Relay Selection in NOMA-Based Cooperative Wireless Backhaul Networks

Noha H. Khattab[1,3], M. Saeed Darweesh[1] and Samy S. Soliman[2,3]

[1]Wireless Intelligent Networks Center (WINC), Nile University, Giza, Egypt
[2]Electronics and Electrical Commuications Engineering Department, Faculty of Engineering, Cairo University, Giza, Egypt
[3]University of Science and Technology, Zewail City of Science, Technology and Innovation, Giza, Egypt

***Abstract*—The joint application of wireless backhaul networks and non-orthogonal multiple access (NOMA) hold the potential to fulfill the increasing demands of fifth-generation (5G) communication networks and beyond. It is usual in wireless backhaul networks to take assistance from small cell base stations acting as intermediate relays to reach the remote destination. This cooperative communication is an acknowledged technique to combat multi-path fading, improve energy efficiency, and enhance the reliability and capacity of wireless networks. This article studies the application of relay selection (RS) in the proposed NOMA-based cooperative wireless backhaul network model. The system performance is analyzed through closed-form and asymptotic expressions of the sum ergodic rate. The paper also presents practical design guidelines of optimum location for small base stations of wireless backhaul networks.**

***Index Terms*—Wireless Backhaul, ergodic capacity, NOMA, UDNs, relay selection, sum rate, node placement.**

I. Introduction

Due to the exponential growth rate of mobile data traffic, fifth-generation (5G) communication networks and beyond face challenges of high connection density, spectrum usage efficiency, energy efficiency, and user fairness [1]. Some promising candidate technologies have been proposed to address these challenges, such as ultra-dense networks, cooperative networks, and non-orthogonal multiple access [2].

Deploying ultra-dense networks (UDNs) can significantly enhance the system capacity by densely using low-power relaying nodes, each called a small base station (SBS), which assist in forwarding the traffic of the base station (BS) [3]. However, one limiting factor in the practical deployment of UDNs is the backhaul network, which is required to handle the forwarding of hundreds of gigabits of traffic to the core network. Moreover, it is not always feasible to provide wired backhaul connectivity between the BS and the large amount of SBSs. Thus, wireless backhaul has been considered a feasible, cost-effective solution for UDNs, which improves network capacity and extends the network's coverage [4], In addition, the choice of a proper power allocation scheme among the SBSs can further improve the overall network capacity [5].

In order to guarantee the massive connectivity in wireless backhaul UDNs, new promising multiple access schemes are necessary. Non-orthogonal multiple access (NOMA) gained much research interest because it supports massive connectivity, delivers higher throughput, and increases spectrum efficiency. In contrast to the conventional orthogonal multiple access (OMA) techniques, the basic principle of NOMA systems is that multiple receivers, with different channel gains, can share the same resource block, such as time, frequency, or code. Different signals are multiplexed at the transmitter using superposition coding, with different power coefficients allocated to various data symbols. At the receiver, successive interference cancellation (SIC) is used to retrieve the desired information by removing undesired symbols from other users based on the power allocation [6], [7]. Several studies have shown that NOMA-based systems can enable massive connectivity while offering higher spectral efficiency and user fairness when compared to OMA-based systems [8]–[11].

In NOMA systems, seeking cooperation from intermediate relays increases the spatial diversity, improve the energy efficiency, and enhance wireless networks' reliability [12]. As a result, a proper relay selection (RS) scheme in such cooperative NOMA networks becomes crucial for achieving spatial diversity since improper relay selection can decrease the network's overall capacity [13], [14].

In addition to the previously mentioned techniques, multiple-input multiple-output (MIMO) systems have proved to increase the system capacities. In this regard, base stations are often equipped with more than one antenna. Accordingly, multi-antenna systems have been combined with cooperative NOMA networks to achieve further capacity enhancement [15]–[17]. In multi-antenna NOMA systems, perfect channel state information at the transmitter (CSIT) is always assumed. However, in the practical scenario of imperfect CSIT, multi-antenna NOMA precoder optimization can be used to overcome CSIT imperfection as studied in [18].

II. Related Works

Driven by the key features of UDNs and Cooperative NOMA, many authors in the literature have investigated the Cooperative NOMA approach in UDNs in order to find a proper architecture suitable for wireless backhaul. In [19]–[21], Authors studied NOMA-based cooperative wireless backhaul, which aims to reduce power consumption and improve the energy efficiency of the wireless backhaul. In [19], authors proposed a two-layer hierarchical model of cooperative wireless backhaul. The optimization problem for maximizing the energy efficiency based on cooperative NOMA and OMA

has been solved and analyzed. The authors in [20] propose another version of cooperative wireless backhaul, where the relays that forward the traffic from the small base stations to the core network are connected through fiber cables. A genetic algorithm combined with a K-means clustering algorithm is introduced to select those forwarding relays' locations that maximize the backhaul capacity. In [21], the authors introduced a new cooperative wireless backhaul model that maximizes the multiplexing gain. The model possesses two SBSs serve as relays between the BS and only one user. Utilizing two relays in the cooperation transmission between BS and user enables the BS to transmit two symbols simultaneously to the destination within only two-time transmission phases. Also, a power allocation scheme that maximizes the overall throughput is introduced. However, this work was only limited to a single-relay scenario.

In this work, motivated by the observations of the literature, a modified NOMA-based cooperative wireless backhaul network has been introduced. Different from the work in [21], this paper considers a network with a single BS, a single destination node, and randomly distributed relaying nodes in between. Following the downlink NOMA principle, the source sends two superimposed symbols with different power allocations to two intermediate SBSs with varying distances from the BS. Then, the SBSs forward the decoded signals to the destination using NOMA in an uplink mode. As a result, the benefits of NOMA in both downlink and uplink transmission can be acquired. While the presence of multiple relaying nodes between the source and the destination allows for multihop relaying, it was shown in [22] that using multiple branches of dual-hop relaying provides better performance than using a single multihop branch. Hence, in the proposed backhaul network, two SBSs, acting as relays, are selected to result two dual-hop branches between the BS and the destination node. A proper RS scheme is proposed for selecting these two SBSs to maximize the destination sum rate.

Also, the impact of increasing the number of SBSs at either optimum or non-optimum relay locations is investigated. Hence, a practical design guideline for optimal locations of SBSs for wireless backhaul networks is investigated. The capacity enhancement of enabling multiple input single output (MISO) technology to the proposed model in the downlink phase has been demonstrated. Finally, the analytical expressions for the sum rate of the proposed system are derived under Rayleigh fading channels.

The remainder of this paper is organized as follows. The system model, with relay selection, is introduced in Section II. Analytical expressions of the ergodic sum rate are presented in Section III. Numerical results are shown in Section IV, and Section V concludes the paper.

III. System Model

Considers a single base station source (S), a single destination (D), and two sets of SBSs relays in between. Each set of relays exist at a different distance from S and D. Each relay exists at a distinct distance from S and D. This forms two communication branches between the source and the destinations, referred to as the upper and lower branches, as shown in Fig. 1. The intermediate relays work to cooperate with the transmission of data symbols from the source to the destination. Assume that R_1 is the best relay selected from the upper branch relays $R_{1,i}$, where $1 \leq i \leq N$, and R_2 is the relay selected from the lower branch relays $R_{2,j}$, where $1 \leq j \leq F$. Accordingly, a pair of symbols is transmitted from the source to the destination over two phases, which may be considered a downlink NOMA phase and an uplink NOMA phase. It is assumed that all other nodes than S, are equipped with a single antenna hence operate in a half-duplex mode, whereas perfect channel state information (CSI) is available at the base station [1]. In our model, decode-and-forward (DF) technique is used at the relays, which is more appropriate for this model than the simpler amplify-and-forward (AF) approach. Because the relays should send only one intended symbol to the destination, whereas AF-based relays would transmit both symbols, complicating the receiving process.

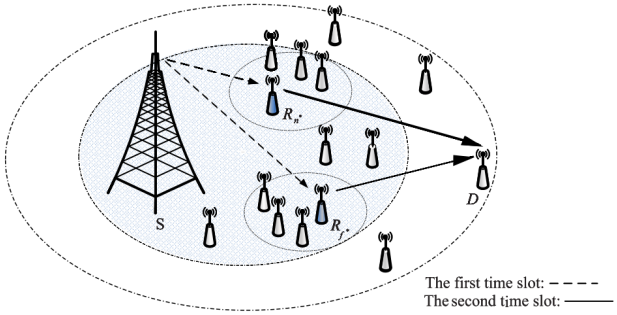

Fig. 1. Two tier of wireless backhaul network model

The following section will introduce the mathematical model and expressions that relate the achievable data rate to the system parameters based on the specified system. The downlink phase is analyzed in both scenarios, single-antenna BS and Multi-antenna BS, to compare the enhancement effect of MIMO systems. However, the analysis has been introduced considering the single-antenna scenario only.

A. The backhaul links downlink NOMA phase

a) Multi-antenna Base station:

In the downlink NOMA phase, the base station S equipped with N_t antennas where the channel matrix between the source and the K SBS relays is denoted by $\mathbf{H} \in \mathbb{C}^{K \times N_t}$, which can be written as $\mathbf{H} = \mathbf{D}^{1/2}\mathrm{G}$ where $\mathrm{G} \in \mathbb{C}^{K \times N_t}$ contains i.i.d $\mathcal{CN}(0,1)$ elements and $\mathbf{D} \in \mathbb{C}^{K \times K}$ is a diagonal matrix represents the path-loss attenuation with $[\mathbf{D}]_k = d_k^{-n}$, where d_k is the distance between the BS and the k^{th} SBS, and n is the path loss exponent.

Two messages m_1 and m_2 are intended for R_1 and R_2, which are independently encoded into the data stream s_1 for

[1]Some common methods can be used to estimate the CSI, such as the least-squares estimator or the maximum likelihood estimator. Also, in a recent work, the authors in [23] investigated NOMA in real-time experimentation to study the practical challenges such as synchronization and SIC error propagation.

R_1 and s_2 for R_2, respectively. The k streams are grouped in a signal vector $\mathbf{s} = [s_1, s_2]^T \in \mathbb{C}^k$, where $\mathcal{E}\{\mathbf{s}\mathbf{s}^H\} = \mathbf{I}$ and $\text{tr}(\mathbf{P}\mathbf{P}^H) \leq P_t$ where P_t is the total power of the base station. Then, symbols are linearly precoded and mapped to the transmit antennas through a precoding matrix defined as $\mathbf{W} = [\mathbf{w}_1, \mathbf{w}_2]$ where $\mathbf{w}_k \in \mathbb{C}^{N\times 1}$ is the k^{th} precoder. The maximum ratio transmission (MRT) is used for precoders initialization. Accordingly, the precoder $\mathbf{w}_k$ is initialized as $\mathbf{w}_k = \alpha_k P_t \frac{\mathbf{h}_k}{\|\mathbf{h}_k\|}$, where α_k is the power allocation coefficient assigned to SBS relay R_k and $k = \{1, 2\}$.

The distance between the source, S, and R_1 is assumed to be shorter than the distance between the source and R_2. By assuming $d_{SR_1} < d_{SR_2}$, it is possible to consider that the norm of the channel vectors can be written as $\|\mathbf{h}_{SR_1}\| \geq \|\mathbf{h}_{SR_2}\|$. Therefore, the superimposed transmitted signal is expressed as

$$\mathbf{x} = \mathbf{W}\mathbf{s} = \mathbf{w}_1 s_1 + \mathbf{w}_2 s_2. \tag{1}$$

Hence, the received signal at the R_k relay is given by

$$y_k = \mathbf{h}_k^H \mathbf{x} + n_k \tag{2}$$

where $\mathbf{h}_k^T \in \mathbb{C}^{N_t\times 1}$ is the $K \times 1$ channel vector from the BS to the k^{th} relay and n_k is the additive white Gaussian noise (AWGN) at the R_k relay, $n_k \sim \mathcal{CN}(0, \sigma_k^2)$.

It is assumed that the first symbol, s_1, is extracted at R_1 through applying SIC by decoding the s_2 first, while R_2 decodes for s_2 directly. Hence, the power allocation coefficients must be such that $\alpha_2 > \alpha_1$. Both relays shall acquire the second symbol s_2 by directly decoded, and s_1 could be treated as noise. Hence, the signal-to-noise ratio at R_k for decoding s_2, denoted by $\gamma_{s_2\to R_k}$, can be expressed as

$$\gamma_{s_2\to R_k}{}^{MISO} = \frac{\left|\mathbf{h}_k^H \mathbf{w}_2\right|^2}{\left|\mathbf{h}_k^H \mathbf{w}_1\right|^2 + 1} \tag{3}$$

In order to consider s_2 successfully decoded, the data rate for s_2 on the S-to-R_1 link,, denoted as $C_{S_{R_1}}^{s_2}$, must be greater than the target data rate, R_{s_2}, for s_2. This condition can be expressed as

$$C_{S_{R_1}}^{s_2 MISO} = \log_2\left(1 + \frac{\left|\mathbf{h}_1^H \mathbf{w}_2\right|^2}{\left|\mathbf{h}_1^H \mathbf{w}_1\right|^2 + 1}\right) \geq R_{s_2}. \tag{4}$$

With this condition fulfilled, SIC is completed to decode s_1. As a result, the signal to noise ratio at R_1 for decoding s_1, denoted by $\gamma_{s_1\to R_1}$, can be written as

$$\gamma_{s_1\to R_1}{}^{MISO} = \left|\mathbf{h}_1^H \mathbf{w}_1\right|^2. \tag{5}$$

Thus, the corresponding data rate achieved for symbol s_1, denoted as $C_{S_{R_1}}^{s_1}$, on the S-to-R_1 link can be expressed as:

$$C_{S_{R_1}}^{s_1 MISO} = \log_2\left(1 + \gamma_{s_1\to R_1}{}^{MISO}\right). \tag{6}$$

Accordingly, from (3), the corresponding data rate achieved for symbol s_2 on the S-to-R_2 link, and denoted $C_{S_{R_2}}^{s_2}$, can be given as

$$\begin{aligned} C_{S_{R_2}}^{s_2 MISO} &= \log_2\left(1 + \gamma_{s_2\to R_2}^{MISO}\right) \\ &= \frac{\left|\mathbf{h}_2^H \mathbf{w}_2\right|^2}{\left|\mathbf{h}_2^H \mathbf{w}_1\right|^2 + 1}. \end{aligned} \tag{7}$$

b) Single Antenna Base Station:

Consider the base Station S equipped with single antenna. The channel coefficients for the links between the source and R_1, the source and R_2 are denoted, respectively, by h_{SR_1} and h_{SR_2}. The channel gains h_i, where $i \in \{SR_1, SR_2\}$ are assumed to be Rayleigh faded, *i.e.* $h_i \sim \mathcal{CN}(0, \lambda_i)$, with zero mean and variance $\lambda_i = d_k^{-n}$ where n is the path loss exponent and d_k is the distance between the BS and the R_k relay.

In the downlink NOMA phase, the source transmits two superimposed data symbols, s_1 and s_2. The transmitted signal representing the superimposed signals, s, is expressed as $s = \sqrt{\alpha_1 P_t} s_1 + \sqrt{\alpha_2 P_t} s_2$, where α_1 and α_2 are the power allocation coefficients. The first symbol, s_1, is to be decoded at R_1, Similarly, the second symbol s_2 is to be decoded at R_2. Both decoded symbols will be transmitted simultaneously, each from one of the two relays, to the destination during the uplink NOMA phase.

The distance between the source, S, and R_1 is assumed to be shorter than the distance between the source and R_2. By assuming $d_{SR_1} < d_{SR_2}$, it is possible to consider that the average channel gains are such that $\lambda_{SR_1} > \lambda_{SR_2}$. Accordingly, $\mathbb{E}\{|h_{SR_1}|^2\} > \mathbb{E}\{|h_{SR_2}|^2\}$.

The received signal at the R_k relay, can be given as

$$\begin{aligned} y_k &= h_{SR_k} s + n_k \\ &= h_{SR_k}(\sqrt{\alpha_1 P_t} s_1 + \sqrt{\alpha_2 P_t} s_2) + n_k \end{aligned} \tag{8}$$

where n_k represents zero mean AWGN, with average noise power σ^2, at R_k. It is assumed that the first symbol, s_1, is extracted at R_1 through applying SIC by decoding the s_2 first, while R_2 decodes for s_2 directly. Hence, the power allocation coefficients must be such that $\alpha_2 > \alpha_1$. Both relays shall acquire the second symbol s_2 by directly decoded, and s_1 could be treated as noise. Hence, the signal-to-noise ratio at R_k for decoding s_2, denoted by $\gamma_{s_2\to R_k}$, can be expressed as

$$\gamma_{s_2\to R_k} = \frac{\alpha_2 \rho_t |h_{SR_k}|^2}{\alpha_1 \rho_t |h_{SR_k}|^2 + 1}, \tag{9}$$

where $\rho_t = \frac{P_d}{\sigma^2}$ represents the ratio between the total transmit signal power to the average noise power at R_k relay.

In order to consider s_2 successfully decoded, the data rate for s_2, denoted as $C_{S_{R_1}}^{s_2}$, must be greater than the target data rate, R_{s_2}, for s_2. This condition can be expressed as

$$C_{S_{R_1}}^{s_2} = \log_2\left(1 + \frac{\alpha_2 \rho_t |h_{SR_k}|^2}{\alpha_1 \rho_t |h_{SR_k}|^2 + 1}\right) \geq R_{s_2}. \tag{10}$$

With this condition fulfilled, SIC is completed to decode s_1. As a result, the signal to noise ratio at R_1 for decoding s_1, denoted by $\gamma_{s_1\to R_1}$, can be written as

$$\gamma_{s_1\to R_1} = \alpha_1 \rho_t |h_{SR_1}|^2. \tag{11}$$

Thus, the corresponding data rate achieved for symbol s_1, denoted as $C_{S_{R_1}}^{s_1}$, on the S-to-$R1$ link can be expressed as:

$$C_{S_{R_1}}^{s_1} = \log_2(1 + \gamma_{s_1\to R_1}) = \log_2(1 + \alpha_1 \rho_t |h_{SR_1}|^2). \tag{12}$$

Accordingly, from (9), the corresponding data rate achieved for symbol s_2 on the S-to-$R2$ link, and denoted $C^{s_2}_{S_{R_2}}$, can be given as

$$\begin{aligned} C^{s_2}_{S_{R_2}} &= \log_2\left(1+\gamma_{s_2\to R_2}\right) \\ &= \log_2\left(1+\frac{\alpha_2\rho_t|h_{SR_2}|^2}{\alpha_1\rho_t|h_{SR_2}|^2+1}\right). \end{aligned} \quad (13)$$

B. The access links - Uplink NOMA Phase

In the second transmission phase, each of the two relays R_1 and R_2, forwards the decoded symbol, s_1 and s_2, respectively, to the destination node. Similar to the single antenna downlink phase, the channel coefficients for the links between R_1 and the destination and R_2 to the destination are denoted, respectively, by h_{R_1D} and h_{R_2D}. The channels are also assumed to be Rayleigh faded. Following the uplink NOMA principle, R_1 transmits the data symbol s_1 with transmit power $\alpha_3 P_t$, while R_2 transmits the data symbol s_2 with transmit power $\alpha_4 P_t$.

Accordingly, the received signal at the destination is expressed as

$$y_d = \underbrace{\sqrt{\alpha_3 P_t}s_1 h_{R_1D}}_{\text{From } R_1} + \underbrace{\sqrt{\alpha_4 P_t}s_2 h_{R_2D}}_{\text{From } R_2 \text{ (dominating)}} + w_3 \quad (14)$$

where w_3 represents zero mean AWGN, with average noise power σ^2, at the destination D. Based on the described model, the relay R_2 is closer to the destination than the relay R_1, implying a higher channel gain for the R_2-to-D link than that for the R_1-to-D link, i.e. $|h_{R_2D}|^2 > |h_{R_1D}|^2$. With appropriate power allocation of $\alpha_3 P_t$ and $\alpha_4 P_t$ such that $\alpha_4 > \alpha_3$, the signal received at the destination is dominated by the signal sent from R_2. This validated the NOMA considerations for the uplink power domain during this phase of transmission. It is worth noting that the total power allocated for the transmission of both symbols s_1 and s_2 is considered to be P_t, which includes the power utilized at the source and both forwarding relays. Accordingly, it is assumed that $\sum_{i=1}^{4}\alpha_i = 1$. It should be noted that optimum power allocation has been an essential aspect of NOMA-based communication models. However, the details of the power allocation algorithms followed at the base station are considered in other work.

Similar to the analysis presented for the downlink NOMA phase, the received signal expressed in (14) is processed to directly decode s_2, the stronger component, treating s_1 as noise. Therefore, the signal to noise ratio, $\gamma_{s_2\to D}$, at the destination for decoding s_2 can be expressed as

$$\gamma_{s_2\to D} = \frac{\alpha_4\rho_t|h_{R_2D}|^2}{\alpha_3\rho_t|h_{R_1D}|^2+1}. \quad (15)$$

The achievable data rate, $C^{s_2}_{R_2D}$, for symbol s_2 on the R_2-to-D link can be written as

$$C^{s_2}_{R_2D} = \log_2\left(1+\frac{\alpha_4\rho_t|h_{R_2D}|^2}{\alpha_3\rho_t|h_{R_1D}|^2+1}\right). \quad (16)$$

Successful decoding of s_2 is considered if the achievable data rate is greater than a determined target data rate R_{s_2}, resulting in the condition

$$C^{s_2}_{R_2D} \geq R_{S_2}, \quad (17)$$

With s_2 decoded, SIC is continued to decode s_1. As a result, the signal to noise ratio at destination for decoding s_1, $\gamma_{s_1\to D}$, is expressed as

$$\gamma_{s_1\to D} = \alpha_3\rho_t|h_{R_1D}|^2. \quad (18)$$

The resulting achievable data rate, $C^{s_1}_{R_1D}$, for symbol s_1 on the R_1-to-D link can be then expressed as

$$\begin{aligned} C^{s_1}_{R_1D} &= \log_2\left(1+\gamma_{s_1\to D}\right) \\ &= \log_2\left(1+\alpha_3\rho_t|h_{R_1D}|^2\right). \end{aligned} \quad (19)$$

C. Overall achievable capacity

The overall system achievable data rate is calculated by adding the achievable rates of both delivered symbols, C_{s_1} and C_{s_2}. According to [12], the end-to-end achievable data rate of conventional cooperative relaying system is limited by the worst link of the source-relay and relay-destination links. From (12) and (19), it can be obviously observed that the achievable rate of s_1 will be limited by the minimum of $C^{s_1}_{S_{R_1}}$ and $C^{s_1}_{R_1D}$. Similarly, from (10), (13) and (16), the achievable rate of s_2 will be limited to the minimum of $C^{s_2}_{S_{R_1}}$, $C^{s_2}_{S_{R_2}}$ and $C^{s_2}_{R_2D}$. Thus, the overall achievable data rate is written as

$$\begin{aligned} C = &\underbrace{\frac{1}{2}\min(C^{s_1}_{S_{R_1}}, C^{s_1}_{R_1D})}_{C_{s_1}} \\ &+ \underbrace{\frac{1}{2}\min(C^{s_2}_{S_{R_1}}, C^{s_2}_{S_{R_2}}, C^{s_2}_{R_2D})}_{C_{s_2}}, \end{aligned} \quad (20)$$

where the factor $\frac{1}{2}$ in (20) accounts for the fact that each symbol requires two phases to reach the destination.

D. Relay Selection Strategy

As proposed in the system model, relay selection can be employed to enhance the performance of the proposed wireless backhaul model. The best relay selection scheme is adopted, in which the pair of relays that can provide the highest end-to-end sum achievable data in (20) are selected to convey symbols to the destination. Accordingly, the $N^{*\text{th}}$ relay is selected from the upper branch available relays, and the $F^{*\text{th}}$ relay is selected from the lower branch available relays if their combined data rate achieves the maximum sum data rate. This can be expressed as

$$\{n^*, f^*\} = \underset{n\in[1,N], f\in[1,F]}{\arg\max}\{C_{n,f}\}. \quad (21)$$

The optimally selected relays are denoted R_1 and R_2 in the analysis presented in (8) - (19). The effect of the numbers of relays, N and F, and the effects of their relative positions to each other as well as relative to the source base station and the destination are studied in the simulation and results section.

In the following section, the ergodic capacity, based on the described model of single-antenna scenario, is defined. Closed-form and asymptotic expressions for the ergodic capacities of both symbols s_1 and s_2, are also derived.

IV. Ergodic Capacity Analysis

In this section, the system performance in terms of the sum ergodic rate is discussed. For the system model discussed in the previous section, the sum ergodic data rate, $\mathbb{E}\{C\}$, *i.e.* the average sum of the achievable data rate of symbols s_1 and s_2 in (20) can be expressed as

$$\mathbb{E}\{C\} = \mathbb{E}\{C_{s_1}\} + \mathbb{E}\{C_{s_2}\}. \tag{22}$$

The ergodic data rate is calculated by averaging the instantaneous capacity achieved through transmission over a varying fading channel. Since the data rate is dependent on the signal to noise ratio, the ergodic data rate is expressed as

$$\mathbb{E}\{C_x\} = \int_0^\infty \log_2(1+x) f_X(x) dx \tag{23}$$

$$= \frac{1}{\ln 2} \int_0^\infty \frac{1 - F_X(x)}{1+x} dx, \tag{24}$$

where $f_X(x)$ and $F_X(x)$ are respectively the probability density function (PDF) and the cumulative distribution function (CDF) of random random variable X. In the following subsections, the ergodic data rate for each of the two symbols s_1 and s_2 is calculated.

A. Ergodic data rate of symbol s_1

The instantaneous data rate achieved for symbol s_1 at the receiver can be obtained by substituting (12) and (19) into the first term of (20), resulting in

$$C_{s_1} = \frac{1}{2} \log_2 \left(1 + \min\left\{\gamma_{s_1 \to R_1}, \gamma_{s_1 \to D}\right\}\right). \tag{25}$$

Consequently, the ergodic capacity for s_1 can be evaluated through

$$\mathbb{E}\{C_{s_1}\} = \frac{1}{2} \frac{1}{\ln 2} \int_0^\infty \frac{1 - F_U(u)}{1+u} du. \tag{26}$$

where $U = \min\{\gamma_{s_1 \to R_1}, \gamma_{s_1 D}\}$. Therefore, using (26) to calculate the ergodic data rate of symbol s_1 requires the calculation of the CDF of U. Using order statistics, the CDF of the random variable U can be obtained as

$$\begin{aligned} F_U(u) =& P(\min\{\gamma_{s_1 \to R_1}, \gamma_{s_1 \to D}\} < u) \\ =& 1 - P(\gamma_{s_1 \to R_1} \geq u) P(\gamma_{s_1 \to D} \geq u) \\ =& 1 - F_{\gamma_{s_1 \to R_1}}(u) F_{\gamma_{s_1 \to D}}(u). \end{aligned} \tag{27}$$

Since the random variables $\gamma_{s_1 \to R_1}$ and $\gamma_{s_1 \to D}$, defined in (11) and (18), follow exponential distribution with means $\alpha_1 \rho_t \lambda_{SR_1}$ and $\alpha_3 \rho_t \lambda_{R1_D}$, hence, their CDFs can be, respectively, expressed as

$$F_{\gamma_{s_1 \to R_1}}(u) = 1 - e^{-\frac{u}{\alpha_1 \rho_t \lambda_{SR1}}} \qquad u > 0. \tag{28}$$

$$F_{\gamma_{s_1 \to D}}(u) = 1 - e^{-\frac{u}{\alpha_3 \rho_t \lambda_{R1D}}} \qquad u > 0. \tag{29}$$

Therefore, substituting (28) and (29) into (27), the CDF of random variable U can be given by

$$F_U(u) = 1 - e^{-\frac{u}{\rho_t}\left(\frac{1}{\alpha_1 \lambda_{SR1}} + \frac{1}{\alpha_3 \lambda_{R1D}}\right)} \qquad u > 0. \tag{30}$$

Substituting into (26), the ergodic capacity of s_1 is then obtained as

$$\mathbb{E}\{C_{s_1}\} = \frac{1}{2 \ln 2} \int_0^\infty \frac{e^{-\frac{u}{\rho_t}\left(\frac{1}{\alpha_1 \lambda_{SR1}} + \frac{1}{\alpha_3 \lambda_{R1D}}\right)}}{1+U} du. \tag{31}$$

The integration in (31) can be solved using [24, eq. (3.352.4)], resulting in

$$\begin{aligned} \mathbb{E}\{C_{s_1}\} = & -\frac{1}{2 \ln 2} e^{\frac{-1}{\rho_t}\left(\frac{1}{\alpha_1 \lambda_{SR1}} + \frac{1}{\alpha_3 \lambda_{R1D}}\right)} \\ & \times Ei\left(\frac{-1}{\rho_t}\left(\frac{1}{\alpha_1 \lambda_{SR1}} + \frac{1}{\alpha_3 \lambda_{R1D}}\right)\right). \end{aligned} \tag{32}$$

where $Ei(.)$ denotes exponential integral function [24, eq. (8.211.1)].

B. Ergodic data rate of symbol s_2

Similar to the aforementioned analysis, by substituting (10), (13) and (16) into the second term of (20), the instantaneously achievable data rate of symbol s_2 can be expressed as

$$C_{s_2} = \frac{1}{2} \log_2 \left(1 + \min\left\{\gamma_{s_2 \to R_1}, \gamma_{s_2 \to R_2}, \gamma_{s_2 \to D}\right\}\right). \tag{33}$$

Similar to (26), (33) can be solved by letting $V = \min\{\gamma_{s_2 \to R_1}, \gamma_{s_2 \to R_2}, \gamma_{s_2 \to D}\}$ and considering order statistics, the CDF of random variable V is given by:

$$F_V(v) = 1 - F_{\gamma_{s_2 \to R_1}}(v) F_{\gamma_{s_2 \to R_2}}(v) F_{\gamma_{s_2 \to D}}(v). \tag{34}$$

The CDFs of $\gamma_{s_2 \to R_1}$, $\gamma_{s_2 \to R_2}$ and $\gamma_{s_2 \to D}$ can be evaluated, respectively, as

$$F_{\gamma_{s_2 \to R_1}}(v) = 1 - e^{-\frac{v}{\alpha_2 - v\alpha_1} \frac{1}{\rho_t \lambda_{SR1}}} \qquad 0 < v < \frac{\alpha_2}{\alpha_1}, \tag{35}$$

$$F_{\gamma_{s_2 \to R_2}}(v) = 1 - e^{-\frac{v}{\alpha_2 - v\alpha_1} \frac{1}{\rho_t \lambda_{SR2}}} \qquad 0 < v < \frac{\alpha_2}{\alpha_1} \tag{36}$$

and

$$\begin{aligned} F_{\gamma_{s_2 \to D}}(v) = & 1 - \frac{\alpha_4 \rho_t \lambda_{R2D}}{\alpha_4 \rho_t \lambda_{R2D} + \alpha_3 \rho_t \lambda_{R1D}} \\ & \times e^{-\frac{v}{\alpha_4 \rho_t \lambda_{R2D}}} \qquad v > 0. \end{aligned} \tag{37}$$

Therefore, the CDF of random variable V can be obtained by substituting (35), (36) and (37) into (34) resulting in

$$\begin{aligned} F_V(v) = & 1 - \frac{\alpha_4 \rho_t \lambda_{R2D}}{\alpha_4 \rho_t \lambda_{R2D} + v\alpha_3 \rho_t \lambda_{R1D}} \times e^{-\frac{v}{\alpha_4 \rho_t \lambda_{R2D}}} \\ & \times e^{-\frac{v}{\alpha_2 - v\alpha_1}\left(\frac{1}{\rho_t \lambda_{SR1}} + \frac{1}{\rho_t \lambda_{SR2}}\right)} \quad 0 < v < \frac{\alpha_2}{\alpha_1}. \end{aligned} \tag{38}$$

The evaluation of the integral resulting from substituting (38) into (24) is analytically complex. Instead, the ergodic capacity of symbol s_2 is evaluated asymptotically for high SNR values. When $\rho_t \to \infty$, the asymptotic ergodic rate of symbol s_2 can be obtained as

$$\begin{aligned} \mathbb{E}\{C_{s_2}\} = & \frac{1}{2 \ln 2} \frac{\alpha_4 \lambda_{R2D}}{\alpha_4 \lambda_{R2D} - \alpha_3 \lambda_{R1D}} \\ & \times \ln\left(\frac{\alpha_4 \lambda_{R2D} + \frac{\alpha_2}{\alpha_1} \alpha_4 \lambda_{R2D}}{\alpha_4 \lambda_{R2D} + \frac{\alpha_2}{\alpha_1} \alpha_3 \lambda_{R1D}}\right). \end{aligned} \tag{39}$$

TABLE I
SIMULATION PARAMETERS, ADOPTED FROM [19]

Parameter	Value
Average noise power	$\sigma^2 = 10^{-10}$ W
BS transmission power	$P_t = 5$ W
Path loss exponent	$n = 3$
Total distance	$r_d = 500$ m

V. SIMULATION AND RESULTS

In this section, system performance and accuracy of the derived analytical expressions are presented. First, Monte Carlo simulations are used to verify the analytical expressions of the ergodic capacity obtained through substituting (32) and (39) into (22). Then, the effects of distances between the source, relays, and destination on the achievable sum rate are investigated using numerical simulations. Simulations are also used to study the impact of the number of available SBS relays in both branches on the system's ergodic capacity. Finally, a comparison between single antenna scenario and multi antenna scenario has been investigated. All the provided results are obtained via Monte Carlo simulations through 10^5 network realizations. The main parameters values used in the simulations are listed in Table I.

In the following simulation examples, a fixed power allocation strategy is adopted, in which $\alpha_1 = 0.05$, $\alpha_2 = 0.45$, $\alpha_3 = 0.05$ and $\alpha_4 = 0.05$.

Fig. 2 shows the achievable ergodic sum rate of the proposed system, in bps/Hz, versus the transmit signal-to-noise ratio, $\rho_t = P_t/\sigma^2$. Although the range of ρ_t may seem large, it is typical for the transmit SNR, as it would be dissipated at the destination due to path loss. As expected, increasing the transmit SNR results in an increase in the achievable ergodic sum rate. As can be seen in Fig. 2, analytical results are in agreement with the Monte Carlo simulation results at high transmit SNR values, *i.e.*, for $\rho_t > 100$ dB. However, the use of the asymptotic equation in (39) causes a slight difference between the analytical and simulation results at low transmit SNR values.

The effect of the placements of the selected SBS relays, R_1 and R_2, is investigated in Fig. 3. Three different placement scenarios are examined, with total sum distances of the source-relay and relay-destination links fixed at $r_d = d_{SR_i} + d_{R_i D} = 500$ m. To keep the distances between the two relays varied, it is also assumed that distance d_{SR_1} is 50 m less that the distance d_{SR_2}. In the first scenario, both relays are closer to the source than to the destination with $d_{SR_1} = 50$ m and $d_{SR_2} = 100$ m. In the second scenario, both relays are near to the source and destination's midpoint with $d_{SR_1} = 250$ m and $d_{SR_2} = 300$ m. Finally, the last scenario with $d_{SR_1} = 400$ m and $d_{SR_2} = 450$ m, both relays are closer to the destination than to the source.

As can be seen from Fig. 3, the scenario in which both relays are near the midpoint between the source and the destination achieves a higher sum data rate than the other two scenarios. This occurs because the end-to-end capacity is dominated

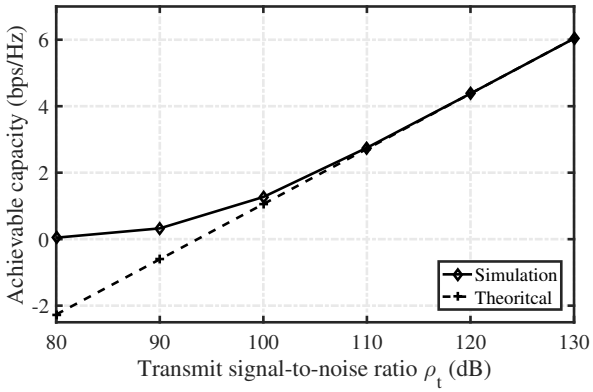

Fig. 2. Achievable ergodic sum rate of the system model. Solid markers shown analytical results and open markers show simulation results.

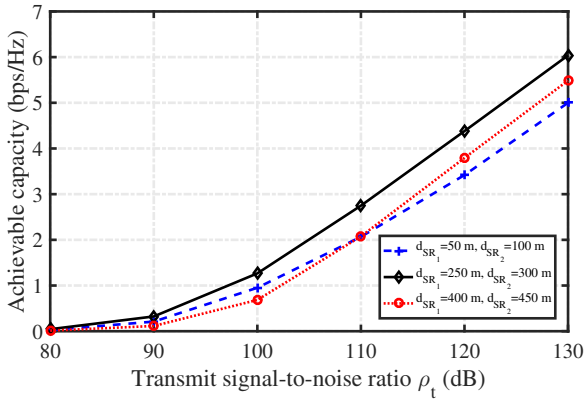

Fig. 3. Achievable sum rate for various relays' positions.

by the minimum of the source-to-relay and the relay-to-destination link capacities, as shown in (20). Consequently, while a shorter communication link positively impacts the link's capacity, a longer link has a dominantly negative effect on the end-to-end capacity. It should be concluded that both relays R_1 and R_2 should be as close as possible to the source and destination's midpoint. It can be then conjectured that there are optimum locations for relays R_1 and R_2 that result in the maximum possible sum ergodic capacity. This could be used as a roadmap for deploying the SBS in ultra-dense networks in practical systems. Fig. 4 shows the achievable sum rate for different values of the distance, d_{SR_1}. The figure is generated for $\rho_t = 107$ dB. Results show that there is an optimum location of R_1 for which a maximum sum rate can be achieved. It can be seen from Fig. 4 that for the simulated example with the assigned power allocation, the optimum capacity is achieved by $d_{SR_1} = 200 \sim 250$ m.

In the next simulation example, the benefits of relay selection and the effect of the number of relays in the selection pools are investigated. In Fig. 5, a comparison of the achievable sum rate based on the number of available relays is introduced. These relays are clustered relatively close together

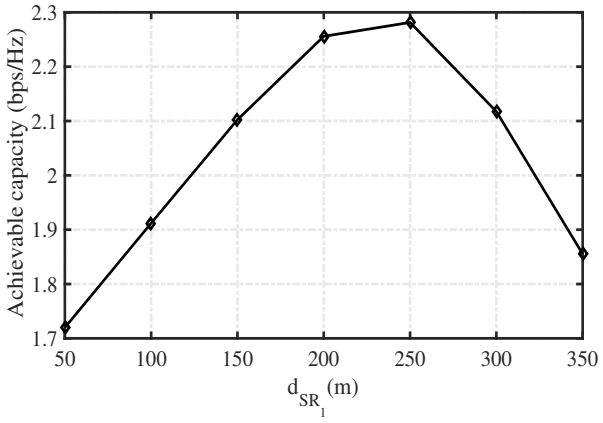

Fig. 4. Achievable sum rate versus various locations of the relay R_1.

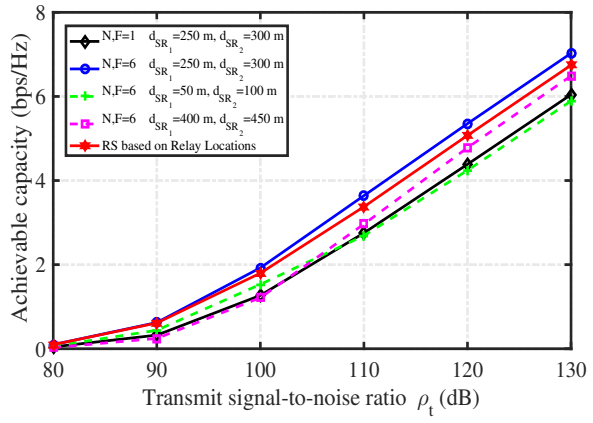

Fig. 6. The effect of number of relays on the achievable capacity.

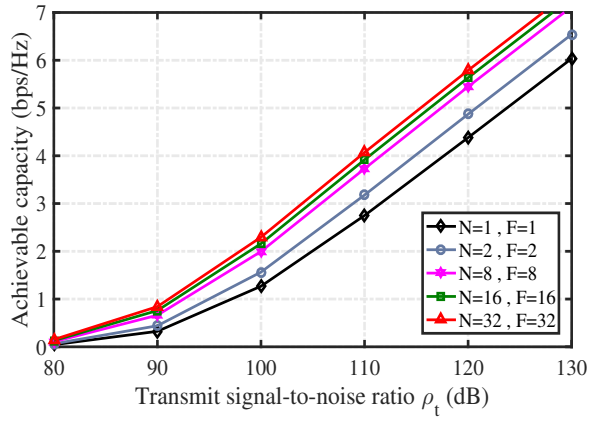

Fig. 5. The achievable sum rate versus the transmit SNR, ρ_t, for various number of relays, $N = F = 1, 2, 8, 16$ and 32.

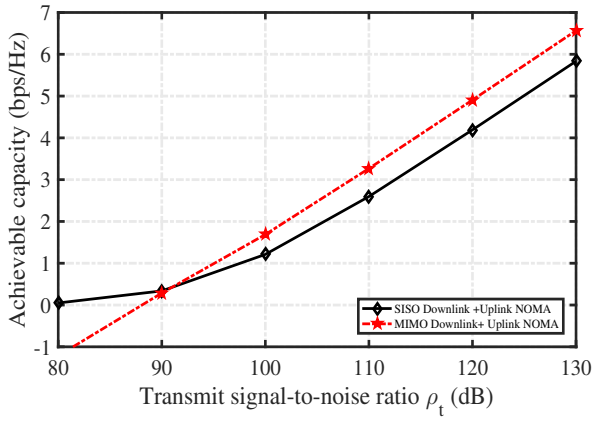

Fig. 7. Comparison of achievable sum-rate while using single antenna versus multi-antenna technology to the proposed system model.

towards the midpoint between the source and the destination with $d_{SR_1} = 250$ m and $d_{SR_2} = 300$ m. The pair of relays that achieve the highest ergodic capacity is selected according to (21). As can be observed, the higher the number of relays, the higher the achievable capacity. It can also be observed that while the achievable capacity improvesas the number of available relays increases, the amount of improvement diminishes for a larger number of relays. For example, at $\rho_t = 110$ dB, the total sum rate achieved is 2.75 bps/Hz for the case when no relay selection is employed. When increasing the number of relays to $N = F = 16$ relays, the overall sum rate increases to 3.917 bps/Hz, representing about 41% improvement in the achievable sum rate. On the other hand, when the number of relays increases to 32, the achievable sum rate is 4.07 bps/Hz, a modest improvement compared to the case of 16 relays. Such observation is helpful during the planning of practical systems while deciding of the density of deployed SBSs in ultra-dense networks.

A more practical scenario is investigated in Fig. 6. In this scenario, the available relays are randomly scattered at various locations, and the relay selection scheme selects the optimal relays location. It can be seen from Fig. that at SNR $\rho_t = 110$ dB, the achievable sum rate for a single pair of relays located at the optimum location between the source and the destination is 2.689 bps/Hz. However, when relay selection is performed from six pairs of relays randomly deployed at various locations in the range $d_{SR_1} \in [50 : 50 : 350]$ m, the achievable sum rate increases to 3.377 bps/Hz. However, if six pairs of relays are all located around the optimum location, the improvement in the data rate is the close to that achieved by randomly deployed relays. It can be concluded then that to benefit from relay selection, it is not necessary to have all relays located at the optimum location, and it is sufficient to use randomly deployed relays.

Fig. 6 shows multiple cases of relay selection compared to each other. Six pairs of relays are relatively clustered around non-optimum locations, either near the source with $d_{SR_1} = 50$ m and $d_{SR_2} = 100$ m or near destination with $d_{SR_1} = 400$ m and $d_{SR_2} = 450$ m. It can be observed that relay selection from a pool of relays located at the optimum location $d_{SR_1} = 200$ m and $d_{SR_2} = 250$ m, results in the highest achievable sum rate.

In addition, randomly deployed relays result in better performance than relay selection from relays located all at a location other than the optimum location. In practical systems, randomly deployed relayed can be used with relay selection to maximize the benefit of NOMA-based wireless backhaul relaying networks. Other relay combining techniques, such as in [25], can be also explored.

Finally in Fig. 7, a comparison of the achievable sum-rate while using single antenna versus multi-antenna technology is shown. In the multi-antenna scheme, the downlink phase is performed using multi-antenna BS equipped with $N_t = 4$ antennas. Similar to the previously mentioned parameters of single antenna scheme, the power coefficients of the downlink phase are considered as follows $\alpha_1 = 0.05$, $\alpha_2 = 0.45$. By substituting (6), (7), and (4) into (20), the hexagram-shaped red line could be obtained, representing the achievable sum-rate of using the MISO downlink phase with the uplink NOMA phase. It is apparent that incorporating multi-antenna networks into the proposed model would improve the achievable sum rate of the wireless backhaul system.

VI. Conclusion

This paper studied a NOMA-based cooperative wireless backhaul network model with a single source, single destination, and multiple randomly deployed decode-and-forward small base station nodes. By utilizing the capabilities of non-orthogonal multiple access and relay selection, the proposed system model can result in higher achievable data rates than those achieved by each alone. The system achievable sum rate was derived, and design guidelines for optimum relay selection policy were provided. Simulation results showed that higher data rates are obtained when the selected pair of relays are such that one is near to the source, and the other is near to the destination. It was also observed that the larger the number of available relays, the higher the achievable data rate. The performance gain obtained from increasing the number of relays at an optimum location has also been investigated. It was shown that such gains diminish with a higher number of relays. The capacity enhancement of enabling multiple input single output technology to the proposed model in the downlink phase has been shown. A future extension of this work will be the joint optimization of relay placement and power allocation of the selected relays. Furthermore, rate splitting non-orthogonal multiple access will be investigated.

References

[1] F. Tariq, M. R. Khandaker, K. K. Wong, M. A. Imran, M. Bennis, and M. Debbah, "A speculative study on 6G," *IEEE Wireless Communications*, vol. 27, no. 4, pp. 118–125, 2020.

[2] M. Z. Chowdhury, S. Ahmed, and Y. M. I. N. Jang, "6G Wireless communication systems : Applications, requirements, technologies, challenges, and research directions," *in IEEE Open Journal of the Communications Society*, vol. 1, pp. 957–975, 2020.

[3] A. L. Rezaabad, H. Beyranvand, J. A. Salehi, and M. Maier, "Ultra-Dense 5G small cell deployment for fiber and wireless backhaul-Aware infrastructures," *IEEE Transactions on Vehicular Technology*, vol. 67, no. 12, pp. 12231–12243, 2018.

[4] B. Tezergil and E. Onur, "Wireless backhaul in 5G and beyond: Issues, challenges and opportunities," *arXiv preprint arXiv:2103.08234*, 2021.

[5] S. Khaled, S. S. Soliman, and Y. Fahmy, "On the performance of power allocation schemes in ultra dense networks," in *14th International Conference on Computer Engineering & Systems (ICCES)*, pp. 419–424, 2019.

[6] M. Xu, F. Ji, M. Wen, and W. Duan, "Novel receiver design for the cooperative relaying system with non-orthogonal multiple access," *IEEE Communications Letters*, vol. 20, no. 8, pp. 1679–1682, 2016.

[7] L. Dai, B. Wang, Z. Ding, Z. Wang, S. Chen, and L. Hanzo, "A survey of non-orthogonal multiple access for 5G," *IEEE Communications Surveys and Tutorials*, vol. 20, no. 3, pp. 2294–2323, 2018.

[8] I. Budhiraja, N. Kumar, S. Tyagi, S. Tanwar, Z. Han, D. Y. Suh, and M. J. Piran, "A systematic review on NOMA variants for 5G and beyond," *IEEE Access*, vol. 9, pp. 85573-85644, 2021.

[9] M. Al-Imari, P. Xiao, M. A. Imran, and R. Tafazolli, "Uplink non-orthogonal multiple access for 5G wireless networks," *11th International Symposium on Wireless Communications Systems(ISWCS)*, pp.781–785, 2014.

[10] S. H. Amin, A. H. Mehana, Y. A. Fahmy, and S. S. Soliman, "Maximizing the number of users in clustered MIMO-NOMA systems under rate constraints," *Mobile Networks and Applications*, vol. 24, no. 2, pp. 618–629, 2019.

[11] S. H. Amin, A. H. Mehana, S. S. Soliman, and Y. A. Fahmy, "User capacity in downlink MISO-NOMA systems," in *IEEE Global Communications Conference (GLOBECOM)*, pp. 1–7, 2018.

[12] M. Zeng, W. Hao, O. A. Dobre, and Z. Ding, "Cooperative NOMA: state of the art, key techniques, and open challenges," *IEEE Network*, vol. 34, no. 5, pp. 205–211, 2020.

[13] P. Xu, Z. Yang, Z. Ding, and Z. Zhang, "Optimal relay selection schemes for cooperative NOMA," *IEEE Transactions on Vehicular Technology*, vol. 67, no. 8, pp. 7851–7855, 2018.

[14] K. Cao, B. Wang, H. Ding, T. Li, and F. Gong, "Optimal relay selection for Secure NOMA systems under untrusted users," *IEEE Transactions on Vehicular Technology*, vol. 69, no. 2, pp. 1942–1955, 2020.

[15] Y. Liu, S. Zhang, X. Mu, Z. Ding, R. Schober, N. Al-Dhahir, E. Hossain, and X. Shen, "Evolution of NOMA toward next generation multiple access (NGMA) for 6G," *IEEE Journal on Selected Areas in Communications*, pp. 1–1, 2022.

[16] G. Chen, L. Qiu, and C. Ren, "On the performance of cluster-Based MIMO-NOMA in multi-cell dense networks," *IEEE Transactions on Communications*, vol. 68, no. 8, pp. 4773–4787, 2020.

[17] Z. Mobini, M. Mohammadi, H. A. Suraweera, and Z. Ding, "Full-Duplex multi-antenna relay assisted cooperative non-orthogonal multiple access," *IEEE Global Communications Conference (GLOBECOM)*, pp. 1–7, 2017.

[18] Q. Zhang, Q. Li, and J. Qin, "Robust beamforming for non-orthogonal multiple access systems in MISO channels," *IEEE Transactions on Vehicular Technology*, vol. 65, no. 12, pp. 10231–10236, 2016.

[19] Q. Zhang, K. Luo, W. Wang, and T. Jiang, "Joint C-OMA and C-NOMA wireless backhaul scheduling in heterogeneous ultra dense networks," *IEEE Transactions on Wireless Communications*, vol. 19, no. 2, pp. 874–887, 2020.

[20] M. Raithatha, A. U. Chaudhry, R. H. Hafez, and J. W. Chinneck, "Locating gateways for maximizing backhaul network capacity of 5G ultra-Dense networks," *Wireless Telecommunications Symposium*, pp. 1–6, 2020.

[21] X. Yang, C. Hua, W. Xu, and P. Gu, "Optimal power allocation for non-orthogonal multiple access in wireless backhaul networks," *IET Communications*, vol. 14, no. 10, pp. 1650–1657, 2020.

[22] S. S. Soliman and N. C. Beaulieu, "Dual-hop Vs multihop AF relaying systems," in *IEEE Global Communications Conference (GLOBECOM)*, pp. 4299–4305, 2013.

[23] Qi, Yue and Zhang, Xinliang and Vaezi, Mojtab,, "Over-the-Air implementation of NOMA: new experiments and future directions, *IEEE Access*, vol. 9, pp. 135828-135844, 2021.

[24] D. Zwillinger, *Table of Integrals, Series, and Products: Eighth Edition*. Academic Press, 2014.

[25] S. S. Soliman, "MRC and selection combining in dual-hop af systems with rician fading," in *10th International Conference on Computer Engineering & Systems (ICCES)*, pp. 314–320, 2015.