{\alpha_k}}{a_{n,k}P_k\eta} - \Theta\left(\frac{z x^{\alpha_k}}{a_{n,k}P_k\eta}\right)\right]dx, \quad (22)$$

and

$$F_{k_{m^*}}\left(z\right) = 1 - \frac{2\pi\lambda_k}{A_k}\int_0^{R_k} x\exp\left[-\frac{\sigma^2 z {R_k}^{\alpha_k}}{\left(a_{m,k}-a_{n,k}z\right)P_k\eta} - \Theta\left(\frac{z {R_k}^{\alpha_k}}{\left(a_{m,k}-a_{n,k}z\right)P_k\eta}\right) + \Lambda\left(x\right)\right]dx. \quad (23)$$

Here $\Lambda\left(x\right) = -\pi\sum\limits_{i=2}^{K}\lambda_i\left(\frac{P_i B_i x^{\alpha_k}}{P_k B_k}\right)^{\delta_i} - \pi\lambda_1\left(\frac{P_1 G_M x^{\alpha_k}}{N a_{n,k}P_k B_k}\right)^{\delta_1}$ *and $\Theta\left(s\right)$ in* (22) *and* (23) *is given by*

$$\Theta\left(s\right) = \lambda_1\pi\delta_1\sum_{p=1}^{N}\binom{N}{p}\left(s\frac{P_1}{N}\eta\right)^p\left(-s\frac{P_1}{N}\eta\right)^{\delta_1-p} \times B\left(-s\frac{P_1}{N}\eta\left[\omega_{1,k}\left(x\right)\right]^{-\alpha_1}; p-\delta_1, 1-N\right) + s\sum_{i=2}^{K}\frac{\lambda_i 2\pi P_i\eta\left(\omega_{i,k}\left(x\right)\right)^{2-\alpha_i}}{\alpha_i\left(1-\delta_i\right)} \times {}_2F_1\left(1, 1-\delta_i; 2-\delta_i; -sP_i\eta\left(\omega_{i,k}\left(x\right)\right)^{-\alpha_i}\right)\bigg], \quad (24)$$

where $\omega_{1,k}\left(x\right) = \left(\frac{P_1 G_M}{a_{n,k}P_k B_k N}\right)^{\frac{\delta_1}{2}} x^{\frac{\alpha_k}{\alpha_1}}$ *and* $\omega_{i,k}\left(x\right) = \left(\frac{P_i B_i}{P_k B_k}\right)^{\frac{\delta_i}{2}} x^{\frac{\alpha_k}{\alpha_i}}$ *are the nearest distance allowed between the typical user associated to the k-th tier small cell and the macro cell BS, and between the typical user and the i-th tier small cell BS, respectively. $B\left(\cdot;\cdot,\cdot\right)$ and ${}_2F_1\left(\cdot,\cdot;\cdot;\cdot\right)$ are the the incomplete Beta function [15, Eq. (8.319)] and Gauss hypergeometric function [15, Eq. (9.142)], respectively.*

Proof: See Appendix A. ■

Lemma 6. *Conditioned on the HPPPs, the achievable ergodic rate of the k-th tier small cell for the far user case can be expressed as follows:*

$$\tau_k^f = \frac{1}{\ln 2}\int_0^{\infty}\frac{1-F_{k_{n^*}}\left(z\right)}{1+z}dz + \frac{1}{\ln 2}\int_0^{\frac{a_{m,k}}{a_{n,k}}}\frac{1-F_{k_m}\left(z\right)}{1+z}dz, \quad (25)$$

where $F_{k_n}\left(z\right)$ and $F_{k_{m^*}}\left(z\right)$ are given in the following equations:

$$F_{k_m}\left(z\right) = 1 - \frac{2\pi\lambda_k}{A_k}\int_{R_k}^{\infty} x\exp\left[-\frac{\sigma^2 z x^{\alpha_k}}{P_k\eta\left(a_{m,k}-a_{n,k}z\right)} - \Theta\left(\frac{z x^{\alpha_k}}{P_k\eta\left(a_{m,k}-a_{n,k}z\right)}\right) + \Lambda\left(x\right)\right]dx, \quad (26)$$

and

$$F_{k_{n^*}}(z) = 1 - \frac{2\pi\lambda_k}{A_k}\int_{R_k}^{\infty} \times \\ x\exp\left[\Lambda(x) - \frac{\sigma^2 z {R_k}^{\alpha_k}}{P_k\eta a_{n,k}} - \Theta\left(\frac{z{R_k}^{\alpha_k}}{P_k\eta a_{n,k}}\right)\right]dx. \tag{27}$$

Proof: The proof procedure is similar to the approach of obtaining (21)*, which is detailed introduced in Appendix A.* ■

Combining (21) and (25), we obtain a general case for the achievable ergodic rate of the k-th tier small cell as

$$\tau_k = \tau_k^n + \tau_k^f. \tag{28}$$

B. Achievable Ergodic Rate of Macro cells

In massive MIMO aided macro cells, the achievable ergodic rate can be significantly improved due to multiple-antenna array gains, but with undertaking more power consumption and high complexity. In order to evaluate the spectrum efficiency of the whole system, we provide a tractable and tight lower bound of throughput of macro cell derived in as follows:

Theorem 2. *where $f_{d_{o,1}}(x)$ is given in* (10) *and $\omega_{i,1}(x) = \left(\frac{a_{n,i}P_iB_iN}{P_1G_M}\right)^{\frac{\delta_i}{2}} x^{\frac{\alpha_1}{\alpha_i}}$ Here $\omega_{i,1}(x)$ is denoted as the nearest distance allowed between i-th tier small cell BS and the typical user associated to the macro cell.*

Proof: With the aid of Jensen's inequality, we can obtain a tight lower bound of the achievable ergodic rate of the macro cell as

$$\mathbb{E}\left\{\log_2(1+\gamma_{r,1})\right\} \geq \tau_{1,L} \\ = \log_2\left\{1+\exp\left[\underbrace{\mathbb{E}\left\{\ln\left(h_{o,1}L(d_{o,1})P_1/N\right)\right\}}_{Q_1} \\ + \underbrace{\mathbb{E}\left\{\ln\left(I_{M,1}+I_{S,1}+\sigma^2\right)^{-1}\right\}}_{Q_2}\right]\right\}, \tag{30}$$

We first calculate the first part Q_1 as

$$Q_1 = \ln\frac{\eta P_1}{N} + \mathbb{E}\left\{\ln\left(d_{o,1}^{-\alpha_1}\right)\right\} + \mathbb{E}\left\{\ln h_{o,1}\right\}. \tag{31}$$

Then we turn our attention to calculate the last two expectation terms. $\mathbb{E}\left\{\ln\left(d_{o,1}^{-\alpha_1}\right)\right\}$ is given by

$$\mathbb{E}\left\{\ln\left(d_{o,1}^{-\alpha_1}\right)\right\} = -\alpha_1\int_0^{\infty} f_{d_{o,1}}(x)\ln x dx, \tag{32}$$

where $f_{d_{o,1}}(x)$ is given by (10)*.*

Recall that $h_{o,1}$ follows Gamma distribution with parameters $(G_M, 1)$, $\mathbb{E}\left\{\ln h_{o,1}\right\}$ can be expressed as

$$\mathbb{E}\left\{\ln h_{o,1}\right\} = \int_0^{\infty} f_{h_{o,1}}(x)\ln x dx, \tag{33}$$

where $f_{h_{o,1}}(x) = \frac{1}{\Gamma(G_M)}x^{G_M-1}e^{-x}$ is the PDF of $h_{o,1}$. Then with the aid of [15, Eq. (4.352.1)], we can obtain

$$\mathbb{E}\left\{\ln h_{o,1}\right\} = \psi(G_M), \tag{34}$$

where $\psi(\cdot)$ is the digamma function.

We then calculate the second part Q_2. With invoking Jensen's inequality again, we have

$$Q_2 \geq \ln\left(\mathbb{E}\left\{I_{M,1}\right\} + \mathbb{E}\left\{I_{S,1}\right\} + \sigma^2\right)^{-1}, \tag{35}$$

where

$$\mathbb{E}\left\{I_{M,1}\right\} = \int_0^{\infty}\mathbb{E}\left\{I_{M,1}\middle| d_{o,1}=x\right\}f_{d_{o,1}}(x)dx \\ \overset{(a)}{=} \int_0^{\infty}\frac{P_1}{N}\eta\mathbb{E}\left\{h_{\ell,1}\right\}\lambda_1\int_{R^2} r^{-\alpha_1}dr f_{d_{o,1}}(x)dx \\ \overset{(b)}{=} 2\pi P_1\eta\lambda_1\int_0^{\infty}\int_x^{\infty} r^{1-\alpha_1}dr f_{d_{o,1}}(x)dx \\ = \frac{2\pi P_1\eta\lambda_1}{\alpha_1-2}\int_0^{\infty} x^{2-\alpha_1}f_{d_{o,1}}(x)dx. \tag{36}$$

Here step (a) is obtained by applying Campbell's theorem, step (b) is obtained since the expectation of $h_{\ell,1}$ is N.

Then we turn to our attention to the expectation, denote $Q_1(x) = \mathbb{E}\left\{I_{M,1}+I_{S,1}\right\}$, with the aid of Campbell's Theorem, we obtain

$$E\left\{I_{S,1}\right\} = \sum_{i=2}^{K}\left(\frac{2\pi\lambda_i P_i\eta}{\alpha_i-2}\right)\int_0^{\infty}\left[\omega_{i,1}(x)\right]^{2-\alpha_i}f_{d_{o,1}}(x)dx, \tag{37}$$

The proof is complete. Substituting Q_1 and Q_2 into (30)*, we can obtain the desired result.* ■

Proposition 1. *When $\alpha_1 = \alpha_k = \alpha$, the lower bound of achievable ergordic rate of the macro cell is given by in closed-form as* (38)*.*

Proof: When $\alpha_1 = \alpha_k = \alpha$, (8) *can be rewritten as*

$$\tilde{A}_1 = \frac{\lambda_1}{b_1}, \tag{39}$$

where $b_1 = \sum\limits_{i=2}^{K}\lambda_i\left(\frac{a_{n,i}P_iB_iN}{P_1G_M}\right)^{\delta} + \lambda_1$. Then we have

$$\tilde{f}_{d_{o,1}}(x) = 2\pi b_1 x\exp\left(-\pi b_1 x^2\right). \tag{40}$$

Substituting the (40) *into* (29)*, we can obtain the desired closed-form expression as* (1)*. The proof is completed.* ■

Remark 1.

C. Spectrum Efficiency

Based on the analysis of last two subsections, a tractable lower bound of spectrum efficiency is given by

$$\tau_{SE,L} = A_1 N\tau_{1,L} + \sum\nolimits_{k=2}^{K} A_k\tau_k, \tag{41}$$

where $N\tau_1$ and $A_k\tau_k$ are the low bound spectrum efficiency of macro cells and exact spectrum efficiency of the k-th tier small cells, respectively.

$$\tau_{1,L} = N\log_2\left\{1+\exp\left[\ln\frac{\eta P_1}{N} - \alpha_1\int_0^\infty f_{d_{o,1}}(x)\ln x dx + \psi(G_M) - \ln\left(\frac{2\pi P_1\eta\lambda_1}{\alpha_1-2}\int_0^\infty x^{2-\alpha_1} f_{d_{o,1}}(x)\,dx + \sum_{i=2}^{K}\left(\frac{2\pi\lambda_i P_i\eta}{\alpha_i-2}\right)\int_0^\infty [\omega_{i,1}(x)]^{2-\alpha_i} f_{d_{o,1}}(x)\,dx + \sigma^2\right)\right]\right\}. \quad (29)$$

$$\tilde{\tau}_{1,L} = N\log_2\left\{1+\exp\left[1+\ln\frac{\eta P_1}{N} - \frac{\alpha}{2}(\psi(1)-\ln(\pi b_1)) + \psi(G_M) - \ln\left(\frac{2\pi P_1\eta\lambda_1}{\alpha_1-2}\Gamma\left(2-\frac{\alpha}{2}\right)(\pi b_1)^{\frac{\alpha}{2}+1} + \sum_{i=2}^{K}\left(\frac{2\pi\lambda_i P_i\eta}{\alpha_i-2}\right)\left(\frac{a_{n,i}P_iB_iN}{P_1G_M}\right)^{\delta-1}\Gamma\left(2-\frac{\alpha}{2}\right)(\pi b_1)^{\frac{\alpha}{2}+1} + \sigma^2\right)\right]\right\}. \quad (38)$$

D. Energy Efficiency

V. Numerical Results

In this section, numerical results are presented to facilitate the performance evaluations of NOMA in the considered massive MIMO aided K-tier HetNets. The BS density of macro cells is set to be $\lambda_1 = \left(500^2 \times \pi\right)^{-1}$. The considered network is assumed to operate at a carrier frequency of 1 GHz. The noise power is set as $\sigma^2 = -90$ dBm. The path loss exponent for macro cells and k-th tier small cells are $\alpha_1 = 3.5$ and $\alpha_k = 4$, respectively. The power allocation coefficients of NOMA are $a_{m,k} = 0.6$ and $a_{n,k} = 0.4$. The distance between small cells and a existing user is $R_k = 50$ m. Monte Carlo simulations are provided to verify the accuracy of our analysis.

Fig. 2 shows the effect of M and bias factor on user association probability, where the tiers of HetNets are set to be $K = 3$, including macro cells and two tiers small cells. The curves representing macro cells and small cells are from (8) and (9), respectively. One can observe that as the number of antennas at macro BSs increases, more users are likely to associate to macro cells. This is because that the massive MIMO aided macro cells are capable of providing larger array gain, which in turns enhance the average received power for the connected users. Another observation is that increasing the bias factor can encourage more users to connect to the small cells, which is an efficient method to extend the coverage of small cells or control loading balance among each tier of HetNets.

Fig. 3 plots the spectrum efficiency of NOMA based and OMA based small cells versus bias factor with different transmit power of small cell BSs, respectively. The solid curves representing the performance of NOMA based small cells are from (28). We can observe that the spectrum efficiency of small cells decreases as the bias factor increases. This behavior can be explained as follows: larger bias factor makes more macro users with low SINR are associated to small cells, which in turn degrades the spectrum efficiency of small cells. It is also worth noting that the performance of NOMA based small cells outperforms the conventional OMA based small cells, which in turn can enhance the spectrum efficiency of the whole HetNets. What is worth pointing out is that optimizing the power allocation between two NOMA users can further enlarge the performance gap over the OMA based scheme [3],

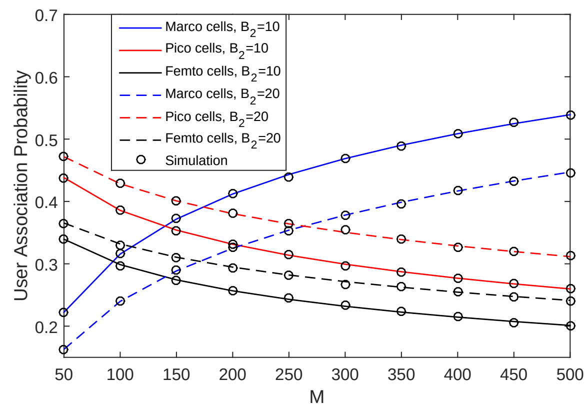

Fig. 2. User association probability of the considered network, with $K = 3$, $N = 15$, $P_1 = 40$ dBm, $P_2 = 30$ dBm and $P_3 = 20$ dBm, $\lambda_2 = \lambda_3 = 20 \times \lambda_1$, and $B_3 = 20B_2$.

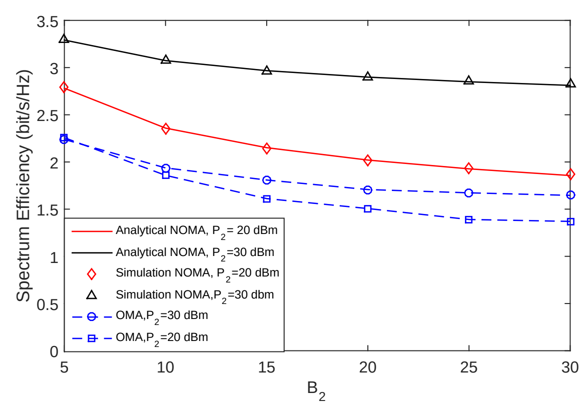

Fig. 3. Comparison of NOMA based and OMA based small cells in terms of spectrum efficiency, with $K = 2$, $M = 200$, $N = 15$, $\lambda_2 = 20 \times \lambda_1$, and $P_1 = 40$ dBm.

which is out of the scope of this paper.

Fig. 4 illustrates the spectrum efficiency versus bias factor with different transmit power of small cell BSs. The curves representing the spectrum efficiency of small cells, macro cells and HetNets are from (28), (**??**) and (41), respectively. We can observe that macro cells can achieve higher spectrum efficiency compared to small cells. This is attributed to the fact

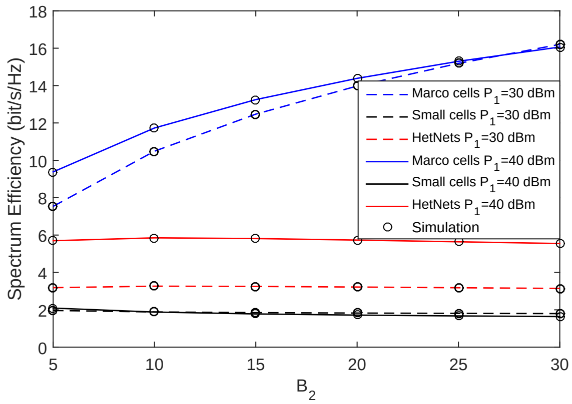

Fig. 4. Spectrum efficiency of the proposed framework, with $K = 2$, $M = 50$, $N = 5$, $P_2 = 20$ dBm, $\lambda_2 = 100 \times \lambda_1$.

that macro BSs are able to serve multiple users simultaneously with offering promising array gains to each user. It is also noted that the spectrum efficiency of macro cells improves as bias factor increases. The reason is again that more low SINR macro cell users are associated to small cells, which in turn makes the spectrum efficiency of macro cells enhance.

VI. Conclusions

In this paper, a novel hybrid massive MIMO and NOMA based HetNets framework has been designed. A flexible massive MIMO and NONA based user association scheme was considered. Stochastic geometry was employed to model the networks and evaluate the corresponding performance. Analytical expressions for spectrum efficiency of the netowrks were derived. It has been demonstrated that NOMA based small cells were able to well-coexist with the current HetNets structure and were capable of achieving higher spectrum efficiency. A promising future direction is to optimize the power allocation among NOMA users to further enhance the spectrum efficiency of the proposed framework.

Appendix A: Proof of Lemma 6

For small cells, the achievable ergodic rate of near user case for the k-th tier can be expressed as

$$\begin{aligned}\tau_k^n &= E\left\{\log_2\left(1+\gamma_{k_{m^*}}\right)+\log_2\left(1+\gamma_{k_n}\right)\right\}\\ &= \frac{1}{\ln 2}\int_0^\infty \frac{1-F_{k_{m^*}}(z)}{1+z}dz + \frac{1}{\ln 2}\int_0^\infty \frac{1-F_{k_n}(z)}{1+z}dz.\end{aligned} \tag{A.1}$$

We first need to obtain the expressions for $F_{k_n}(z)$. Based on (4), we can obtain

$$\begin{aligned}F_{k_n}(z) &= \int_0^{R_k} \Pr\left[\frac{a_{n,k}P_k g_{o,k}\eta x^{-\alpha_k}}{I_{M,k}+I_{S,k}+\sigma^2}\le z\right] f_{d_{o,k}}(x)\,dx\\ &=1-\int_0^{R_k}\exp\left(-\frac{\sigma^2 z x^{\alpha_k}}{a_{n,k}P_k\eta}\right)\mathcal{L}_{I_{k*}}\left(\frac{zx^{\alpha_k}}{a_{n,k}P_k\eta}\right) f_{d_{o,k}}(x)\,dx,\end{aligned} \tag{A.2}$$

where we denote $I_{k*} = I_{M,k} + I_{S,k}$. Then we turn to our attention to the laplace transform of I_{k*} with utilizing it as $\mathcal{L}_{I_{k*}}(s) = \mathcal{L}_{I_{M,k}}(s)\,\mathcal{L}_{I_{S,k}}(s)$. We then derive these two parts in the following:

$$\begin{aligned}\mathcal{L}_{I_{M,k}}(s) &= \mathbb{E}_{I_{M,k}}\left[\exp\left(-s\sum_{\ell\in\Phi_1}\frac{P_1}{N}g_{\ell,1}L\left(d_{\ell,1}\right)\right)\right]\\ &= \mathbb{E}_{\Phi_1}\left[\prod_{\ell\in\Phi_1}\mathbb{E}_{g_{\ell,1}}\left[\exp\left(-s\frac{P_1}{N}g_{\ell,1}\eta d_{\ell,1}^{-\alpha_1}\right)\right]\right]\\ &\overset{(a)}{=}\exp\left(-\lambda_1 2\pi\int_{\omega_{i,1}(x)}^\infty\left(1-\mathbb{E}_{g_{\ell,1}}\left[e^{-\frac{sP_1 g_{\ell,1}\eta}{Nr^{\alpha_1}}}\right]\right)rdr\right),\end{aligned} \tag{A.3}$$

where (a) is obtained with the aid of invoking generating functional of poisson point process (PPP). Recall that the $g_{\ell,1}$ follows Gamma distribution with parameter $(N,1)$. With the aid of Laplace transform for the Gamma distribution, we obtain $E_{g_{\ell,1}}\left[\exp\left(-s\frac{P_1}{N}g_{\ell,1}\eta r^{-\alpha_1}\right)\right] = \mathcal{L}_{g_{\ell,1}}\left(s\frac{P_1}{N}\eta r^{-\alpha_1}\right) = \left(1+s\frac{P_1}{N}\eta r^{-\alpha_1}\right)^{-N}$. As such, we can rewrite (A.3) as

$$\begin{aligned}&\mathcal{L}_{I_{M,k}}(s) =\\ &\exp\left(-\lambda_1 2\pi\int_{\omega_{1,k}(x)}^\infty\left(1-\left(1+\frac{sP_1\eta}{Nr^{\alpha_1}}\right)^{-N}\right)rdr\right).\end{aligned} \tag{A.4}$$

Applying binomial expression and after some mathematical manipulations, we obtain the Laplace transform of $I_{M,k}$ as

$$\begin{aligned}\mathcal{L}_{I_{M,k}}(s) = \exp&\left[-\lambda_1\pi\delta_1\sum_{p=1}^N\binom{N}{p}\left(s\frac{P_1}{N}\eta\right)^p\left(-s\frac{P_1}{N}\eta\right)^{\delta_1-p}\right.\\ &\left.\times B\left(-s\frac{P_1}{N}\eta[\omega_{i,1}(x)]^{-\alpha_1};p-\delta_1,1-N\right)\right],\end{aligned} \tag{A.5}$$

where $B(\cdot;\cdot,\cdot)$ is the is the incomplete Beta function. Following the similar procedure to obtain (A.5), with the aid of [15, Eq. (3.194). 2], we can express $I_{s,k}$ as

$$\begin{aligned}\mathcal{L}_{I_{S,k}}(s) = \exp&\left[-s\sum_{i=2}^K\frac{\lambda_i 2\pi P_i\eta(\omega_{i,k}(x))^{2-\alpha_i}}{\alpha_i(1-\delta_i)}\right.\\ &\left.\times\ {}_2F_1\left(1,1-\delta_i;2-\delta_i;-sP_i\eta(\omega_{i,k}(x))^{-\alpha_i}\right)\right]\end{aligned} \tag{A.6}$$

where ${}_2F_1(\cdot,\cdot;\cdot;\cdot)$ is the is the Gauss hypergeometric function. Then combining (A.5) and (A.6), we can obtain the Laplace transform of I_{k*} as $\mathcal{L}_{I_{k*}}(s) = \exp(-\Theta(s))$, where $\Theta(s)$ is given in (24). Then plugging (11) and $\mathcal{L}_{I_{k*}}(s)$ into (A.2), we obtain the CDF of $F_{k_n}(z)$ in (22). Then we turn to our attention to derive the CDF of $F_{k_{m^*}}(z)$. Based on (5), we can obtain $F_{k_{m^*}}(z)$ as

$$\begin{aligned}F_{k_{m^*}}(z) &= \int_0^{R_k} f_{d_{o,k}}(x)\times\\ &\Pr\left[\left(a_{m,k}-a_{n,k}z\right)g_{o,k}\le\frac{\left(I_{M,k}+I_{S,k}+\sigma^2\right)z}{P_k\eta R_k{}^{-\alpha_k}}\right]dx.\end{aligned} \tag{A.7}$$

Note that for the case $z \ge \frac{a_{m,k}}{a_{n,k}}$, it is easy to observe that $F_{k_{m^*}}(z) = 1$. For the case $z \le \frac{a_{m,k}}{a_{n,k}}$, following the similar procedure of deriving (22), we can obtain the ergodic rate of the existing user for the near user case as (23). The proof is complete.

REFERENCES

[1] Y. Saito, Y. Kishiyama, A. Benjebbour, T. Nakamura, A. Li, and K. Higuchi, "Non-orthogonal multiple access (NOMA) for cellular future radio access," in *Proc. Vehicular Technology Conference (VTC Spring)*, June Dresden, Germany, Jun. 2013, pp. 1–5.

[2] Z. Ding, Z. Yang, P. Fan, and H. V. Poor, "On the performance of non-orthogonal multiple access in 5G systems with randomly deployed users," *IEEE Signal Process. Lett.*, vol. 21, no. 12, pp. 1501–1505, Dec. 2014.

[3] S. Timotheou and I. Krikidis, "Fairness for non-orthogonal multiple access in 5G systems," *IEEE Signal Process. Lett.*, vol. 22, no. 10, pp. 1647–1651, Oct. 2015.

[4] Y. Liu, Z. Ding, M. Elkashlan, and H. V. Poor, "Cooperative non-orthogonal multiple access with simultaneous wireless information and power transfer," *IEEE J. Sel. Areas Commun.*, vol. 34, no. 4, April 2016.

[5] J. G. Andrews, S. Buzzi, W. Choi, S. V. Hanly, A. Lozano, A. C. Soong, and J. C. Zhang, "What will 5G be?" *IEEE J. Sel. Areas Commun.*, vol. 32, no. 6, pp. 1065–1082, 2014.

[6] A. Adhikary, H. S. Dhillon, and G. Caire, "Massive-MIMO meets hetnet: Interference coordination through spatial blanking," *IEEE J. Sel. Areas Commun.*, vol. 33, no. 6, pp. 1171–1186, Jun. 2015.

[7] Q. Ye, O. Y. Bursalioglu, H. C. Papadopoulos, C. Caramanis, and J. G. Andrews, "User association and interference management in massive MIMO hetnets," *arXiv preprint arXiv:1509.07594*, 2015.

[8] W. Liu, S. Jin, C. K. Wen, M. Matthaiou, and X. You, "A tractable approach to uplink spectral efficiency of two-tier massive MIMO cellular hetnets," *IEEE Commun. Lett.*, vol. 20, no. 2, pp. 348–351, Feb. 2016.

[9] A. He, L. Wang, M. Elkashlan, Y. Chen, and K. K. Wong, "Spectrum and energy efficiency in massive MIMO enabled hetnets: A stochastic geometry approach," *IEEE Commun. Lett.*, vol. 19, no. 12, pp. 2294–2297, Dec. 2015.

[10] Z. Ding, Y. Liu, J. Choi, Q. Sun, M. Elkashlan, C.-L. I, and H. V. Poor, "Application of non-orthogonal multiple access in LTE and 5G networks," *IEEE Commun. Mag.*, submitted. [Online]. Available: http://arxiv.org/abs/1511.08610

[11] H. Huh, A. M. Tulino, and G. Caire, "Network MIMO with linear zero-forcing beamforming: Large system analysis, impact of channel estimation, and reduced-complexity scheduling," *IEEE Trans. Inf. Theory*, vol. 58, no. 5, pp. 2911–2934, 2012.

[12] 3rd Generation Partnership Project (3GPP), "Study on downlink multiuser superposition transmission for LTE," Mar. 2015.

[13] 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, 2012.

[14] K. Hosseini, W. Yu, and R. S. Adve, "Large-scale MIMO versus network MIMO for multicell interference mitigation," *IEEE J. Sel. Topics Signal Process.*, vol. 8, no. 5, pp. 930–941, 2014.

[15] I. S. Gradshteyn and I. M. Ryzhik, *Table of Integrals, Series and Products*, 6th ed. New York, NY, USA: Academic Press, 2000.