Chapter 3
Compatibility in NOMA

In this chapter, the compatibility of NOMA will be introduced by discussing the applications of NOMA to various techniques, such as heterogeneous networks (HetNets), cognitive radio networks (CRNs), and multiple-input multiple-output (MIMO). Particularly, the average performance of NOMA enabled HetNets will be provided as an example.

3.1 NOMA in Heterogeneous Networks

HetNets and massive multiple-input multiple-output (MIMO), as two of the "big three" technologies, laid the fundamental structure for future network designs. 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 (Adhikary et al. 2015; Ye et al. 2015; Liu et al. 2016b). In Adhikary et al. (2015), the interference coordination issue of massive MIMO enabled HetNets was addressed by utilizing the spatial blanking of macro cells. In Ye et al. (2015), 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 massive MIMO-aided HetNets was evaluated in Liu et al. (2016b).

Among the recent research contributions towards 5G and the beyond, 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

© The Author(s), under exclusive license to Springer Nature Switzerland AG 2020

Y. Liu et al., *Non-Orthogonal Multiple Access for Massive Connectivity*,
SpringerBriefs in Computer Science, https://doi.org/10.1007/978-3-030-30975-6_3

cells in massive MIMO enabled HetNets—can be a new highly rewarding candidate, which will contribute to the design of a more promising future wireless networks due to the following key advantages:

- High spectrum efficiency: NOMA improves the spectrum efficiency with multiplexing users in power domain and invoking successive 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, 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.

NOMA-based HetNets 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.

3.1.1 Network Model

3.1.1.1 Network Description

Motivated by the aforementioned potential benefits, we 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 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 ($k \in \{1, \ldots, K\}$) 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 channels 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

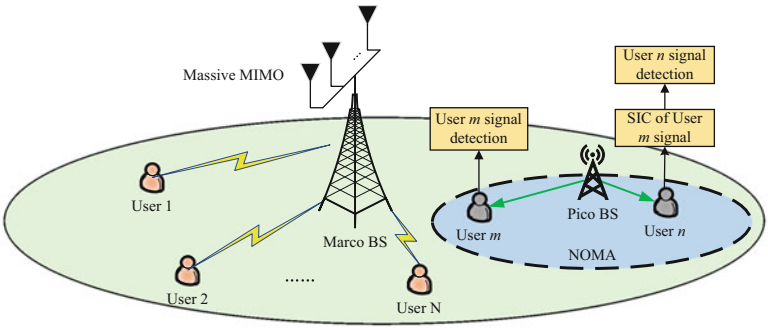

Fig. 3.1 Illustration of NOMA and massive MIMO-based hybrid HetNets

MIMO technologies to macro cells and NOMA transmission to small cells in this work. As shown in Fig. 3.1, 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. 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 (Liu et al. 2016c; Qin et al. 2018b).

3.1.1.2 NOMA and Massive MIMO-Based User Association

In this work, a user is allowed to access any tier BS, which provides the best coverage. We consider the flexible user association which is based on the maximum average received power of each tier.

Different from the convectional user association in OMA, NOMA exploits the power sparsity for multiple access by allocating different powers to different users. Due to the random spatial topology of the stochastic geometry model, the space information of users are not predetermined (Qin et al. 2019). The user association policy for the NOMA enhanced small cells assumes that near user is chosen as the typical one first. As such, at the i-th tier small cell, the averaged power received at users connecting to the i-th tier BS j (where $j \in \Phi_i$) is given by

$$P_{r,i} = a_{n,i} P_i L\left(d_{j,i}\right) B_i, \tag{3.1}$$

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

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

In macro cells, as the macro BS is equipped with multiple antennas, macro cell users experience large array gains. By adopting ZFBF transmission scheme, the array gain obtained at macro users is $G_M = M - N + 1$ (Huh et al. 2012; Hosseini et al. 2014). As a result, the average power received at users connecting to macro BS ℓ (where $\ell \in \Phi_M$) is given by

$$P_{r,1} = G_M P_1 L\left(d_{\ell,1}\right)/N, \tag{3.2}$$

where P_1 is the transmit power of a macro BS, $L\left(d_{\ell,1}\right) = \eta d_{\ell,1}^{-\alpha_1}$ is large-scale path loss, $d_{\ell,1}$ is the distance between the user and a macro BS, η is the frequency dependent factor, and B_i is the identical bias factor which is useful for offloading data traffic in HetNets.

3.1.1.3 Channel Model

In small cells, without loss of generality, we consider that each small cell BS is associated with one user in the previous round of user association process. With applying NOMA protocol, we aim to squeeze a typical user into a same small cell to improve the spectral efficiency. For simplicity, we assume that the distances between the associated users and the connected small cell BSs are the same, which can be arbitrary values and are denoted as r_k; future work will relax this assumption. The distance between atypical user and the connected small cell BS is random. Due to the fact that the path loss is more stable and dominant compared to the instantaneous small-scale fading, we assume that the SIC operation always happened at the near user. We denote that d_{o,k_m} and d_{o,k_n} are the distances from the k-th tier small cell BS to user m and user n, respectively. Since it is not predetermined that atypical user is a near user n or a far user m, we have the following cases.

Near User Case When atypical 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 the connected user m^* to the same BS with the following SINR

$$\gamma_{k_{n\to m*}} = \frac{a_{m,k} P_k g_{o,k} L\left(d_{o,k_n}\right)}{a_{n,k} P_k g_{o,k} L\left(d_{o,k}\right) + I_{M,k} + I_{S,k} + \sigma^2}, \tag{3.3}$$

where $a_{m,k}$ and $a_{n,k}$ are the power sharing coefficients for two users in the k-th layer, σ^2 is the additive white Gaussian noise (AWGN) power, $L\left(d_{o,k_n}\right) = \eta d_{o,k_n}^{-\alpha_i}$ is the large-scale path loss, $I_{M,k} = \sum_{\ell\in\Phi_1} \frac{P_1}{N} g_{\ell,1} L\left(d_{\ell,1}\right)$ is the interference from macro cells, $I_{S,k} = \sum_{i=2}^{K} \sum_{j\in\Phi_i\setminus B_{o,k}} P_i g_{j,i} L\left(d_{j,i}\right)$ is the interference from small cells,

$g_{o,k}$ and d_{o,k_n} refer the small-scale fading coefficients and distance between atypical user and the connected BS in the k-th tier, $g_{\ell,1}$ and $d_{\ell,1}$ refer the small-scale fading coefficients and distance between a typical user and connected BS ℓ in the macro cell, respectively, $g_{j,i}$ and $d_{j,i}$ refer the small-scale fading coefficients and distance between a typical user and its connected BS j except the serving BS $B_{o,k}$ in the i-th tier small cell, respectively. Here, $g_{o,k}$ and $g_{j,i}$ follow exponential distributions with unit mean. $g_{\ell,1}$ follows Gamma distribution with parameters $(N, 1)$.

If the information of user m^* can be decoded successfully, user n then decodes its own message. As such, the SINR at atypical user n, which connects with the k-th tier small cell, can be expressed as

$$\gamma_{k_n} = \frac{a_{n,k} P_k g_{o,k} L\left(d_{o,k_n}\right)}{I_{M,k} + I_{S,k} + \sigma^2}. \tag{3.4}$$

For the connected far user m^* to the same BS, the signal can be decoded by treating the message of user n as interference. Therefore, the SINR that for the connected user m^* to the same BS in the k-th tier small cell can be expressed as

$$\gamma_{k_{m^*}} = \frac{a_{m,k} P_k g_{o,k} L\left(r_k\right)}{I_{k,n} + I_{M,k} + I_{S,k} + \sigma^2}, \tag{3.5}$$

where $I_{k,n} = a_{n,k} P_k g_{o,k} L\left(r_k\right)$, and $L\left(r_k\right) = \eta {r_k}^{-\alpha_k}$.

Far User Case When atypical user is the far user m $(x > r_k)$, we have $d_{o,k_n} = r_k$. As such, for the connected near user n^*, it will first decode the information of user m with the following SINR

$$\gamma_{k_{n^* \to m}} = \frac{a_{m,k} P_k g_{o,k} L\left(r_k\right)}{a_{n,k} P_k g_{o,k} L\left(r_k\right) + I_{M,k} + I_{S,k} + \sigma^2}. \tag{3.6}$$

Once user m is decoded successfully, the interference from atypical user m can be canceled, by applying the SIC technology. Therefore, the SINR at the connected user n^* to the same BS in the k-th tier small cell is given by

$$\gamma_{k_{n^*}} = \frac{a_{n,k} P_k g_{o,k} L\left(r_k\right)}{I_{M,k} + I_{S,k} + \sigma^2}. \tag{3.7}$$

For user m that connects to the k-th tier small cell, the SINR can be expressed as

$$\gamma_{k_m} = \frac{a_{m,k} P_k g_{o,k} L\left(d_{o,k_m}\right)}{I_{k,n^*} + I_{M,k} + I_{S,k} + \sigma^2}, \tag{3.8}$$

where $I_{k,n^*} = a_{n,k} P_k g_{o,k} L\left(d_{o,k_m}\right)$, $L\left(d_{o,k_m}\right) = \eta d_{o,k_m}^{-\alpha_k}$, d_{o,k_n} is the distance between atypical user m and the connected BS in the k-th tier.

Without loss of generality, we assume that a typical user is located at the origin of an infinite two-dimensional plane. Based on (3.1) and (3.2), the SINR at atypical user that connects with a macro BS at a random distance $d_{o,1}$ can be expressed as

$$\gamma_{r,1} = \frac{\frac{P_1}{N} h_{o,1} L\left(d_{o,1}\right)}{I_{M,1} + I_{S,1} + \sigma^2}, \tag{3.9}$$

where $I_{M,1} = \sum_{\ell \in \Phi_1 \setminus B_{o,1}} \frac{P_1}{N} h_{\ell,1} L\left(d_{\ell,1}\right)$ 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, $h_{o,1}$ is the small-scale fading coefficient between atypical user and the connected macro BS, $h_{\ell,1}$ and $d_{\ell,1}$ refer the small-scale fading coefficients and distance between a typical user and the connected macro BS ℓ except the serving BS $B_{o,1}$ in the macro cell, respectively, $h_{j,i}$ and $d_{j,i}$ refer the small-scale fading coefficients and distance between atypical user and connected BS j in the i-th tier small cell, 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.

3.1.2 *Coverage Probability of Non-orthogonal Multiple-Access-Based Small Cells*

In this subsection, we focus our attention on analyzing the coverage probability of a typical user associated with the NOMA enhanced small cells, which is different from the conventional OMA-based small cells due to the channel ordering of two users. The analysis of coverage probability of a typical user associated with the massive MIMO-aided macro cells is the same as the conventional massive MIMO-aided OMA small cells.

3.1.2.1 User Association Probability and Distance Distributions

The user association of the proposed framework is based on maximizing the biased average received power at users. As such, based on (3.1) and (3.2), the user association of macro cells and small cells is given in the following. For simplicity, we denote $\tilde{B}_{ik} = \frac{B_i}{B_k}, \tilde{\alpha}_{ik} = \frac{\alpha_i}{\alpha_k}, \tilde{\alpha}_{1k} = \frac{\alpha_1}{\alpha_k}, \tilde{\alpha}_{i1} = \frac{\alpha_i}{\alpha_1}, \tilde{P}_{1k} = \frac{P_1}{P_k}, \tilde{P}_{i1} = \frac{P_i}{P_1}$, and $\tilde{P}_{ik} = \frac{P_i}{P_k}$ in the following parts of this treatise.

Lemma 3.1 *The user association probability that a typical user connects to NOMA enhanced small cell BSs in the k-th tier and to macro BSs can be calculated as*

$$A_k = 2\pi\lambda_k \int_0^\infty r \exp\left[-\pi \sum_{i=2}^{K} \lambda_i \left(\tilde{P}_{ik} \tilde{B}_{ik}\right)^{\delta_i} r^{\frac{2}{\tilde{\alpha}_{ik}}} - \pi\lambda_1 \left(\frac{\tilde{P}_{1k} G_M}{N a_{n,k} B_k}\right)^{\delta_1} r^{\frac{2}{\tilde{\alpha}_{1k}}}\right] dr., \tag{3.10}$$

and

$$A_1 = 2\pi\lambda_1 \int_0^{\infty} r \exp\left[-\pi \sum_{i=2}^{K} \lambda_i \left(\frac{a_{n,i}\tilde{P}_{i1}B_i N}{G_M}\right)^{\delta_i} r^{\frac{2}{\tilde{\alpha}_{i1}}-\pi\lambda_1 r^2}\right] dr., \tag{3.11}$$

respectively, where $\delta_1 = \frac{2}{\alpha_1}$ *and* $\delta_i = \frac{2}{\alpha_i}$.

Proof *Using the similar method as Lemma 1 of Jo et al. (2012),* (3.10) *and* (3.11) *can be easily obtained.*

Corollary 3.1 *For the special case that each tier has the same path loss exponent, i.e.,* $\alpha_1 = \alpha_k = \alpha$*, the user association probability of the NOMA enhanced small cells in the k-th tier and macro cells can be expressed in closed form as*

$$\tilde{A}_k = \frac{\lambda_k}{\sum_{i=2}^{K} \lambda_i \left(\tilde{P}_{ik}\tilde{B}_{ik}\right)^{\delta} + \lambda_1 \left(\frac{\tilde{P}_{1k}G_M}{N a_{n,k} B_k}\right)^{\delta}}, \tag{3.12}$$

and

$$\tilde{A}_1 = \frac{\lambda_1}{\sum_{i=2}^{K} \lambda_i \left(\frac{a_{n,i}\tilde{P}_{i1}B_i N}{G_M}\right)^{\delta} + \lambda_1}, \tag{3.13}$$

respectively, where $\delta = \frac{2}{\alpha}$.

Remark 3.1 The derived results in (3.12) and (3.13) demonstrate that by increasing the number of antennas at the macro cell BSs, the user association probability of the macro cells increases and the user association probability of the small cells decreases. This is due to the large array gains brought by the macro cells to the served users. It is also worth noting that increasing the power sharing coefficient, a_n, results in higher association probability of small cells. As $a_n \rightarrow 1$, the user association becomes the same as in the conventional OMA-based approach.

Then we consider the probability density function (PDF) of the distance between a typical user and its connected small cell BS in the k-th tier. Based on (3.10), we obtain

$$f_{d_{o,k}}(x) = \frac{2\pi\lambda_k x}{A_k} \exp\left[-\pi \sum_{i=2}^{K} \lambda_i \left(\tilde{P}_{ik}\tilde{B}_{ik}\right)^{\delta_i} x^{\frac{2}{\tilde{\alpha}_{ik}}} - \pi\lambda_1 \left(\frac{\tilde{P}_{1k}G_M}{N a_{n,k} B_k}\right)^{\delta_1} x^{\frac{2}{\tilde{\alpha}_{1k}}}\right]. \tag{3.14}$$

We then calculate the PDF of the distance between a typical user and its connected macro BS. Based on (3.11), we obtain

$$f_{d_{o,1}}(x) = \frac{2\pi\lambda_1 x}{A_1} \exp\left[-\pi \sum_{i=2}^{K} \lambda_i \left(\frac{a_{n,i}\tilde{P}_{i1}B_i N}{G_M}\right)^{\delta_i} x^{\frac{2}{\tilde{\alpha}_{i1}}} - \pi\lambda_1 x^2\right]. \tag{3.15}$$

3.1.2.2 Laplace Transform of Interferences

The next step is to derive the Laplace transform of a typical user. We denote that $I_k = I_{S,k} + I_{M,k}$ is the total interference to the typical user in the k-th tier. The Laplace transform of I_k is $\mathscr{L}_{I_k}(s) = \mathscr{L}_{I_{S,k}}(s)\mathscr{L}_{I_{M,k}}(s)$. We first calculate the Laplace transform of interference from the small cell BS to a typical user $\mathscr{L}_{I_{S,k}}(s)$ in the following lemma.

Lemma 3.2 *The Laplace transform of interferences from the small cell BSs to a typical user can be expressed as*

$$\mathscr{L}_{I_{S,k}}(s) = \exp\left\{-s\sum_{i=2}^{K} \frac{\lambda_i 2\pi P_i \eta\left(\omega_{i,k}(x_0)\right)^{2-\alpha_i}}{\alpha_i(1-\delta_i)} \times {}_2F_1\left(1, 1-\delta_i; 2-\delta_i; -sP_i\eta\left(\omega_{i,k}(x_0)\right)^{-\alpha_i}\right)\right\}, \tag{3.16}$$

where ${}_2F_1(\cdot,\cdot;\cdot;\cdot)$ is the Gauss hypergeometric function (Gradshteyn and Ryzhik 2000, Eq. (9.142)), and $\omega_{i,k}(x_0) = \left(\tilde{B}_{ik}\tilde{P}_{ik}\right)^{\frac{\delta_i}{2}} x_0^{\frac{1}{\tilde{\alpha}_{ik}}}$ is the nearest distance allowed between the typical user and its connected small cell BS in the k-th tier.

Then we calculate the Laplace transform of interference from the macro cell to a typical user $\mathscr{L}_{I_{M,k}}(s)$ in the following lemma.

Lemma 3.3 *The Laplace transform of interference from the macro cell BSs to a typical user can be expressed as*

$$\mathscr{L}_{I_{M,k}}(s) = \exp\left[-\lambda_1\pi\delta_1 \sum_{p=1}^{N} \binom{N}{p} \left(s\frac{P_1}{N}\eta\right)^p \left(-s\frac{P_1}{N}\eta\right)^{\delta_1-p} \times B\left(-s\frac{P_1}{N}\eta\left[\omega_{1,k}(x_0)\right]^{-\alpha_1}; p-\delta_1, 1-N\right)\right], \tag{3.17}$$

where $B(\cdot;\cdot,\cdot)$ is the incomplete Beta function (Gradshteyn and Ryzhik 2000, Eq. (8.319)), and $\omega_{1,k}(x_0) = \left(\frac{\tilde{P}_{1k}G_M}{a_{n,k}B_k N}\right)^{\frac{\delta_1}{2}} x^{\frac{1}{\tilde{\alpha}_{1k}}}$ is the nearest distance allowed between a typical user and its connected BS in the macro cell.

3.1.2.3 Coverage Probability

The coverage probability is defined as that a typical user can successfully transmit signals with targeted data rate R_t. 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 when the following two conditions hold:

1. The typical user can decode the message of the connected user served by the same BS.
2. After the SIC process, the typical user can decode its own message.

As such, the coverage probability of the typical user on the condition of the distance x_0 in the k-th tier is:

$$P_{cov,k}\left(\tau_c, \tau_t, x_0\right)\big|_{x_0 \le r_k} = \Pr\left\{\gamma_{k_{n\to m*}} > \tau_c, \gamma_{k_n} > \tau_t\right\}, \tag{3.18}$$

where $\tau_t = 2^{R_t} - 1$ and $\tau_c = 2^{R_c} - 1$. Here R_c is the targeted data rate of the connected user served by the same BS.

Based on (3.18), for the near user case, we can obtain the expressions for the conditional coverage probability of a typical user in the following lemma.

Lemma 3.4 *If $a_{m,k} - \tau_c a_{n,k} \ge 0$ holds, the conditional coverage probability of a typical user for the near user case is expressed in closed form as*

$$\begin{aligned} P_{cov,k}\left(\tau_c, \tau_t, x_0\right)\big|_{x_0 \le r_k} = \exp\Bigg\{ & -\frac{\varepsilon^*\left(\tau_c, \tau_t\right) x_0^{\alpha_k} \sigma^2}{P_k \eta} \\ & - \lambda_1 \delta_1 \pi \left(\tilde{P}_{1k} \varepsilon^*\left(\tau_c, \tau_t\right)/N\right)^{\delta_1} x_0^{\frac{2}{\tilde{\alpha}_{1k}}} Q_{1,t}^n\left(\tau_c, \tau_t\right) \\ & - \sum_{i=2}^{K} \frac{\lambda_i \delta_i \pi \left(\tilde{B}_{ik}\right)^{\frac{2}{\alpha_i}-1} \left(\tilde{P}_{ik}\right)^{\frac{2}{\alpha_i}} x_0^{\frac{2}{\tilde{\alpha}_{ik}}}}{1-\delta_i} Q_{i,t}^n\left(\tau_c, \tau_t\right) \Bigg\}. \end{aligned} \tag{3.19}$$

Otherwise, $P_{cov,k}\left(\tau_c, \tau_t, x_0\right)\big|_{x_0 \le r_k} = 0$. *Here,* $\varepsilon_t^n = \frac{\tau_t}{a_{n,k}}$, $\varepsilon_c^f = \frac{\tau_c}{a_{m,k} - \tau_c a_{n,k}}$, $\varepsilon^*\left(\tau_c, \tau_t\right) = \max\left\{\varepsilon_c^f, \varepsilon_t^n\right\}$, $Q_{i,t}^n\left(\tau_c, \tau_t\right) = \varepsilon^*\left(\tau_c, \tau_t\right) {}_2F_1\Big(1, 1-\delta_i; 2-\delta_i; -\frac{\varepsilon^*(\tau_c,\tau_t)}{\tilde{B}_{ik}}\Big)$, *and* $Q_{1,t}^n\left(\tau_c, \tau_t\right) = \sum_{p=1}^{N} \binom{N}{p} (-1)^{\delta_1 - p} \times B\Big(-\frac{\varepsilon^*(\tau_c,\tau_t) a_{n,k} B_k}{G_M}; p - \delta_1, 1 - N\Big)$.

Far User Case For the far user case, $x_0 > r_k$, the success decoding will happen if the typical user can decode its own message by treating the connected user served by the same BB as noise. The conditional coverage probability of a typical user for far user case is calculated in the following lemma.

Lemma 3.5 *If $a_{m,k} - \tau_t a_{n,k} \geq 0$ holds, the coverage probability of a typical user for the far user case is expressed in closed form as*

$$P_{cov,k}(\tau_t, x_0)\big|_{x_0>r_k} = \exp\left\{-\frac{\varepsilon_t^f x_0^{\alpha_k}\sigma^2}{P_k\eta} - \lambda_1\delta_1\pi\left(\tilde{P}_{1k}\varepsilon_t^f/N\right)^{\delta_1} x_0^{\frac{2}{\tilde{\alpha}_{1k}}} Q_{1,t}^f(\tau_t) - \sum_{i=2}^{K}\frac{\lambda_i\delta_i\pi\left(\tilde{B}_{ik}\right)^{\frac{2}{\alpha_i}-1}\left(\tilde{P}_{ik}\right)^{\frac{2}{\alpha_i}} x_0^{\frac{2}{\tilde{\alpha}_{ik}}}}{1-\delta_i} Q_{i,t}^f(\tau_t)\right\}. \tag{3.20}$$

Otherwise, $P_{cov,k}(\tau_t, x_0)\big|_{x_0>r_k} = 0$. Here $\varepsilon_t^f = \frac{\tau_t}{a_{m,k}-\tau_t a_{n,k}}$, and

$$Q_{1,t}^f(\tau_t) = \sum_{p=1}^{N}\binom{N}{p}(-1)^{\delta_1-p} B\left(-\frac{\varepsilon_t^f a_{n,k} B_k}{G_M}; p-\delta_1, 1-N\right)$$

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

Based on Lemmas 3.4 and 3.5, we can calculate the coverage probability of a typical user in the following theorem.

Theorem 3.1 *The coverage probability of a typical user associated with the k-th tier small cells is expressed as*

$$P_{cov,k}(\tau_c, \tau_t) = \int_0^{r_k} P_{cov,k}(\tau_c, \tau_t, x_0)\big|_{x_0\leq r_k} f_{d_{o,k}}(x_0)\, dx_0 + \int_{r_k}^{\infty} P_{cov,k}(\tau_t, x_0)\big|_{x_0>r_k} f_{d_{o,k}}(x_0)\, dx_0, \tag{3.21}$$

where $P_{cov,k}(\tau_c, \tau_t, x_0)\big|_{x_0\leq r_k}$ is given in (3.19), $P_{cov,k}(\tau_t, x_0)\big|_{x_0>r_k}$ is given in (3.20), and $f_{d_{o,k}}(x_0)$ is given in (3.14).

Although (3.21) has provided the exact analytical expression for the coverage probability of typical user, it is hard to directly obtain insights from this expression. Driven by this, we provide one special case with considering that each tier is with the same path loss exponents. As such, we have $\tilde{\alpha}_{1k} = \tilde{\alpha}_{ik} = 1$. In addition, we consider the interference limited case, where the thermal noise can be neglected. Then based on (3.21), we can obtain the closed-form coverage probability of a typical user in the following corollary.

Corollary 3.2 *With $\alpha_1 = \alpha_k = \alpha$ and $\sigma^2 = 0$, the coverage probability of a typical user can be expressed in closed form as follows:*

$$\tilde{P}_{cov,k}(\tau_c, \tau_t) = \frac{b_k\left(1 - e^{-\pi\left(b_k + c_1^n(\tau_c,\tau_t) + c_2^n(\tau_c,\tau_t)\right)r_k^2}\right)}{b_k + c_1^n(\tau_c, \tau_t) + c_2^n(\tau_c, \tau_t)} + \frac{b_k e^{-\pi\left(b_k + c_1^f(\tau_t) + c_2^f(\tau_t)\right)r_k^2}}{b_k + c_1^f(\tau_t) + c_2^f(\tau_t)}, \tag{3.22}$$

where $b_k = \sum_{i=2}^{K} \lambda_i \left(\tilde{P}_{ik}\tilde{B}_{ik}\right)^{\delta} + \lambda_1\left(\frac{\tilde{P}_{1k}G_M}{Na_{n,k}B_k}\right)^{\delta}$, $c_1^n(\tau_c, \tau_t) = \lambda_1\delta_1\left(\frac{\tilde{P}_{1k}\varepsilon^*(\tau_c,\tau_t)}{N}\right)^{\delta}\tilde{Q}_{1,t}^n(\tau_c, \tau_t)$, $c_2^n(\tau_c, \tau_t) = \sum_{i=2}^{K} \frac{\lambda_i\delta_i\left(\tilde{B}_{ik}\right)^{\frac{2}{\alpha}-1}\left(\tilde{P}_{ik}\right)^{\frac{2}{\alpha}}}{1-\delta_i}\tilde{Q}_{i,t}^n(\tau_c, \tau_t)$, $c_1^f(\tau_t) = \lambda_1\delta_1\left(\frac{\tilde{P}_{1k}\varepsilon_t^f}{N}\right)^{\delta_1}\tilde{Q}_{1,t}^f(\tau_t)$, *and* $c_2^f(\tau_t) = \sum_{i=2}^{K} \frac{\lambda_i\delta_i\left(\tilde{B}_{ik}\right)^{\frac{2}{\alpha}-1}\left(\tilde{P}_{ik}\right)^{\frac{2}{\alpha}}}{1-\delta}\tilde{Q}_{i,t}^f(\tau_t)$. *Here,* $\tilde{Q}_{1,t}^n(\tau_c, \tau_t)$, $\tilde{Q}_{i,t}^n(\tau_c, \tau_t)$, $\tilde{Q}_{1,t}^f(\tau_t)$, *and* $\tilde{Q}_{i,t}^f(\tau_t)$ *are based on interchanging the same path loss exponents, i.e.,* $\alpha_1 = \alpha_k = \alpha$, *for each tier from* $Q_{1,t}^n(\tau_c, \tau_t)$, $Q_{i,t}^n(\tau_c, \tau_t)$, $Q_{1,t}^f(\tau_t)$, *and* $Q_{i,t}^f(\tau_t)$.

Remark 3.2 The derived results in (3.22) demonstrate that the coverage probability of a typical user is determined by the target rate of itself as well as the target rate of the connected user. Additionally, inappropriate power allocation such as, $a_{m,k} - \tau_t a_{n,k} < 0$, will lead to the coverage probability always being zero.

3.1.3 Spectrum Efficiency

To evaluate the spectrum efficiency of the proposed NOMA enhanced hybrid HetNets framework, we calculate the spectrum efficiency of each tier in this section.

3.1.3.1 Ergodic Rate of NOMA Enhanced Small Cells

Different from calculating the coverage probability of the case with fixed targeted rate, the achievable ergodic rate for NOMA enhanced small cells is opportunistically determined by the channel conditions of users. It is also easy to verify that if the far user can decode the message of itself, the near user can definitely decode the message of far user since it has a better channel condition (Ding et al. 2014). Recall that the distance order between the connected BS and the two users are not predetermined, as such, we calculate the achievable ergodic rate of small cells for both the near user case and far user case in the following lemmas.

Lemma 3.6 *The achievable ergodic rate of the k-th tier small cell for the near user case can be expressed as follows:*

$$\tau_k^n = \frac{2\pi\lambda_k}{A_k \ln 2}\left[\int_0^{\frac{a_{m,k}}{a_{n,k}}} \frac{\bar{F}_{\gamma_{k_{m*}}}(z)}{1+z}dz + \int_0^{\infty} \frac{\bar{F}_{\gamma_{k_n}}(z)}{1+z}dz\right], \tag{3.23}$$

where $\bar{F}_{\gamma_{k_{m*}}}(z)$ *and* $\bar{F}_{\gamma_{k_n}}(z)$ *are given by*

$$\bar{F}_{\gamma_{k_{m*}}}(z) = \int_0^{r_k} x \exp\left[-\frac{\sigma^2 z {r_k}^{\alpha_k}}{\left(a_{m,k} - a_{n,k}z\right)P_k\eta} - \Theta\left(\frac{z{r_k}^{\alpha_k}}{\left(a_{m,k} - a_{n,k}z\right)P_k\eta}\right) + \Lambda(x)\right]dx, \tag{3.24}$$

and

$$\bar{F}_{\gamma_{k_n}}(z) = \int_0^{r_k} x \exp\left[\Lambda(x) - \frac{\sigma^2 z x^{\alpha_k}}{a_{n,k}P_k\eta} - \Theta\left(\frac{zx^{\alpha_k}}{a_{n,k}P_k\eta}\right)\right]dx. \tag{3.25}$$

Here $\Lambda(x) = -\pi\sum_{i=2}^{K}\lambda_i\left(\tilde{P}_{ik}\tilde{B}_{ik}\right)^{\delta_i} x^{\frac{2}{\tilde{\alpha}_{ik}}} - \pi\lambda_1\left(\frac{\tilde{P}_{1k}G_M}{Na_{n,k}B_k}\right)^{\delta_1} x^{\frac{2}{\tilde{\alpha}_{1k}}}$ *and* $\Theta(s)$ *is given by*

$$\begin{aligned}\Theta(s) = {} & \lambda_1\pi\delta_1\sum_{p=1}^{N}\binom{N}{p}\left(s\frac{P_1}{N}\eta\right)^p\left(-s\frac{P_1}{N}\eta\right)^{\delta_1-p} \\ & \times B\left(-s\frac{P_1}{N}\eta\left[\omega_{1,k}(x)\right]^{-\alpha_1}; p-\delta_1, 1-N\right) \\ & + s\sum_{i=2}^{K}\frac{\lambda_i 2\pi P_i\eta\left(\omega_{i,k}(x)\right)^{2-\alpha_i}}{\alpha_i(1-\delta_i)} \\ & \times {}_2F_1\left(1, 1-\delta_i; 2-\delta_i; -sP_i\eta\left(\omega_{i,k}(x)\right)^{-\alpha_i}\right).\end{aligned} \tag{3.26}$$

Lemma 3.7 *The achievable ergodic rate of the k-th tier small cell for the far user case can be expressed as follows:*

$$\tau_k^f = \frac{2\pi\lambda_k}{A_k \ln 2}\left[\int_0^{\infty} \frac{\bar{F}_{\gamma_{k_{n*}}}(z)}{1+z}dz + \int_0^{\frac{a_{m,k}}{a_{n,k}}} \frac{\bar{F}_{\gamma_{k_m}}(z)}{1+z}dz\right], \tag{3.27}$$

where $\bar{F}_{\gamma_{k_m}}(z)$ *and* $\bar{F}_{\gamma_{k_{n*}}}(z)$ *are given by*

$$\bar{F}_{\gamma_{k_m}}(z) = \int_{r_k}^{\infty} x \exp\left[-\frac{\sigma^2 z x^{\alpha_k}}{P_k \eta \left(a_{m,k} - a_{n,k} z\right)} - \Theta\left(\frac{z x^{\alpha_k}}{P_k \eta \left(a_{m,k} - a_{n,k} z\right)}\right) + \Lambda(x) \right] dx, \tag{3.28}$$

and

$$\bar{F}_{\gamma_{k_{n*}}}(z) = \int_{r_k}^{\infty} x \exp\left[\Lambda(x) - \frac{\sigma^2 z {r_k}^{\alpha_k}}{P_k \eta a_{n,k}} - \Theta\left(\frac{z {r_k}^{\alpha_k}}{P_k \eta a_{n,k}}\right) \right] dx. \tag{3.29}$$

Theorem 3.2 *Conditioned on the HPPPs, the achievable ergodic rate of the small cells can be expressed as follows:*

$$\tau_k = \tau_k^n + \tau_k^f, \tag{3.30}$$

where τ_k^n *and* τ_k^f *are obtained from* (3.23) *and* (3.27).

Note that the derived results in (3.30) is a double integral form, since even for some special cases, it is challenging to obtain closed form solutions. However, the derived expression is still much more efficient and also more accurate compared to using the approach of Monte Carlo simulations, which highly depends on the repeated iterations of random sampling.

3.1.3.2 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 more power consumption and high complexity. However, the exact analytical results require high order derivatives of Laplace transform. When the number of antennas goes large, it becomes mathematical intractable to calculate the derivatives due to the unacceptable complexity. In order to evaluate the spectrum efficiency of the whole system, we provide a tractable lower bound of throughput for macro cells in the following theorem.

Theorem 3.3 *The lower bound of achievable ergodic rate of the macro cells can be expressed as follows:*

$$\tau_{1,L} = \log_2\left(1 + \frac{P_1 G_M \eta}{N \int_0^{\infty} \left(Q_1(x) + \sigma^2\right) x^{\alpha_1} f_{d_{o,1}}(x)\, dx}\right), \tag{3.31}$$

where $f_{d_{o,1}}(x)$ *is given in* (3.15), $Q_1(x) = \frac{2P_1\eta\pi\lambda_1}{\alpha_1-2}x^{2-\alpha_1} + \sum_{i=2}^{K} 2\pi\lambda_i \left(\frac{P_i\eta}{\alpha_i-2}\right) \left[\omega_{i,1}(x)\right]^{2-\alpha_i}$, *and* $\omega_{i,1}(x) = \left(\frac{a_{n,i}\tilde{P}_{i1}B_iN}{G_M}\right)^{\frac{\delta_i}{2}} x^{\frac{1}{\alpha_{i1}}}$ *is denoted as the nearest distance allowed between the* i*-th tier small cell BS and the typical user that is associated with the macro cell.*

Corollary 3.3 *If* $\alpha_1 = \alpha_k = \alpha$ *holds, the lower bound of achievable ergordic rate of the macro cell is given by in closed form as*

$$\tilde{\tau}_{1,L} = \log_2\left(1 + \frac{P_1G_M\eta/N}{\psi(\pi b_1)^{-1} + \sigma^2\Gamma\left(\frac{\alpha}{2}+1\right)(\pi b_1)^{-\frac{\alpha}{2}}}\right), \tag{3.32}$$

where $\psi = \frac{2P_1\eta\pi\lambda_1}{\alpha-2} + \sum_{i=2}^{K}\left(\frac{2\pi\lambda_i P_i\eta}{\alpha-2}\right)\left(\frac{a_{n,i}\tilde{P}_{i1}B_iN}{G_M}\right)^{\delta-1}$ *and* $b_1 = \sum_{i=2}^{K}\lambda_i\left(\frac{a_{n,i}\tilde{P}_{i1}B_iN}{G_M}\right)^{\delta} + \lambda_1$.

Remark 3.3 The derived results in (3.32) demonstrate that achievable ergordic rate of the macro cell can be enhanced by increasing the number of antennas at the macro cell BSs. This is because the users in macro cells can experience larger array gains.

3.1.3.3 Spectrum Efficiency of the Proposed Hybrid Hetnets

Based on the analysis of last two subsections, a tractable lower bound of spectrum efficiency can be given in the following proposition.

Proposition 3.1 *The spectrum efficiency of the proposed hybrid Hetnets is*

$$\tau_{\mathrm{SE,L}} = A_1N\tau_{1,L} + \sum_{k=2}^{K} A_k\tau_k, \tag{3.33}$$

where $N\tau_1$ *and* τ_k *are the low bound spectrum efficiency of macro cells and exact spectrum efficiency of the* k*-th tier small cells, respectively. Here,* A_k *and* A_1 *are obtained from* (3.10) *and* (3.11)*, and* τ_k *and* $\tau_{1,L}$ *are obtained from* (3.30) *and* (3.31)*, respectively.*

3.1.4 Energy Efficiency

In this section, we proceed to investigate the performance of the proposed hybrid HetNets framework from the perspective of energy efficiency, due to the fact that energy efficiency is an important performance metric in the 5G systems.

3.1.4.1 Power Consumption Model

To calculate the energy efficiency of the considered networks, we first need to model the power consumption parameter of both small cell BSs and macro cell BSs. The power consumption of small cell BSs is given by

$$P_{i,total} = P_{i,static} + \frac{P_i}{\varepsilon_i}, \tag{3.34}$$

where $P_{i,static}$ is the static hardware power consumption of small cell BSs in the i-th tier, and ε_i is the efficiency factor for the power amplifier of small cell BSs in the i-th tier.

The power consumption of macro cell BSs is given by

$$P_{1,total} = P_{1,static} + \sum_{a=1}^{3} \left(N^a \Delta_{a,0} + N^{a-1} M \Delta_{a,1} \right) + \frac{P_1}{\varepsilon_1}, \tag{3.35}$$

where $P_{1,static}$ is the static hardware power consumption of macro cell BSs, ε_1 is the efficiency factor for the power amplifier of macro cell BSs, and $\Delta_{a,0}$ and $\Delta_{a,1}$ are the practical parameters which depended on the chains of transceivers, precoding, coding/decoding, etc.

3.1.4.2 Energy Efficiency of NOMA Enhanced Small Cells and Macro Cells

The energy efficiency is defined as

$$\Theta_{\text{EE}} = \frac{\text{Total data rate}}{\text{Total energy consumption}}. \tag{3.36}$$

Therefore, based on (3.36) and the power consumption model for small cells that we have provided in (3.34), the energy efficiency of the k-th tier of NOMA enhanced small cells is expressed as

$$\Theta_{\text{EE}}^{k} = \frac{\tau_k}{P_{k,total}}, \tag{3.37}$$

where τ_k is obtained from (3.30).

Based on (3.35) and (3.36), the energy efficiency of macro cell is expressed as

$$\Theta_{\text{EE}}^{1} = \frac{N \tau_{1,L}}{P_{1,total}}, \tag{3.38}$$

where $\tau_{1,L}$ is obtained from (3.31).

3.1.4.3 Energy Efficiency of the Proposed Hybrid Hetnets

According to the derived results of energy efficiency of NOMA enhanced small cells and macro cells, we can express the energy efficiency in the following proposition.

Proposition 3.2 *The energy efficiency of the proposed hybrid Hetnets is as follows:*

$$\Theta_{\mathrm{EE}}^{\mathrm{Hetnets}} = A_1 \Theta_{\mathrm{EE}}^{1} + \sum\nolimits_{k=2}^{K} A_k \Theta_{\mathrm{EE}}^{k}, \tag{3.39}$$

where A_k *and* A_1 *are obtained from* (3.10) *and* (3.11)*, and* Θ_{EE}^{k} *and* Θ_{EE}^{1} *are obtained from* (3.37) *and* (3.38)*.*

3.1.5 Numerical Results

In this section, numerical results are presented to facilitate the performance evaluations of NOMA enhanced hybrid K-tier HetNets. The noise power is $\sigma^2 = -170 + 10 \times \log_{10}(BW) + N_f$. The power sharing coefficients of NOMA for each tier are same as $a_{m,k} = a_m$ and $a_{n,k} = a_n$ for simplicity. BPCU is short for bit per channel use. Monte Carlo simulations marked as "○" are provided to verify the accuracy of our analysis. Table 3.1 summarizes the simulation parameters used in this section.

3.1.5.1 User Association Probability and Coverage Probability

Figure 3.2 shows the effect of number of antennas equipped at each macro BS, M, and bias factor on user association probability, where the tiers of HetNets are set to

Table 3.1 Table of parameters

Monte Carlo simulations repeated	10^5 times
The radius of the plane	10^4 m
Carrier frequency	1 GHz
The BS density of macro cells	$\lambda_1 = \left(500^2 \times \pi\right)^{-1}$
Pass loss exponent	$\alpha_1 = 3.5, \alpha_k = 4$
The noise figure	$N_f = 10$ dB
The noise power	$\sigma^2 = -90$ dBm
Static hardware power consumption	$P_{1,total} = 4$ W, $P_{i,total} = 2$ W
Power amplifier efficiency factor	$\varepsilon_1 = \varepsilon_i = 0.4$
Precoding power consumption	$\Delta_{1,0} = 4.8, \Delta_{2,0} = 0$
	$\Delta_{3,0} = 2.08 \times 10^{-8}$
	$\Delta_{1,1} = 1, \Delta_{2,1} = 9.5 \times 10^{-8}$
	$\Delta_{3,1} = 6.25 \times 10^{-8}$

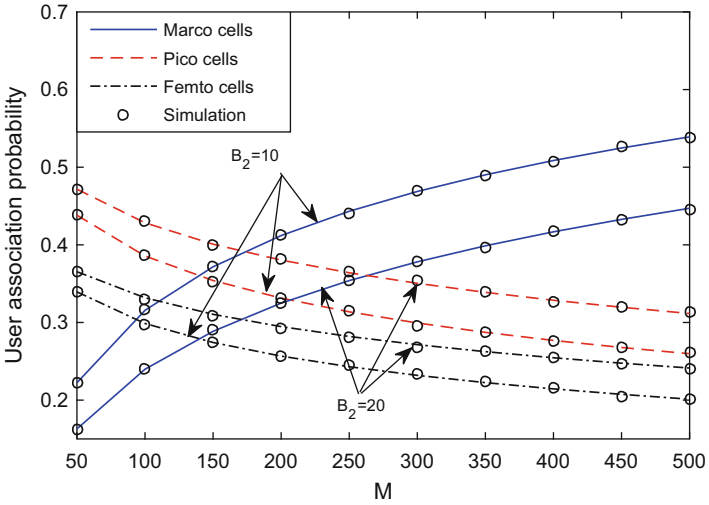

Fig. 3.2 User association probability versus antenna number with different bias factor, with $K = 3$, $N = 15$, $P_1 = 40\,\text{dBm}$, $P_2 = 30\,\text{dBm}$ and $P_3 = 20\,\text{dBm}$, $r_k = 50\,\text{m}$, $a_m = 0.6$, $a_n = 0.4$, $\lambda_2 = \lambda_3 = 20 \times \lambda_1$, and $B_3 = 20 \times B_2$

be $K = 3$, including macro cells and two tiers of small cells. The analytical curves representing small cells and macro cells are from (3.10) and (3.11), respectively. One can observe that as the number of antennas at each macro BS increases, more users are likely to associate with macro cells. This is because that the massive MIMO-aided macro cells are capable of providing larger array gain, which in turn enhance the average received power for the connected users. This observation is consistent with Remark 3.1. 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.

Figure 3.3 plots the coverage probability of a typical user associated with the k-tier NOMA enhanced small cells versus bias factor. The solid curves representing the analytical results of NOMA are from (3.21). One can observe that the coverage probability decreases as bias factor increases, which means that the unbiased user association outperforms the biased one, i.e., when $B_2 = 1$, the scenario becomes unbiased user association. This is because by invoking biased user association, users cannot be always associated with the BS which provides the highest received power. But the biased user association is capable of offering more flexibility for users as well as the whole networks, especially for the case that cells are fully overload. We also demonstrate that NOMA has superior behavior over OMA scheme.[2]

[2]The OMA benchmark adopted in this treatise is that by dividing the two users in equal time/frequency slots.

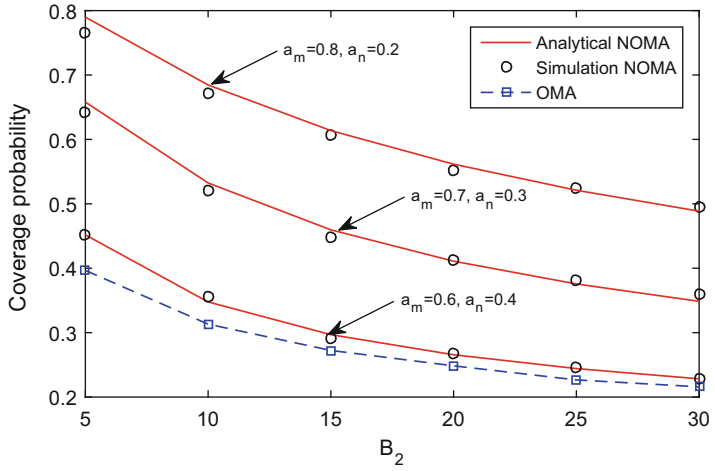

Fig. 3.3 Coverage probability comparison of NOMA- and OMA-based small cells. $K = 2$, $M = 200$, $N = 15$, $\lambda_2 = 20 \times \lambda_1$, $R_t = R_c = 1$ BPCU, $r_k = 10$ m, $P_1 = 40$ dBm, and $P_2 = 20$ dBm

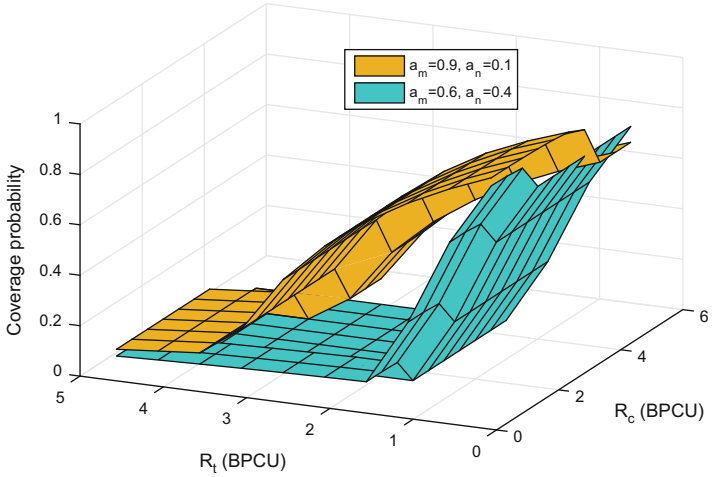

Fig. 3.4 Successful probability of typical user versus targeted rates of R_t and R_c, with $K = 2$, $M = 200$, $N = 15$, $\lambda_2 = 20 \times \lambda_1$, $r_k = 15$ m, $B_2 = 5$, $P_1 = 40$ dBm, and $P_2 = 20$ dBm

Figure 3.4 plots the coverage probability of a typical user associated with the k-tier NOMA enhanced small cells versus both R_t and R_c. We observe that there is a cross between these two plotted surfaces, which means that there exists an optimal power sharing allocation scheme for the given targeted rate. In contrast, for fixed power sharing coefficients, e.g., $a_m = 0.9, a_n = 0.1$, there also exist optimal targeted rates of two users for coverage probability. This figure also illustrates that for inappropriate power and targeted rate selection, the coverage probability is always zero, which also verifies our obtained insights in Remark 3.2.

3.1.5.2 Spectrum Efficiency

Figure 3.5 plots the spectrum efficiency of small cells with NOMA and OMA versus bias factor, B_2, with different transmit power of small cell BSs, P_2. The curves representing the performance of NOMA enhanced small cells are from (3.30). The performance of conventional OMA-based small cells is illustrated as a benchmark to demonstrate the effectiveness of our proposed framework. We observe that the spectrum efficiency of small cells decreases as the bias factor increases. This behavior can be explained as follows: larger bias factor associates more macro users with low SINR to small cells, which in turn degrades the spectrum efficiency of small cells. It is also worth noting that the performance of NOMA enhanced small cells outperforms the conventional OMA-based small cells, which in turn enhances the spectrum efficiency of the whole HetNets.

Figure 3.6 plots the spectrum efficiency of the proposed whole HetNets versus bias factor, B_2, with different transmit power, P_1. The curves representing the spectrum efficiency of small cells, macro cells, and HetNets are from (3.33). We can observe that macro cells can achieve higher spectrum efficiency compared to small cells. This is attributed to the fact that macro BSs are able to serve multiple users simultaneously with offering promising array gains to each user, which has been analytically demonstrated in Remark 3.3. It is also noted that the spectrum efficiency of macro cells improves as bias factor increases. The reason is again that when more low SINR macro cell users are associated with small cells, the spectrum efficiency of macro cells can be enhanced.

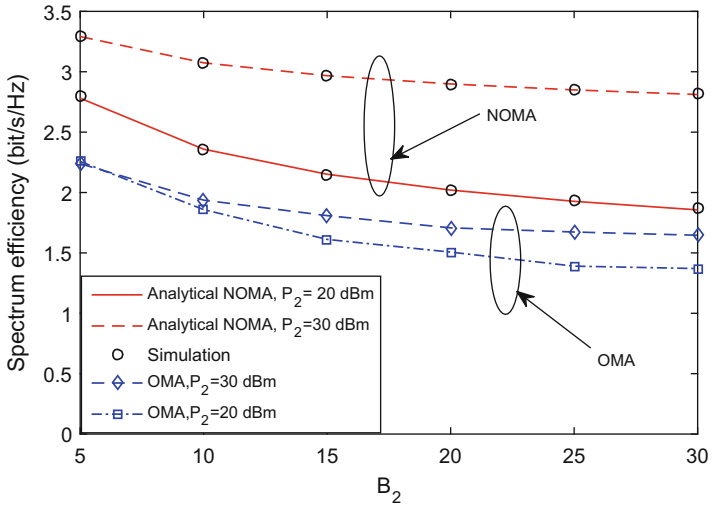

Fig. 3.5 Spectrum efficiency comparison of NOMA- and OMA-based small cells. $K = 2$, $M = 200$, $N = 15$, $r_k = 50$ m, $a_m = 0.6$, $a_n = 0.4$, $\lambda_2 = 20 \times \lambda_1$, and $P_1 = 40$ dBm

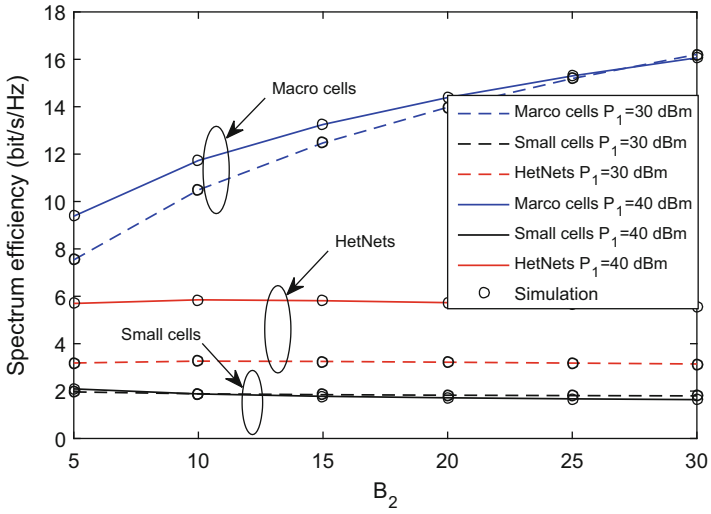

Fig. 3.6 Spectrum efficiency of the proposed framework. $r_k = 50$ m, $a_m = 0.6$, $a_n = 0.4$, $K = 2$, $M = 50$, $N = 5$, $P_2 = 20$ dBm, and $\lambda_2 = 100 \times \lambda_1$

3.1.5.3 Energy Efficiency

Figure 3.7 plots the energy efficiency of the proposed whole HetNets versus bias factor, B_2, with different transmit antenna of macro cell BSs, M. Several observations are as follows: (1) One observation is that the energy efficiency of the macro cells decreases as the number of antenna increases. Enlarging the number of antenna at the macro BSs is capable of offering a larger array gain, which in turn enhances the spectrum efficiency. Such operations also bring significant power consumption from the baseband signal processing of massive MIMO, which results in decreased energy efficiency. (2) Another observation is that NOMA enhanced small cells can achieve higher energy efficiency than the massive MIMO-aided macro cells. It means that from the perspective of energy consumption, densely deploying BSs in NOMA enhanced small cell is a more effective approach. (3) It is also worth noting that the number of antennas at the macro cell BSs almost has no effect on the energy efficiency of the NOMA enhanced small cells. (4) It also demonstrates that NOMA enhanced small cells have superior behavior than conventional OMA-based small cells in terms of energy efficiency. Such observations above demonstrate the benefits of the proposed NOMA enhanced hybrid HetNets and provide insightful guidelines for designing the practical large-scale networks.

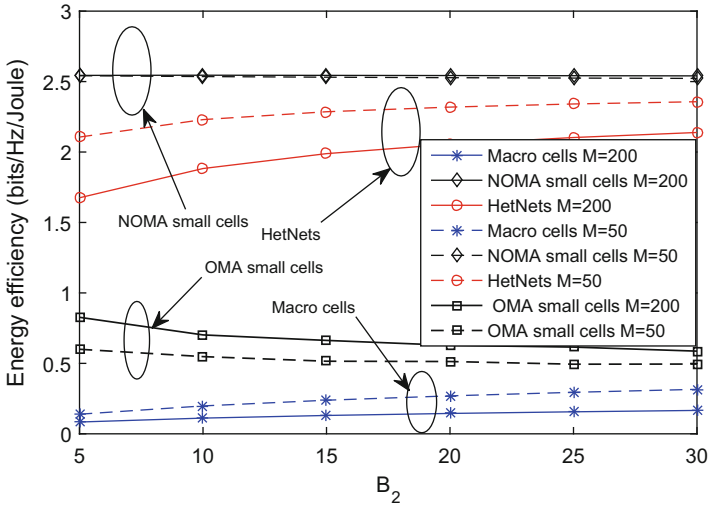

Fig. 3.7 Energy efficiency of the proposed framework. $K = 2$, $r_k = 10$ m, $a_m = 0.6$, $a_n = 0.4$, $N = 15$, $P_1 = 30$ dBm, $P_2 = 20$ dBm, and $\lambda_2 = 20 \times \lambda_1$

3.2 NOMA in Cognitive Radio Networks

The 2010s have witnessed the rapidly increasing penetration of mobile devices (e.g., smart phones, tablets, and laptops) all over the world, which gave rise to increasing demand for spectral resources. As reported by the Federal Communications Commission (FCC), there are significant temporal and spatial variations in the exploitation of the allocated spectrum. Given this fact, the CR concept inspired the community to mitigate the spectrum scarcity problem. The basic concept of CR is that at a certain time of the day or in a geographic region, the unlicensed secondary users (SUs) are allowed to opportunistically access the licensed spectrum of primary users (PUs). These CR techniques may be categorized into the interweave, overlay, and underlay paradigms:

- **Interweave**: The interweave CR can be regarded as an interference avoidance paradigm, where the SUs are required to sense the temporary slivers of the space-frequency domain of PUs before they access the channels (Qin et al. 2016a,b, 2018a). The concurrent transmission of SUs and PUs is not allowed under the interweave paradigm.
- **Overlay**: The overlay paradigm essentially constitutes an interference mitigation technique. With the aid of the classic dirty paper encoding technique, overlay CR ensures that a cognitive user becomes capable of transmitting simultaneously with a noncognitive PU (Goldsmith et al. 2009). Additionally, SUs are capable of forwarding the information of PUs to the PU receivers, while superimposing their own signals as a reward for their relaying services.

- **Underlay**: The underlay CR operates like an intelligent interference control paradigm, where the SUs are permitted to access the spectrum allocated to PUs as long as the interference power constraint at the PUs is satisfied.

One of the core challenges in both CR and NOMA networks is the interference management, while improving the bandwidth efficiency. Hence it is natural to link them for achieving an improved bandwidth efficiency. The application of NOMA in large-scale underlay CR networks has been investigated by using the stochastic geometry model (Liu et al. 2016d). The diversity order of the NOMA users was characterized analytically in two scenarios. The classic OMA-based underlay CR was also used as a benchmark to show the benefits of the proposed CR-NOMA scheme. Ding et al. (2016c) has proposed a novel power allocation (PA) policy for NOMA, namely the CR-inspired NOMA PA, which constitutes a beneficial amalgam of NOMA and underlay CR.

To the best of our knowledge, CR-NOMA studies only exist in the context of the underlay CR paradigm. Hence both the interweave and overlay CR paradigms have to be investigated in NOMA networks. It is worth pointing out that a significant research challenge of NOMA is to dynamically cluster/pair the NOMA users first, followed by dynamically allocating the clusters/pairs to different orthogonal sub-channels. In the context of the interweave paradigm, intelligent sensing has to be applied first, followed by user clustering/pairing of NOMA users, depending on the specific channel conditions sensed.

3.3 NOMA with MIMO

Multiple-antenna techniques are of significant importance, since they offer the extra dimension of the spatial domain, for further performance improvements. The application of multiple-antenna techniques in NOMA has attached substantial interest both from academia (Ding et al. 2016b,d; Hanif et al. 2016; Choi 2016; Kim et al. 2013; Qureshi et al. 2016; Liu et al. 2016a; Chen et al. 2016) and from industry (Higuchi and Kishiyama 2013; Higuchi and Benjebbour 2015; Benjebbour et al. 2013; Saito et al. 2013). The distinct NOMA features such as channel ordering and PA inevitably require special attention in the context of multiple antennas. More specifically, in contrast to the SISO-NOMA scenarios whose channels are all scalars, the channels of MIMO-NOMA scenarios are represented in form of matrices, which makes the power-based ordering of users rather challenging. As a consequence, conceiving an appropriate beamforming/precoding design is essential for multi-antenna-aided NOMA systems. NOMA relying on beamforming (BF) constitutes an efficient technique of improving the bandwidth efficiency by exploiting both the power domain and the angular domain. There are two popular MIMO-NOMA designs, namely the (1) Cluster-based (CB) MIMO-NOMA design; and the (2) Beamformer-based (BB) MIMO-NOMA design, which will be introduced in the following.

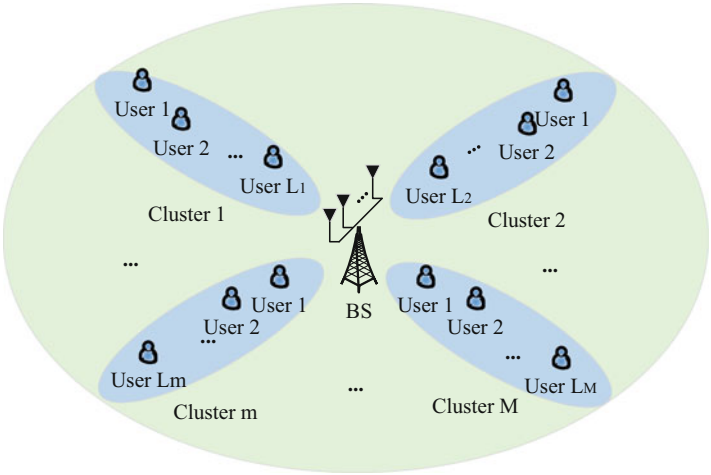

Fig. 3.8 Illustration of cluster-based MIMO-NOMA

3.3.1 Cluster-Based MIMO-NOMA

One of the popular NOMA designs is associated with the cluster-based structure, partitioning users into several different clusters. Explicitly, as shown in Fig. 3.8, the NOMA users are partitioned into M clusters and each cluster consists of L_m users, where $m \in \{1, 2, \ldots, M\}$. Then we design appropriate beams for the corresponding clusters. Upon applying effective transmit precoding and detector designs, it becomes possible to guarantee that the beam associated with a particular cluster is orthogonal to the channels of users in other clusters. Hence the inter-cluster interference can be efficiently suppressed. When considering each cluster in isolation, there is a difference among the users' channel conditions, hence we are faced again with the conventional NOMA scenarios. Thus, SIC can be readily invoked for mitigating the intra-cluster interference between users of the same cluster. Recently, many important research contributions investigated beamforming aided NOMA (Liu et al. 2016e).

Specifically, Choi (2015) proposed two-stage multicast zero-forcing (ZF)-based beamforming for downlink inter-group/cluster interference mitigation, where the total transmit power of each group/cluster was minimized during the second stage. Higuchi and Benjebbour (2015) utilized receive beamformers at the NOMA users and a transmit beamformer at the BS. Higuchi and Kishiyama (2013) then proposed a novel scheme, which combined open-loop random beamforming in conjunction with intra-beam SIC for downlink NOMA transmission. However, random beamforming fails to guarantee a constant QoS at the users' side. To overcome this limitation, in Ding et al. (2016a), Ding et al. proposed a TPC and detection scheme combination for a cluster-based downlink MIMO-NOMA scenario relying on fixed PA. By adopting this design, their MIMO-NOMA system can be decomposed

into several independent single-input single-output (SISO) NOMA arrangements. Furthermore, in order to establish a more general framework considering both downlink and uplink MIMO-NOMA scenarios, the so-called signal alignment (SA) technique was proposed in Ding et al. (2016d). Stochastic geometry-based tools were invoked to model the impact of the NOMA users' locations. In contrast to the research contributions in Ding et al. (2016a,d), which are inter-cluster interference free design, an inter-cluster interference allowance design for CB MIMO-NOMA was proposed in Ali et al. (2017). Note that the existing NOMA designs have routinely relied on assuming different channel conditions for the different users, which is however a somewhat restrictive assumption. In order to circumvent this restriction, Ding et al. (2016b) designed a new MIMO-NOMA scheme, which distinguishes the users according to their QoS requirements with particular attention on IoT scenarios for the sake of supporting the SIC operation. Furthermore, they compared this new MIMO-NOMA design to two NOMA schemes, which order users according to the prevalent channel conditions. More particularly, the ZF-NOMA scheme of Ding et al. (2016a) and the SA-NOMA scheme (Ding et al. 2016d) were used as benchmarks in Ding et al. (2016b). Figure 3.9 illustrates the outage probability defined as the probability of erroneously detecting the message intended for User m in the i-th data stream, $i = 1, 2, 3$ at User n, where the QR decomposition is used to augmenting the differences between the users' effective channel conditions according to the associated QoS requirements. As shown in Fig. 3.9, the QR-based MIMO-NOMA scheme is capable of outperforming both ZF-NOMA and SA-NOMA[3] as well as MIMO-OMA,[4] since it exploits the heterogeneous QoS requirements of different users and applications. In Liu et al. (2016e), the fairness issues of the MIMO-NOMA scenario were addressed by applying appropriate user allocation algorithms among the clusters and dynamic PA algorithms within each cluster.

3.3.2 *Beamformer-Based MIMO-NOMA*

Another technique of implementing MIMO-NOMA is to assign different beams to different users, as shown in Fig. 3.10. By doing so, the QoS can be satisfied by calculating the beamformer-weights in a predefined order, commencing with the most demanding QoS requirement. By adopting this approach, several contributions have been made in terms of MIMO-NOMA. Considering the illustration of Fig. 3.10 as an example, User 1 to User N occupy the same RB, similar to user $(N+1)$ to user $(N+M)$. Again, within the same RB we may employ SIC at each user, according to

[3]Note that when $M = N$, ZF-NOMA achieves the same performance as SA-NOMA (Ding et al. 2016b).

[4]Figure 3.9 is focused on the performance of User n, since the QoS requirements have been guaranteed with the aid of appropriate PA (Ding et al. 2016b).

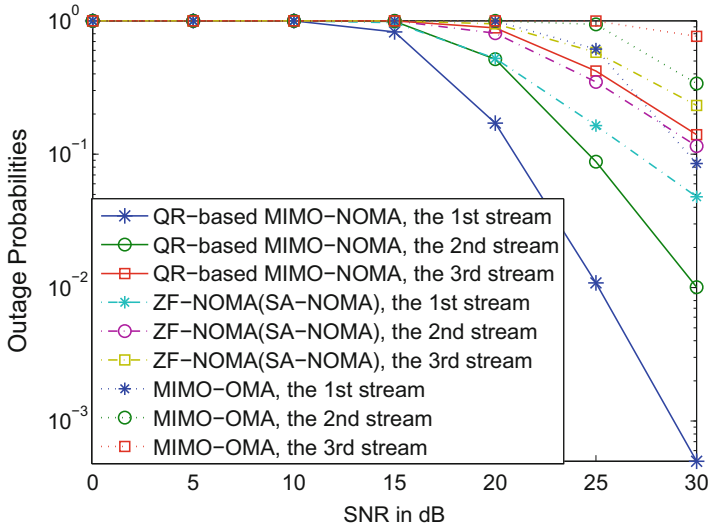

Fig. 3.9 Comparison of MIMO systems with various multiple-access techniques using a two-user case. $M = N = 3$, where M is the number of antennas at transmitters and N is the number of antennas at receivers. As such, each user is to receive three data streams from the BS. $R_m = 1.2$ bit per channel use (BPCU) and $R_n = 5$ BPCU are the targeted data rates of User m and User n, respectively

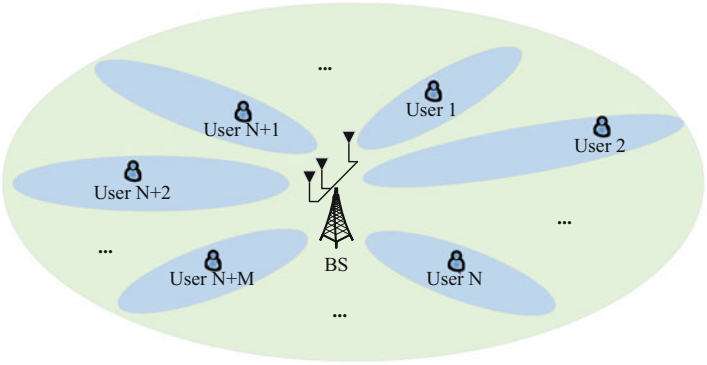

Fig. 3.10 Illustration of beamformer-based MIMO-NOMA

the particular ordering of the different users' received signal power. Sun et al. (2015) first investigated the power optimization problem constructed for maximizing the ergodic capacity and then showed that their proposed MIMO-NOMA schemes are capable of achieving significantly better performance than OMA. In an effort to reduce the decoding complexity imposed at the users, a layered transmission-based MIMO-NOMA scheme was proposed by Choi (2016), who also investigated

the associated PA problem. It was demonstrated that upon invoking this layered transmission scheme, the achievable sum rate increases linearly with the number of antennas.

3.3.3 Massive-MIMO-NOMA

Massive MIMO may be considered as one of the key technologies Andrews et al. (2014) in 5G systems as a benefit of improving both the received SNR and the bandwidth efficiency. It was shown in Larsson et al. (2014) that massive MIMO is capable of substantially increasing both the capacity and the energy efficiency. These compelling benefits of massive MIMO sparked off the interest of researchers also in the context of NOMA. In Ding and Poor (2016), Ding and Poor conceived a two-stage TPC design for implementing massive-MIMO-NOMA. More particularly, a beamformer was adopted for serving to a cluster of angularly similar users and then they decomposed the MIMO-NOMA channels into a number of SISO-NOMA channels within the same cluster. A one-bit CSI feedback scheme was proposed for maintaining a low feedback overhead and a low implementation complexity.

3.3.4 Cognitive Radio Inspired Power Control

The objective of CR inspired power control relying on NOMA is to guarantee the QoS of weak users by constraining the power allocated to the strong user. Inspired by the CR concept Goldsmith et al. (2009), NOMA can be regarded as a special case of CR networks (Ding et al. 2016c; Yang et al. 2016). More specifically, still considering a downlink scenario supporting two users, Fig. 3.11 compares conventional CR and CR inspired NOMA. The BS can be viewed as the combination of a primary transmitter (PT) and a secondary transmitter (ST), which transmits the superimposed signals. The strong user (User n) and the weak user (User m) can be regarded as a secondary receiver (SR) and a primary receiver (PR), respectively. By doing so, the strong User n becomes capable of accessing the spectrum occupied by the weak User m under predetermined interference constraints, which is the key feature of the classic underlay CR. The concept of CR-inspired PA in NOMA was proposed by Ding et al. (2016c), who investigated the PA of user-pairing-based NOMA systems.

The key advantages of cognitive PA are summarized as follows:

- **Guaranteed QoS**: by applying cognitive PA, the QoS requirements of the weak user are guaranteed, which is especially vital in real-time safety-critical applications.

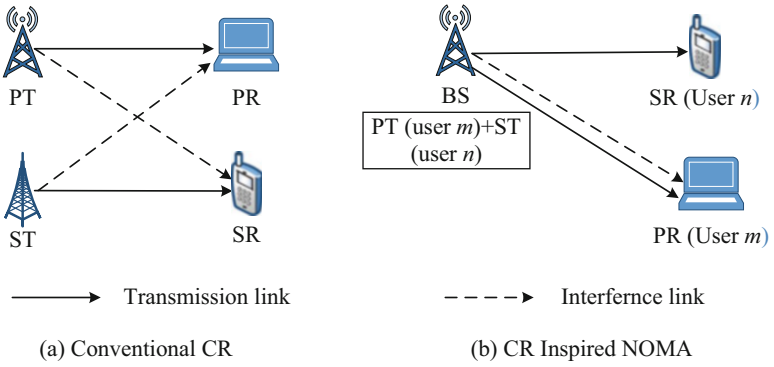

Fig. 3.11 Comparison of convectional CR (**a**) and CR inspired NOMA (**b**)

- **Fairness/throughput tradeoff**: cognitive PA is capable of striking a beneficial tradeoff between the overall system throughput and the individual user fairness, where the targeted data rate of the weak user has to be satisfied by appropriate PA.
- **High flexibility**: cognitive PA offers a high degree of freedom for the BS to explore the opportunistic support of the strong user.
- **Low complexity**: compared to the optimal PA approach, cognitive PA imposes a lower complexity during PA. This becomes particularly useful when the channel ordering and PA constraints are not convex and hence finding an appropriate PA scheme becomes a challenge, especially in multiple-antenna-aided NOMA scenarios.

Motivated by its advantages mentioned above, the cognitive PA policy was invoked for characterizing MIMO-NOMA systems. More particularly, in addition to investigating the convectional downlink cognitive PA conceived for MIMO-NOMA scenarios, the authors of Ding et al. (2016d) also designed a more sophisticated CR NOMA PA scheme for uplink MIMO-NOMA scenarios. In Ding et al. (2016b), in an effort to find a PA strategy suitable for SU-MIMO IoT scenarios, a cognitive PA policy was designed for ensuring that SIC may indeed be carried out at the strong user.

3.3.5 NOMA-Based Device-to-Device Communications

Due to the recent rapid increase in the demand for local area services under the umbrella of cellular networks, an emerging technique, namely device-to-device (D2D) communication, may be invoked for supporting direct communications among devices without the assistance of cellular BSs. The main advantages of integrating D2D communications into cellular networks are: (1) low-power support of

proximity services for improving the energy efficiency; (2) reusing the frequency of the over-sailing cellular networks in an effort to increase the bandwidth efficiency; and (3) the potential to facilitate new types of peer-to-peer (P2P) services (Ma et al. 2015).

Note that one of the common features of both D2D and NOMA is that of enhancing the bandwidth efficiency by managing the interference among users within each RB. Motivated by this, it is desirable to invoke intelligent joint interference management approaches for fully exploiting the potential benefits of both D2D and NOMA. In Zhao et al. (2016), a novel NOMA-based D2D communication scheme has been designed, where several D2D groups were permitted to share the same RB with the cellular users. In contrast to the conventional D2D pair's transmission, the novel "D2D group" concept was introduced, where a D2D transmitter was able to simultaneously communicate with multiple D2D receivers with the aid of NOMA. It was demonstrated that the proposed NOMA-based D2D scheme is capable of delivering higher throughput than conventional D2D communications.

3.4 Summary

This chapter discusses the compatibility of NOMA when it is applied to the HetNets, MIMO, and CRNs techniques.

References

Adhikary, A., Dhillon, H. S., & Caire, G. (2015). Massive-MIMO meets HetNet: Interference coordination through spatial blanking. *IEEE Journal on Selected Areas in Communications, 33*, 1171–1186.

Ali, S., Hossain, E., & Kim, D. I. (2017). Non-orthogonal multiple access (NOMA) for downlink multiuser MIMO systems: User clustering, beamforming, and power allocation. *IEEE Access, 5*, 565–577.

Andrews, J. G., Buzzi, S., Choi, W., Hanly, S. V., Lozano, A., Soong, A. C., et al. (2014). What will 5G be? *IEEE Journal on Selected Areas in Communications, 32*, 1065–1082.

Benjebbour, A., Saito, Y., Kishiyama, Y., Li, A., Harada, A., & Nakamura, T. (2013). Concept and practical considerations of non-orthogonal multiple access (NOMA) for future radio access. In *Proceedings of IEEE Intelligent Signal Processing and Communications Systems (ISPACS)* (pp. 770–774).

Chen, Z., Ding, Z., Dai, X., & Karagiannidis, G. K. (2016). On the application of quasi-degradation to MISO-NOMA downlink. *IEEE Transactions on Signal Processing, 64*, 6174–6189.

Choi, J. (2015). Minimum power multicast beamforming with superposition coding for multiresolution broadcast and application to NOMA systems. *IEEE Transactions on Communications, 63*, 791–800.

Choi, J. (2016). On the power allocation for MIMO-NOMA systems with layered transmissions. *IEEE Transactions on Wireless Communications, 15*, 3226–3237.

Ding, Z., Adachi, F., & Poor, H. V. (2016a). The application of MIMO to non-orthogonal multiple access. *IEEE Transactions on Wireless Communications, 15*, 537–552.

Ding, Z., Dai, L., & Poor, H. V. (2016b). MIMO-NOMA design for small packet transmission in the internet of things. *IEEE Access, 4*, 1393–1405.

Ding, Z., Fan, P., & Poor, H. V. (2016c). Impact of user pairing on 5G non-orthogonal multiple access. *IEEE Transactions on Vehicular Technology, 65*, 6010–6023.

Ding, Z., & Poor, H. V. (2016). Design of massive-MIMO-NOMA with limited feedback. *IEEE Signal Processing Letters, 23*, 629–633.

Ding, Z., Schober, R., & Poor, H. V. (2016d). A general MIMO framework for NOMA downlink and uplink transmission based on signal alignment. *IEEE Transactions on Wireless Communications, 15*, 4438–4454.

Ding, Z., Yang, Z., Fan, P., & Poor, H. V. (2014). On the performance of non-orthogonal multiple access in 5G systems with randomly deployed users. *IEEE Signal Processing Letters, 21*, 1501–1505.

Goldsmith, A., Jafar, S. A., Maric, I., & Srinivasa, S. (2009). Breaking spectrum gridlock with cognitive radios: An information theoretic perspective. *Proceedings of the IEEE, 97*, 894–914.

Gradshteyn, I. S., & Ryzhik, I. M. (2000). *Table of integrals, series and products* (6th edn.). New York: Academic.

Hanif, M. F., Ding, Z., Ratnarajah, T., & Karagiannidis, G. K. (2016). A minorization-maximization method for optimizing sum rate in the downlink of non-orthogonal multiple access systems. *IEEE Transactions on Signal Processing, 64*, 76–88.

Higuchi, K., & Benjebbour, A. (2015). Non-orthogonal multiple access (NOMA) with successive interference cancellation for future radio access. *IEICE Transactions on Communications, 98*, 403–414.

Higuchi, K., & Kishiyama, Y. (2013). Non-orthogonal access with random beamforming and intra-beam SIC for cellular MIMO downlink. In *Proceedings of IEEE Vehicular Technology Conference (VTC Fall)* (pp. 1–5).

Hosseini, K., Yu, W., & Adve, R. S. (2014). Large-scale MIMO versus network MIMO for multicell interference mitigation. *IEEE Journal of Selected Topics in Signal Processing, 8*, 930–941.

Huh, H., Tulino, A. M., & Caire, G. (2012). Network MIMO with linear zero-forcing beamforming: Large system analysis, impact of channel estimation, and reduced-complexity scheduling. *IEEE Transactions on Information Theory, 58*, 2911–2934.

Jo, H.-S., Sang, Y. J., Xia, P., & Andrews, J. G. (2012). Heterogeneous cellular networks with flexible cell association: A comprehensive downlink SINR analysis. *IEEE Transactions on Wireless Communications, 11*, 3484–3495.

Kim, B., Lim, S., Kim, H., Suh, S., Kwun, J., Choi, S., et al. (2013). Non-orthogonal multiple access in a downlink multiuser beamforming system. In *Proceedings of Military Communications Conference (MILCOM)*, pp. 1278–1283.

Larsson, E., Edfors, O., Tufvesson, F., & Marzetta, T. (2014). Massive MIMO for next generation wireless systems. *IEEE Communications Magazine, 52*, 186–195.

Liu, L., Yuen, C., Guan, Y. L., & Li, Y. (2016a). Capacity-achieving iterative LMMSE detection for MIMO-NOMA systems. In *IEEE Proceedings of International Communication Conference (ICC), Kuala Lumpur, Malaysia*.

Liu, W., Jin, S., Wen, C. K., Matthaiou, M., & You, X. (2016b). A tractable approach to uplink spectral efficiency of two-tier massive MIMO cellular HetNets. *IEEE Communications Letters, 20*, 348–351.

Liu, Y., Ding, Z., Elkashlan, M., & Poor, H. V. (2016c). Cooperative non-orthogonal multiple access with simultaneous wireless information and power transfer. *IEEE Journal on Selected Areas in Communications, 34*(4), 938–953, April 2016.

Liu, Y., Ding, Z., Elkashlan, M., & Yuan, J. (2016d). Non-orthogonal multiple access in large-scale underlay cognitive radio networks. *IEEE Transactions on Vehicular Technology, 65*, 10152–10157.

Liu, Y., Elkashlan, M., Ding, Z., & Karagiannidis, G. K. (2016e). Fairness of user clustering in MIMO non-orthogonal multiple access systems. *IEEE Communications Letters, 20*, 1465–1468.

Ma, C., Wu, W., Cui, Y., & Wang, X. (2015). On the performance of successive interference cancellation in D2D-enabled cellular networks. In *Proceedings of IEEE International Conference on Computer Communication (INFOCOM), Kowloon, Hong Kong* (pp. 37–45)

Qin, Z., Fan, J., Liu, Y., Gao, Y., & Li, G. Y. (2018a). Sparse representation for wireless communications: A compressive sensing approach. *IEEE Signal Processing Magazine, 35*, 40–58.

Qin, Z., Gao, Y., & Parini, C. G. (2016a). Data-assisted low complexity compressive spectrum sensing on real-time signals under sub-Nyquist rate. *IEEE Transactions on Wireless Communications, 15*, 1174–1185.

Qin, Z., Gao, Y., Plumbley, M., & Parini, C. (2016b). Wideband spectrum sensing on real-time signals at sub-Nyquist sampling rates in single and cooperative multiple nodes. *IEEE Transactions on Signal Processing, 64*, 3106–3117.

Qin, Z., Liu, Y., Li, Y., & McCann, J. A. (2019). Performance analysis of clustered LoRa networks. In *IEEE Transactions on Vehicular Technology, 68*(8), 7616–7629, Aug. 2019.

Qin, Z., Yue, X., Liu, Y., Ding, Z., & Nallanathan, A. (2018b). User association and resource allocation in unified NOMA enabled heterogeneous ultra dense networks. *IEEE Communications Magazine, 56*, 86–92.

Qureshi, S., Kim, H., & Hassan, S. A. (2016). MIMO uplink NOMA with successive bandwidth division. In *Proceedings of IEEE Wireless Communication and Networking Conference, MILCOM, Doha*

Saito, Y., Kishiyama, Y., Benjebbour, A., Nakamura, T., Li, A., & Higuchi, K. (2013). Non-orthogonal multiple access (NOMA) for cellular future radio access. In *IEEE Proceedings of Vehicle Technology Conference (VTC), Dresden* (pp. 1–5).

Sun, Q., Han, S., I, C.-L., & Pan, Z. (2015). On the ergodic capacity of MIMO NOMA systems. *IEEE Wireless Communications Letters, 4*, 405–408.

Yang, Z., Ding, Z., Fan, P., & Al-Dhahir, N. (2016). A general power allocation scheme to guarantee quality of service in downlink and uplink NOMA systems. *IEEE Transactions on Wireless Communications, 15*, 7244–7257.

Ye, Q., Bursalioglu, O. Y., Papadopoulos, H. C., Caramanis, C., & Andrews, J. G. (2015). User association and interference management in massive MIMO HetNets. arXiv preprint arXiv:1509.07594.

Zhao, J., Liu, Y., Chai, K. K., Chen, Y., Elkashlan, M., & Alonso-Zarate, J. (2016). NOMA-based D2D communications towards 5G. In *IEEE Proceedings of Global Communication Conference (GLOBECOM)*, Washington (pp. 1–6).