Energy Efficient Framework for Multiuser Downlink MIMO-NOMA Systems

Abdelsalam Sayed-Ahmed*, Maha Elsabrouty*, Ahmed H. Abd El-Malek*, Mohammed Abo-Zahhad*†
*Department of Electronics and Communications Engineering
Egypt-Japan University of Science and Technology
New Borg El-Arab City, Alexandria, Egypt
E-mails: {abdelsalam.ahmed, maha.elsabrouty, ahmed.abdelmalek, mohammed.zahhad}@ejust.edu.eg
†Department of Electrical and Electronics Engineering, Assiut University, Assiut, Egypt

***Abstract*—Non-Orthogonal Multiple Access (NOMA) is a promising multiple access technique for the 5th generation (5G) mobile communication systems. In this paper, an energy efficient power allocation scheme is derived for a general number of users per cluster multiple input multiple output (MIMO) downlink NOMA system. The proposed scheme idea is based on converting the difficult energy efficiency power allocation problem into an equivalent spectral efficiency power allocation problem and then dividing this equivalent problem into multiple simple cluster sum rate maximization problems. In addition, a user clustering scheme is proposed to maximize the NOMA system energy efficiency by first designing the detection vectors to convert the MIMO users channel matrices into their equivalent channel vectors and distribute the users on the clusters based on their equivalent channel gains. The users are clustered such that those with the largest equivalent channel gains are selected as cluster heads and the rest of users in each cluster are selected to maximize the equivalent channel gain difference between each other. Numerical results show that the proposed framework improves the system energy efficiency of the MIMO NOMA system at different values of the total transmit power, the minimum required date rates, the number of users per cluster and at different users' distance distribution scenarios.**

Keywords*—Non-Orthogonal Multiple Access (NOMA); Multiple-Input Multiple-Output (MIMO); power allocation; energy efficiency, 5th generation (5G) mobile communication systems

I. Introduction

Fifth generation mobile communications system (5G) has evolved recently to catch up the massive increase in the data traffic, which is expected to be 1000-fold by 2020 [1]. In order to achieve the huge required system capacity, non-orthogonal multiple access (NOMA) is proposed as a multiple access technique candidate for 5G communications system [2]. The power-domain NOMA, or here after called NOMA, utilizes different power levels to serve several users, simultaneously, over the same channel. Consequently, NOMA boosts the individual users' data rates and the number of served users within the same amount of available radio resources. In the downlink transmission, the NOMA transmitter sends the message signals using superposition coding by scaling each user's signal with its power allocation factor and adding the scaled signals together. The NOMA receivers extract their signals using successive interference cancellation (SIC). When comparing NOMA with conventional orthogonal multiple access (OMA) techniques, NOMA can achieve higher spectral efficiency (SE) than OMA [3]. Early research efforts in NOMA were directed mainly towards the spectral efficiency improvement for the single-input single-output (SISO) NOMA system. In [4], the effect of user pairing strategy on two-user SISO NOMA systems was studied. In [5], the analytical and numerical results proved that NOMA attains greater ergodic capacity than OMA when considering a cellular downlink scenario with randomly deployed users.

Recently, research attention is drawn significantly towards integrating MIMO and NOMA concepts to further boost the spectral efficiency. In MIMO-NOMA systems, users are usually clustered together to minimize the SIC detection complexity at the NOMA receivers. In order to remove the inter-cluster interference, precoding and detection vectors should be designed such as users in the same cluster share a common precoding vector while each user in each cluster has a unique detection vector. In [6]–[8], power allocation and user selection schemes were proposed for the maximization of the spectral efficiency of the multi-user multiple-input single-output (MU-MISO) case assuming that there are two users per cluster. For the MISO-NOMA case, detection vectors are not required and only the precoding vectors are designed using zero-forcing beamforming (ZFBF) based on the channel vectors of the cluster heads. The proposed schemes in [6], [7] tried to maximize the system sum rate while the work proposed in [8] aimed to conserve the fairness among selected users when maximizing the system sum rate. The work in [9] and [10] demonstrate that MIMO-NOMA attains larger system sum rate than MIMO-OMA when taking into consideration the two users per cluster scenario. The work in [11] and [12] extend the validation of the performance superiority of MIMO-NOMA on MIMO-OMA to the multiple users per cluster scenario. In [13], the precoding design for MIMO NOMA system with multiple cells is investigated. simulation results indicate that the proposed precoding design for NOMA in [13] maximizes both the edge users' individual data rates and the system sum rate compared with conventional OMA techniques.

In addition to spectral efficiency maximization, optimizing the energy efficiency (EE) has attracted also remarkable attention because of the global concerns for reducing the energy consumption. In [14], energy efficiency optimization was carried out in a fading MIMO-NOMA system scenario taking into account the existence of only two users in each cluster. The work in [15] proposed joint subchannel assignment and power allocation to enhance the energy efficiency for a two users per cluster multi-carrier NOMA system. Their results indicate that NOMA attains higher spectral efficiency and energy efficiency than OMA. Therefore, the proposed power allocation schemes in [14] and [15] are applicable only for the two-users per cluster scenario. In [16], the energy efficiency is investigated for a single carrier multi-user NOMA system with a minimum quality of service (QoS) data rate requirement for each user and a power allocation algorithm was proposed based on non-convex fractional programming. In both [15] and [16], SISO-NOMA scenario is considered, which is a large defect in these schemes since future communication systems depend on the MIMO structures. In [17], the energy efficiency maximization problem is investigated for a multiple users per cluster MIMO-NOMA system considering a minimum QoS data rate constraint for each user. A power allocation strategy among and inside clusters was proposed to solve the energy efficiency maximization problem when it is feasible. When the problem is infeasible, the work in [17] proposed a user admission protocol, which admits users one by one following the ascending order of the required transmit power to satisfy the minimum QoS data rate requirements. The main drawback in [17] is considering the power allocation process among and inside clusters as one problem, which increases the complexity of solving it analytically. The power allocation scheme proposed in [17] includes some computationally extravagant operations such as searching for the roots of the derivative of the obtained energy efficiency expression.

The proposed work in this paper considers the multiple users per cluster MIMO-NOMA scenario. We try to simplify the power allocation problem considered in [17] by converting it to its equivalent spectral efficiency problem and then dividing this equivalent problem into an easy to solve sum rate optimization problem in each cluster. Therefore, closed form expressions for the power allocation factors are derived to maximize the NOMA system energy efficiency under the minimum QoS data rate constraint at each user.

The contributions of this paper can be listed as follows: First, a user clustering scheme is proposed in which the users are grouped in the clusters based on their equivalent channel gains taking into consideration conserving the channel gain differences among the clustered users. Second, an energy efficient power allocation scheme is derived for a general value of the number of users per cluster. Simulation results show that the proposed scheme for power allocation and user clustering provides system energy efficiency better than that of the OMA system for different values of the total transmit power, the minimum required QoS data rate, the number of users per cluster and different users' distance distribution scenarios.

The rest of this paper is organized as follows: Section II gives a brief description of the considered system model. Section III presents the adopted beamforming method and the proposed user clustering scheme. The signal model and the energy efficiency problem formulation are discussed in section IV. Section V introduces the solution of the energy efficiency problem. In section VI, Numerical results are presented. Finally, conclusions are summarized in section VII.

Notations: Matrices and vectors are written in uppercase and lowercase boldface letters, respectively. $\mathbf{H}^H$ and $\mathbf{h}^H$ stand for the conjugate transpose (Hermitian) of a matrix $\mathbf{H}$ and a vector $\mathbf{h}$, respectively. $|\mathbf{h}|$ stands for the Euclidean vector norm of $\mathbf{h}$.

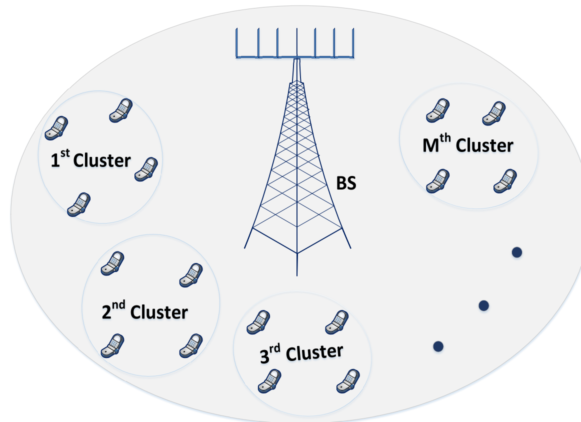

Fig. 1: The system model considered: A single cell with multiple users per cluster employing MIMO-NOMA

II. System Model

The system model considered in this paper is illustrated in Fig.1, which consists of a single cell. A base station (BS) is installed in the center of the cell and equipped with M transmit antennas. The user equipment around the BS are grouped into clusters whose number is less than or equal to the number of BS transmit antennas and each cluster includes L users. Each user equipment has N receive antennas. The users are assumed to be distributed randomly within the cell and will be denoted by (m, l), $m \in 1, ..., M$ and $l \in 1, ..., L$, to refer to the l-th user in the m-th cluster. The distance between each user and the BS is $d_{m,l}$. The matrix of the channel coefficients between the BS transmit antennas and the user equipment (m, l) receive antennas is $\mathbf{H}_{m,l} \in \mathbb{C}^{N \times M}$, $\mathbf{H}_{m,l} = \frac{\mathbf{G}_{m,l}}{\sqrt{P_L}}$, where $\mathbf{G}_{m,l}$ is assumed to be quasi-static independent and identically distributed (i.i.d.) and P_L is the path loss factor, which depends on the distance $d_{m,l}$ and the path loss exponent according to the path loss equation expressed in Table II. The noise is assumed to be Gaussian with zero mean and variance σ_n^2.

III. Beamforming and User Clustering

The design of the precoding matrix $\mathbf{P} \in \mathbb{C}^{M \times M}$ and the detection vectors $\mathbf{v}_{m,l} \in \mathbb{C}^{N \times 1}$ can be done either using the beamforming scheme in [9] or [11]. Both methods apply the concepts of the zero-forcing (ZF) based detection design, which means that the inter-cluster interference can be removed completely even when there exist multiple users in each cluster. The difference between the two methods is the relation between N and M in order to find the detection vectors. In [9], the constraint on the number of receive antennas is: $N \geq \frac{M}{2}$ while in [11], it is $N \geq M$. This means that the condition of the application of the method proposed in [9] is more relaxed than that proposed in [11] at the expense of the computational complexity. In this paper, the beamforming method of [11] is adopted because it is suitable for our proposed clustering scheme. In this method, the precoding matrix is set to be $\mathbf{P} = I_M$, where I_M is the $M \times M$ identity matrix and the detection vector is designed such that: $\mathbf{v}_{m,l}^H \tilde{\mathbf{H}}_{m,l} = 0$ where $\tilde{\mathbf{H}}_{m,l} \in \mathbb{C}^{M \times (M-1)}$ is a submatrix of $\mathbf{H}_{m,l}$ formed by removing the m-th column.

The proposed user clustering scheme can be explained as follows: Assume that there are K candidate users around the BS, each with channel matrix $\mathbf{H}_k$ for $k = 1, ..., K$. The selection process starts by calculating the equivalent channel gain $\left|\mathbf{v}_{m,k}^H \mathbf{H}_k \mathbf{p}_m\right|$ for each user when it is put in each of the M clusters and select the cluster which provides the best equivalent channel gain for each user. Therefore, we have now M groups of users, each of size K_m, containing the users which best fits each cluster. The users in each group are sorted in descending order starting by the largest equivalent channel gain in each group. From each of these M sorted groups, L users are selected to be grouped in each cluster such that the selected users should have the most distinctive equivalent channel gains [18]. This can be obtained by dividing each of the M sorted groups into L sub-groups, each sub-group has a size of $\lfloor \frac{Km}{L} \rfloor$ and selecting the first user from each of the upper $\lceil \frac{L}{2} \rceil$ sub-groups and the last user from each of the lower $\lfloor \frac{L}{2} \rfloor$ sub-groups.

After applying the beamforming and user clustering processes, each of the selected users has an effective channel gain $\left|\mathbf{v}_{m,l}^H \mathbf{H}_{m,l} \mathbf{p}_m\right|$ on which the selected users are ranked in each cluster in order for the SIC detection to be accomplished perfectly where $\mathbf{p}_m$ is the m-th column of $\mathbf{P}$.

$$\begin{gathered} \left|\mathbf{v}_{m,1}^H \mathbf{H}_{m,1} \mathbf{p}_m\right|^2 > ... > \left|\mathbf{v}_{m,L}^H \mathbf{H}_{m,L} \mathbf{p}_m\right|^2 \\ \forall \quad m = 1, ..., M. \end{gathered} \tag{1}$$

IV. Signal Model and Problem Formulation

A. Signal Model

The received signal at the user (m, l) can be expressed as:

$$\begin{gathered} \mathbf{y}_{m,l} = \mathbf{H}_{m,l} \mathbf{P} \mathbf{x} + \mathbf{n}_{m,l} \\ \forall \;\; l = 1, ..., L \;\; and \;\; m = 1, ..., M. \end{gathered} \tag{2}$$

where $\mathbf{y}_{m,l} \in \mathbb{C}^{N \times 1}$ is the user's observation vector and $\mathbf{n}_{m,l} \in \mathbb{C}^{N \times 1}$ is the noise received at the user (m, l). $\mathbf{x}$ is the transmitted signals of all L users in the m-th cluster and is given by:

$$\mathbf{x} = \begin{pmatrix} \sum_{l=1}^{L} \sqrt{\alpha_{1,l} P_1} s_{1,l} \\ ... \\ \sum_{l=1}^{L} \sqrt{\alpha_{m,l} P_m} s_{m,l} \\ ... \\ \sum_{l=1}^{L} \sqrt{\alpha_{M,l} P_M} s_{M,l} \end{pmatrix} \tag{3}$$

where $s_{m,l}$ is the signal intended for user (m, l), $\alpha_{m,l}$ is the power allocation factor for the user (m, l), which is defined here as the ratio of the cluster power allocated to the (m, l) user. This definition of $\alpha_{m,l}$ will serve in the optimization problem formulation break down into a group of per-cluster power allocation optimization problems. This is presented in more details in Section IV-B. P_m is the transmit power allocated to the m-th cluster and $P_{tot} = \sum_{m=1}^{M} P_m$ is the total BS transmit power. User (m, l) multiplies the Hermitian of its detection vector $\mathbf{v}_{m,l}$ by the observation vector and the resulting observation can be rewritten as follows:

$$\begin{aligned} y_{m,l} &= \mathbf{v}_{m,l}^H \mathbf{y}_{m,l} = \mathbf{v}_{m,l}^H \mathbf{H}_{m,l} \mathbf{P} \mathbf{x} + \mathbf{v}_{m,l}^H \mathbf{n}_{m,l} \\ &= \mathbf{v}_{m,l}^H \mathbf{H}_{m,l} \mathbf{p}_m \sum_{l=1}^{L} \sqrt{\alpha_{m,l} P_m} \\ &+ \sum_{k=1, k \neq m}^{M} \mathbf{v}_{m,l}^H \mathbf{H}_{m,l} \mathbf{p}_k \sum_{l=1}^{L} \sqrt{\alpha_{k,l} P_k} + \mathbf{v}_{m,l}^H \mathbf{n}_{m,l} \\ &\forall \;\; l = 1, ..., L \;\; and \;\; m = 1, ..., M. \end{aligned} \tag{4}$$

Due to the the ZF-based detection design, the following condition must be satisfied.

$$\mathbf{v}_{m,l}^H \mathbf{H}_{m,l} \mathbf{p}_k = 0 \quad \forall \quad k \neq m. \tag{5}$$

Hence, the detected observation signal can be written as:

$$\begin{gathered} y_{m,l} = \mathbf{v}_{m,l}^H \mathbf{H}_{m,l} \mathbf{p}_m \sum_{l=1}^{L} \sqrt{\alpha_{m,l} P_m} + \mathbf{v}_{m,l}^H \mathbf{n}_{m,l} \\ \forall \;\; l = 1, ..., L \;\; and \;\; m = 1, ..., M. \end{gathered} \tag{6}$$

Therefore, the data rate of the (m, l) user in the NOMA system can be expressed as:

$$\begin{aligned} R_{m,l} &= \log_2 \left(1 + \frac{\left|\mathbf{v}_{m,l}^H \mathbf{H}_{m,l} \mathbf{p}_m\right|^2 \alpha_{m,l} P_m}{\left|\mathbf{v}_{m,l}^H \mathbf{H}_{m,l} \mathbf{p}_m\right|^2 P_m \sum_{j=1}^{l-1} \alpha_{m,j} + \sigma_n^2} \right) \\ &= \log_2 (1 + \frac{\gamma_{m,l} \alpha_{m,l}}{\gamma_{m,l} \sum_{j=1}^{l-1} \alpha_{m,j} + 1}) \\ &\forall \;\; l = 1, ..., L \;\; and \;\; m = 1, ..., M. \end{aligned} \tag{7}$$

where $\gamma_{m,l} = \frac{\left|\mathbf{v}_{m,l}^H \mathbf{H}_{m,l} \mathbf{p}_m\right|^2 P_m}{\sigma_n^2}$.

For the MIMO-OMA scheme, the same power coefficients are allocated to the L users per cluster as for the case of MIMO-NOMA for the sake of fair comparison [12]. From *theorem 1* in [12], the sum rate in the m-th cluster of the MIMO-OMA system is upper bounded by:

$$R_m^{OMA} \leq \log_2 (1 + \sum_{l=1}^{L} \gamma_{m,l} \alpha_{m,l}) \quad \forall \;\; m = 1, ..., M. \tag{8}$$

B. Energy Efficiency Problem Formulation

The main target of the considered MIMO-NOMA system is to maximize the system energy efficiency, which is defined as the ratio between the system sum rate R_{sum} and the total power consumption under two constraints, which are: 1) The total power constraint and 2) The minimum QoS data rate requirements for each user. The total power consumption is comprised of two parts: the fixed circuit power consumption P_c, and the flexible transmit power $P_t = \sum_{m=1}^{M} \sum_{l=1}^{L} \alpha_{m,l} P_m$. Therefore, the energy efficiency can be defined as [16], [17]:

$$\eta_{EE} = \frac{R_{sum}}{P_c + P_t} \tag{9}$$

The total power constraint will be:

$$\sum_{l=1}^{L} \alpha_{m,l} \leq 1 \quad \forall \ m = 1, ..., M. \tag{10}$$

The minimum required user data rate constraint states that: $R_{m,l} \geqslant \bar{R}_{m,l}$ where $\bar{R}_{m,l}$ is the minimum QoS data rate required for the (m, l) user. Consequently, this constraint can be rewritten as:

$$\log_2(1 + \frac{\gamma_{m,l}\alpha_{m,l}}{\gamma_{m,l}\sum\limits_{j=1}^{l-1}\alpha_{m,l}+1}) \geqslant \bar{R}_{m,l} \qquad \forall \ l = 1, ..., L \ \ and \ \ m = 1, ..., M. \tag{11}$$

Therefore, equation (11) can be reformulated as follows:

$$\frac{\gamma_{m,l}\alpha_{m,l}}{\gamma_{m,l}\sum\limits_{j=1}^{l-1}\alpha_{m,j}+1} \geqslant (\varphi_{m,l} - 1) \qquad \forall \ l = 1, ..., L \ \ and \ \ m = 1, ..., M. \tag{12}$$

where $\varphi_{m,l} = 2^{\bar{R}_{m,l}}$. Hence, the user data rate constraints can be expressed as:

$$\gamma_{m,l}\alpha_{m,l} - (\gamma_{m,l}\sum_{j=1}^{l-1}\alpha_{m,j} + 1)(\varphi_{m,l} - 1) \geqslant 0 \qquad \forall \ l = 1, ..., L \ \ and \ \ m = 1, ..., M. \tag{13}$$

The previous equation can be rewritten as:

$$\alpha_{m,l} \geqslant (\varphi_{m,l} - 1)(\sum_{j=1}^{l-1}\alpha_{m,j} + \frac{1}{\gamma_{m,l}}) \qquad \forall \ l = 1, ..., L \ \ and \ \ m = 1, ..., M. \tag{14}$$

Therefore, $\alpha_{m,l}^{min}$ can be expressed as:

$$\alpha_{m,l}^{min} = (\varphi_{m,l} - 1)(\sum_{j=1}^{l-1}\alpha_{m,j}^{min} + \frac{1}{\gamma_{m,l}}) \qquad \forall \ l = 1, ..., L \ \ and \ \ m = 1, ..., M. \tag{15}$$

Consequently, the minimum power allocation factors $\alpha_{m,l}^{min}$ required to satisfy the minimum QoS data rate requirement for the (m, l) user can be derived by successively substituting for l in (15) as shown in more details in Table I in the necessary conditions column.

Therefore, the minimum required transmit power can be written as:

$$\bar{P}_m = \sum_{l=1}^{L}\alpha_{m,l}^{min}P_m \quad \forall \ m = 1, ..., M. \tag{16}$$

and the minimum required BS total transmit power can be expressed as:

$$P_{req} = \sum_{m=1}^{M}\sum_{l=1}^{L}\alpha_{m,l}^{min}P_m \tag{17}$$

Therefore, the optimization problem can be formulated as:

$$\begin{aligned} \max_{\alpha_{m,l}} \quad & \eta_{EE} \\ s.t. \quad & \sum_{l=1}^{L}\alpha_{m,l} \leq 1 \quad \forall \ m = 1, ..., M. \\ & R_{m,l} \geqslant \bar{R}_{m,l} \quad \forall \ l = 1, ..., L \ \ and \ \ m = 1, ..., M. \end{aligned} \tag{18}$$

The optimization problem in (18) is feasible in the m-th cluster when $\bar{P}_m \leq P_m$.

V. Energy efficiency Problem Solution

Case I: *When (18) is feasible.*

When (18) is feasible, the energy efficiency maximization problem can be replaced by their corresponding spectral efficiency maximization problems in all the M clusters R_m^{NOMA} under any given transmit power $P_f^{(m)}$, $P_f^{(m)} \in [\bar{P}_m, P_m]$ for all $m = 1, ..., M$, and then we have to select the appropriate values of $P_f^{(m)}$ [17].

$$\begin{aligned} \max_{\alpha_{m,l}} \quad & R_m^{NOMA} = \sum_{l=1}^{L}R_{m,l} \quad \forall \ m = 1, ..., M. \\ s.t. \quad & \sum_{l=1}^{L}\alpha_{m,l}P_m = P_f^{(m)} \\ & R_{m,l} \geqslant \bar{R}_{m,l} \qquad \forall \ l = 1, ..., L \end{aligned} \tag{19}$$

By substituting from (7) and (13) in (19), we can get the final form of the optimization problem, which will be solved at each cluster:

$$\begin{aligned} \max_{\alpha_{m,l}} & \sum_{l=1}^{L}\log_2(1 + \frac{\gamma_{m,l}\alpha_{m,l}}{\gamma_{m,l}\sum\limits_{j=1}^{l-1}\alpha_{m,j} + 1}) \quad \forall \ m = 1, ..., M. \\ s.t. \quad & \beta_m - \sum_{l=1}^{L}\alpha_{m,l} = 0 \\ & \gamma_{m,l}\alpha_{m,l} - (\gamma_{m,l}\sum_{j=1}^{l-1}\alpha_{m,j} + 1)(\varphi_{m,l} - 1) \geqslant 0 \\ & \forall \ l = 1, ..., L \end{aligned} \tag{20}$$

where $\beta_m = \frac{P_f^{(m)}}{P_m}$.

The Lagrangian form of this optimization problem at the m-th cluster can be expressed as:

$$\begin{aligned} \mathfrak{L}(\alpha, \lambda, \mu) = & \sum_{l=1}^{L}\log_2(1 + \frac{\gamma_{m,l}\alpha_{m,l}}{\gamma_{m,l}\sum\limits_{j=1}^{l-1}\alpha_{m,j}+1}) + \lambda(\beta_m - \sum_{l=1}^{L}\alpha_{m,l}) \\ & + \sum_{l=1}^{L}\mu_l\left\{\gamma_{m,l}\alpha_{m,l} - (\gamma_{m,l}\sum_{j=1}^{l-1}\alpha_{m,j} + 1)(\varphi_{m,l} - 1)\right\} \end{aligned} \tag{21}$$

where λ and μ_l are the Lagrange multipliers, $\forall\ l = 1, ..., L$. The Karush-Kuhn-Tucker (KKT) conditions are given as,

$$\frac{\partial \mathfrak{L}}{\partial \alpha^*_{m,1}} = 0 \tag{22}$$
$$\frac{\partial \mathfrak{L}}{\partial \alpha^*_{m,l}} = 0 \qquad \forall\ l = 2, ..., L. \tag{23}$$
$$\lambda^* \frac{\partial \mathfrak{L}}{\partial \lambda^*} = 0 \tag{24}$$
$$\mu^*_l \frac{\partial \mathfrak{L}}{\partial \mu^*_l} = 0 \qquad \forall\ l = 1, ..., L. \tag{25}$$
$$\mu^*_l \geqslant 0 \qquad \forall\ l = 2, ..., L. \tag{26}$$

By taking the derivatives of the Lagrangian with respect to the power allocation factors and Lagrange multipliers, a system of algebraic equations is obtained as follows.

$$\begin{aligned}\frac{\partial \mathfrak{L}}{\partial \alpha^*_{m,l}} &= \frac{1}{\ln(2)} \left\{ \frac{\gamma_{m,l}}{\gamma_{m,l} \sum_{j=1}^{l} \alpha_{m,j} + 1} \right\} - \lambda + \mu_l \gamma_{m,l} \\ &+ \frac{1}{\ln(2)} \left\{ \sum_{j=l+1}^{L} \frac{\gamma^2_{m,j} \alpha_{m,j}}{(\gamma_{m,j} \sum_{i=1}^{j} \alpha_{m,i} + 1)(\gamma_{m,j} \sum_{i`=1}^{j-1} \alpha_{m,i`} + 1} \right\} \\ &- \sum_{j=l+1}^{L} \mu_j \gamma_{m,j} (\varphi_{m,j} - 1) = 0 \quad \forall\ l = 1, ..., L. \end{aligned} \tag{27}$$

$$\lambda^* \frac{\partial \mathfrak{L}}{\partial \lambda^*} = \lambda^* (\beta_m - \sum_{l=1}^{L} \alpha_{m,l}) = 0 \tag{28}$$

$$\mu^*_l \frac{\partial \mathfrak{L}}{\partial \mu^*} = \mu^*_l (\gamma_{m,l} \alpha_{m,l} - (\gamma_{m,l} \sum_{j=1}^{l-1} \alpha_{m,j} + 1)(\varphi_{m,l} - 1)) = 0 \quad \forall\ l = 1, ..., L. \tag{29}$$

$$\mu^*_l \geqslant 0 \quad \forall\ l = 2, ..., L. \tag{30}$$

By solving the linear equations (28) to (30), the power allocation factors will be obtained. Equation (27) can be used to obtain the values of the Lagrange multipliers. Table I shows the optimal power allocation factors expressions for the following values of the number of users per cluster: $L = 2, 3, 4$. Note that the parameters $\alpha_{m,l}$, $\gamma_{m,l}$ and $\varphi_{m,l}$ are replaced by α_l, γ_l and φ_l in Table I due to the limited space. By using the mathematical induction, the general form of the optimal power allocation factors can be obtained as:

$$\begin{aligned}\alpha_{m,1} &= \frac{\beta_m}{\prod_{j=2}^{L} \varphi_{m,j}} - \sum_{j=2}^{L} \frac{\varphi_{m,j} - 1}{\gamma_{m,j} \prod_{k=2}^{j} \varphi_{m,k}} \\ \alpha_{m,l} &= \frac{(\varphi_{m,l} - 1)\beta_m}{\prod_{j=l}^{L} \varphi_{m,j}} - \sum_{j=l}^{L} \frac{(\varphi_{m,j} - 1)(\varphi_{m,l} - 1)}{\gamma_{m,j} \prod_{k=l}^{j} \varphi_{m,k}} \\ &+ \frac{\varphi_{m,l} - 1}{\gamma_{m,l}} \qquad \forall\ l = 2, ..., L. \end{aligned} \tag{31}$$

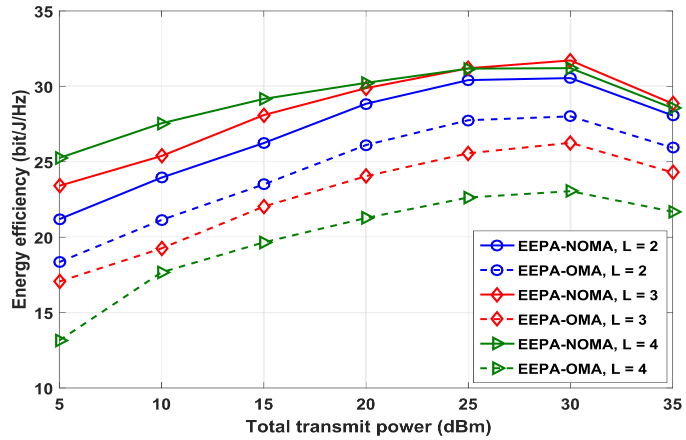

Fig. 2: The variation of the energy efficiency with the total transmit power at different values of the number of users per cluster: $d_{m,1} = 40$ m, $d_{m,2} = 80$ m, $d_{m,3} = 120$ m, $d_{m,4} = 160$ m, $\bar{R}_{m,l} = R_{min} = 2$ bps/Hz.

The value of the system energy efficiency is dependent on the power allocation factors, which in turn depends on the value of $P^{(m)}_f$ within the interval $[\bar{P}_m, P_m]$ for all $m = 1, ..., M$. To optimally select this value, we have to search for the optimal value of $P^{(m)}_f$. In order to reduce the search complexity, we divide the search interval equally into T power points and search for which one of them that will provide the largest energy efficiency starting from the minimum interval limit.

Case II: *When (18) is infeasible.*

In this case, the clusters in which the optimization problem is infeasible will turn on to the MIMO-OMA case with equal power allocation factors. Therefore, the upper limit of the m-th cluster sum rate in this case, R^{inf}_m, can be written as:

$$R^{inf}_m \leq \log_2(1 + \tfrac{1}{L} \sum_{l=1}^{L} \gamma_{m,l}) \quad \forall\ infeasible\ clusters. \tag{32}$$

and the transmit power for each of the infeasible clusters will be P_m. The system sum rate and the total transmit power will be computed by accumulating the sum rates and transmit powers of both the feasible and infeasible clusters.

VI. Simulation Results

In this section, simulation results are presented to verify the advantage of the energy efficiency performance of the proposed MIMO-NOMA power allocation scheme "EEPA-NOMA" over that of the MIMO-OMA case "EEPA-OMA". The values of the simulation parameters are presented in Table II [12]. The MIMO-OMA energy efficiency curves are obtained by accumulating the MIMO-OMA cluster rates calulated using (8) and then dividing the result by the sum of the fixed circuit power and the total BS transmit power. Comparisons are held versus the total available BS transmit power, minimum QoS data rate, the number of users per cluster and the distance distribution of the users in the clusters. In order to fairly measure the energy efficiency of the MIMO-NOMA system using simulations, the energy efficiency of the trials in which the total transmit power cannot support the minimum QoS data rate requirements for all users is set to zero.

TABLE I: Optimal power allocation factors and corresponding necessary conditions for the 2, 3, and 4 users per cluster MIMO-NOMA scenarios.

Cluster	Optimal power allocation factors	Necessary conditions
2-user	$\alpha_1 = \frac{1}{\varphi_2}(\beta_m - \frac{\varphi_2-1}{\gamma_2})$ $\alpha_2 = \frac{\varphi_2-1}{\varphi_2}(\beta_m + \frac{1}{\gamma_2})$	$\frac{(\varphi_1-1)}{\gamma_1} < \alpha_1 < \alpha_2$ $\frac{(\varphi_2-1)(\varphi_1-1)}{\gamma_1} + \frac{(\varphi_2-1)}{\gamma_2} < \alpha_2 < 1$
3-user	$\alpha_1 = \frac{\beta_m}{\varphi_2\varphi_3} - \frac{\varphi_2-1}{\varphi_2\gamma_2} - \frac{\varphi_3-1}{\varphi_2\varphi_3\gamma_3}$ $\alpha_2 = \frac{(\varphi_2-1)\beta_m}{\varphi_2\varphi_3} + \frac{\varphi_2-1}{\varphi_2\gamma_2} - \frac{(\varphi_2-1)(\varphi_3-1)}{\varphi_2\varphi_3\gamma_3}$ $\alpha_3 = \frac{(\varphi_3-1)\beta_m}{\varphi_3} + \frac{\varphi_3-1}{\varphi_3\gamma_3}$	$\frac{(\varphi_1-1)}{\gamma_1} < \alpha_1 < \alpha_2$ $\frac{(\varphi_2-1)(\varphi_1-1)}{\gamma_1} + \frac{(\varphi_2-1)}{\gamma_2} < \alpha_2 < \alpha_3$ $\frac{(\varphi_3-1)\varphi_2(\varphi_1-1)}{\gamma_1} + \frac{(\varphi_3-1)(\varphi_2-1)}{\gamma_2} + \frac{(\varphi_3-1)}{\gamma_3} < \alpha_3 < 1$
4-user	$\alpha_1 = \frac{\beta_m}{\varphi_2\varphi_3\varphi_4} - \frac{\varphi_2-1}{\varphi_2\gamma_2} - \frac{\varphi_3-1}{\varphi_2\varphi_3\gamma_3} - \frac{\varphi_4-1}{\varphi_2\varphi_3\varphi_4\gamma_4}$ $\alpha_2 = \frac{(\varphi_2-1)\beta_m}{\varphi_2\varphi_3\varphi_4} + \frac{\varphi_2-1}{\varphi_2\gamma_2} - \frac{(\varphi_2-1)(\varphi_3-1)}{\varphi_2\varphi_3\gamma_3} - \frac{(\varphi_2-1)(\varphi_4-1)}{\varphi_2\varphi_3\varphi_4\gamma_4}$ $\alpha_3 = \frac{(\varphi_3-1)\beta_m}{\varphi_3\varphi_4} + \frac{\varphi_3-1}{\varphi_3\gamma_3} - \frac{(\varphi_3-1)(\varphi_4-1)}{\varphi_3\varphi_4\gamma_4}$ $\alpha_4 = \frac{(\varphi_4-1)\beta_m}{\varphi_4} + \frac{\varphi_4-1}{\varphi_4\gamma_4}$	$\frac{(\varphi_1-1)}{\gamma_1} < \alpha_1 < \alpha_2$ $\frac{(\varphi_2-1)(\varphi_1-1)}{\gamma_1} + \frac{(\varphi_2-1)}{\gamma_2} < \alpha_2 < \alpha_3$ $\frac{(\varphi_3-1)\varphi_2(\varphi_1-1)}{\gamma_1} + \frac{(\varphi_3-1)(\varphi_2-1)}{\gamma_2} + \frac{(\varphi_3-1)}{\gamma_3} < \alpha_3 < \alpha_4$ $\frac{(\varphi_4-1)\varphi_3\varphi_2(\varphi_1-1)}{\gamma_1} + \frac{(\varphi_4-1)\varphi_3(\varphi_2-1)}{\gamma_2} + \frac{(\varphi_4-1)(\varphi_3-1)}{\gamma_3} + \frac{(\varphi_4-1)}{\gamma_4} < \alpha_4 < 1$

TABLE II: Simulation Parameters.

Parameter	Value
Number of antennas	$M = 3, N = 3$
Circuit power consumption	30 dBm
Channel bandwidth	10 MHz
Thermal noise density	-174 dBm/Hz
Path loss model	$114 + 38\log_{10}(d_{m,l})$, $d_{m,l}$ in km

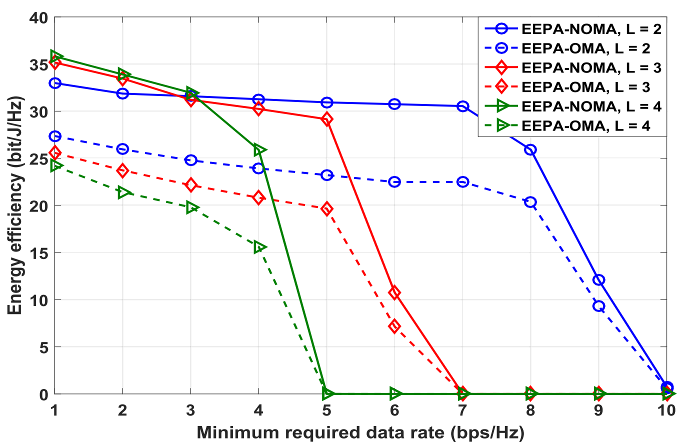

Fig. 3: The variation of the energy efficiency of the MIMO-OMA system with the minimum QoS data rate at different values of the number of users per cluster: $d_{m,1} = 40$ m, $d_{m,2} = 80$ m, $d_{m,3} = 120$ m, $d_{m,4} = 160$ m, $P_{tot} = 25$ dBm.

In Fig.2, the variation of the average system energy efficiency with the total BS transmit power P_{tot} for both the MIMO-NOMA and MIMO-OMA cases at different values of the number of users per cluster L is presented. Results in Fig.2 show that, at all values of L, the proposed MIMO-NOMA power allocation scheme "EEPA-NOMA" provides improved average system energy efficiency performance than that of MIMO-OMA "EEPA-OMA" scheme. As it is observed, the amount of improvement in the energy efficiency of the MIMO-NOMA case over that of the MIMO-OMA case increases with increasing the number of users per cluster from $L = 2$ to $L = 4$ since the available BS transmit power is sufficient enough to serve the minimum data rate requirements of four users per cluster system.

The case when the BS transmit power is not enough for serving the minimum QoS requirements of all users is

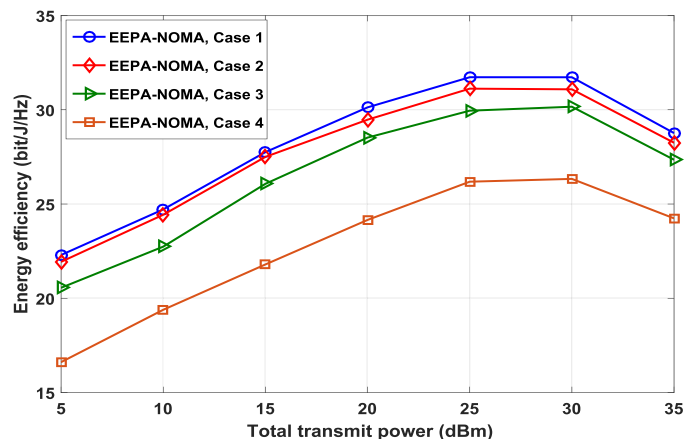

Fig. 4: The variation of the energy efficiency with the total transmit power at $L = 2$ and $R_{min} = 2$ bps/Hz for the following scenarios: Case 1:$d_{m,1} = 40$ m, $d_{m,2} = 50$ m and $d_{av} = 45$ m. Case 2: $d_{m,1} = 40$ m, $d_{m,2} = 55$ m and $d_{av} = 47.5$ m. Case 3: $d_{m,1} = 50$ m, $d_{m,2} = 55$ m and $d_{av} = 52.5$ m. Case 4: $d_{m,1} = 80$ m, $d_{m,2} = 80$ m and $d_{av} = 80$ m.

explained by Fig.3. It presents the variation of the average system energy efficiency with the minimum QoS data rate $R_{m,l} = R_{min}$ for both the MIMO-NOMA and MIMO-OMA cases at different values of the number of users per cluster L. The obtained results indicate that the average system energy efficiency decreases with increasing R_{min} at all values of L because, as R_{min} increases, the system reaches the infeasibility case more rapidly and becomes unable to satisfy R_{min} for all the system users. It is obvious also that the energy efficiency decays more rapidly for large L because the BS transmit power will not be sufficient to serve more users with large R_{min}. Therefore, the system reaches the infeasibility case more rapidly for large L. We can observe also that, at any value of L, the gap between the MIMO-NOMA and MIMO-OMA energy efficiency values shrinks as R_{min} increases because more clusters turns to the infeasibility case with increasing R_{min}.

Fig.4 and Fig.5 show the variation of the average system energy efficiency with the total BS transmit power P_{tot} for the MIMO-NOMA system at different users' distance distribution scenarios for $L = 2$ and $L = 3$, respectively. The obtained results indicate that the average system energy efficiency is affected by the average cluster distance d_{av} to the BS where

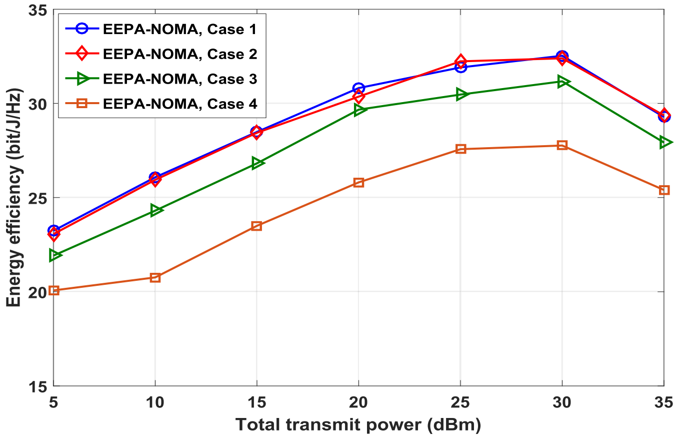

Fig. 5: The variation of the energy efficiency with the total transmit power at $L = 3$ and $R_{min} = 2$ bps/Hz for the following scenarios: Case 1: $d_{m,1} = 40$ m, $d_{m,2} = 50$ m, $d_{m,3} = 60$ m and $d_{av} = 50$ m. Case 2: $d_{m,1} = 40$ m, $d_{m,2} = 55$ m, $d_{m,3} = 70$ m and $d_{av} = 55$ m. Case 3: $d_{m,1} = 50$ m, $d_{m,2} = 55$ m, $d_{m,3} = 60$ m and $d_{av} = 55$ m. Case 4: $d_{m,1} = 80$ m, $d_{m,2} = 80$ m, $d_{m,3} = 80$ m and $d_{av} = 80$ m.

the larger the average distance of the users in the cluster, the lower the resulting energy efficiency. Furthermore, by comparing the scenarios of case 2 and case 3, we can observe that the channel gain of the strongest user in the cluster has the largest impact on the achieved energy efficiency. It is also noticed from the results that the distance difference improves the system energy efficiency because the distance difference helps in differentiating the users at the receiver and facilitate the SIC detection process.

VII. Conclusion

In this paper, energy efficient power allocation and user clustering schemes are proposed for multiple users per cluster downlink MIMO-NOMA system. The proposed power allocation scheme converts the original energy efficiency optimization problem into its equivalent spectral efficiency one and then divides the equivalent optimization problem into multiple ones to be solved at each cluster. The optimization problems at the clusters can be solved using KKT conditions to obtain closed form expressions of the optimal power allocation factors in terms of the selected cluster power. The selected cluster power is optimized by searching for its optimal values that maximize the energy efficiency within its feasible interval. Simulation results verify the improved performance of the proposed MIMO-NOMA power allocation scheme over the MIMO-OMA one for different values of the BS total transmit power, the minimum QoS data rates, the number of users per cluster and at different users' distribution scenarios inside the clusters.

Acknowledgment

This work has been supported by Egypt Japan University of Science and Technology (E-JUST), and the Egyptian Ministry of Higher Education (MOHE). The authors would like to express their gratitude to the National Telecommunication Regulatory Authority (NTRA), Ministry of Communications and Information Technology in Egypt for providing the financial support for this research.

References

[1] H. Kayama and H. Jiang, "Evolution of LTE and New Radio Access Technologies for FRA (Future Radio Access)," in *2014 48th Asilomar Conference on Signals, Systems and Computers*, Nov 2014, pp. 1944–1948.

[2] A. Benjebbour, Y. Saito, Y. Kishiyama, A. Li, A. Harada, and T. Nakamura, "Concept and Practical Considerations of Non-Orthogonal Multiple Access (NOMA) for Future Radio Access," in *2013 International Symposium on Intelligent Signal Processing and Communication Systems*, Nov 2013, pp. 770–774.

[3] Q. C. Li, H. Niu, A. T. Papathanassiou, and G. Wu, "5G Network Capacity: Key Elements and Technologies," *IEEE Vehicular Technology Magazine*, vol. 9, no. 1, pp. 71–78, March 2014.

[4] Z. Ding, P. Fan, and H. V. Poor, "Impact of User Pairing on 5G Non-Orthogonal Multiple Access Downlink Transmissions," *IEEE Transactions on Vehicular Technology*, vol. 65, no. 8, pp. 6010–6023, Aug 2016.

[5] Z. Ding, Z. Yang, P. Fan, and H. V. Poor, "On the Performance of Non-Orthogonal Multiple Access in 5G Systems with Randomly Deployed Users," *IEEE Signal Processing Letters*, vol. 21, no. 12, pp. 1501–1505, Dec 2014.

[6] S. Liu, C. Zhang, and G. Lyu, "User Selection and Power Schedule for Downlink Non-Orthogonal Multiple Access (NOMA) System," in *2015 IEEE International Conference on Communication Workshop (ICCW)*, June 2015, pp. 2561–2565.

[7] A. Sayed-Ahmed and M. Elsabrouty, "Capacity-Based User Selection Algorithm for Downlink Beamforming Non-Orthogonal Multiple Access System," in *2017 24th International Conference on Telecommunications (ICT)*, May 2017, pp. 1–5.

[8] ——, "User Selection and Power Allocation for Guaranteed SIC Detection in Downlink Beamforming Non-Orthogonal Multiple Access," in *2017 Wireless Days*, March 2017, pp. 188–193.

[9] Z. Ding, R. Schober, and H. V. Poor, "A General MIMO Framework for NOMA Downlink and Uplink Transmission Based on Signal Alignment," *IEEE Transactions on Wireless Communications*, vol. 15, no. 6, pp. 4438–4454, June 2016.

[10] M. Zeng, A. Yadav, O. A. Dobre, G. I. Tsiropoulos, and H. V. Poor, "On the Sum Rate of MIMO-NOMA and MIMO-OMA Systems," *IEEE Wireless Communications Letters*, vol. 6, no. 4, pp. 534–537, Aug 2017.

[11] Z. Ding, F. Adachi, and H. V. Poor, "The Application of MIMO to Non-Orthogonal Multiple Access," *IEEE Transactions on Wireless Communications*, vol. 15, no. 1, pp. 537–552, Jan 2016.

[12] M. Zeng, A. Yadav, O. A. Dobre, G. I. Tsiropoulos, and H. V. Poor, "Capacity Comparison Between MIMO-NOMA and MIMO-OMA With Multiple Users in a Cluster," *IEEE Journal on Selected Areas in Communications*, vol. 35, no. 10, pp. 2413–2424, Oct 2017.

[13] V. D. Nguyen, H. D. Tuan, T. Q. Duong, H. V. Poor, and O. S. Shin, "Precoder Design for Signal Superposition in MIMO-NOMA Multicell Networks," *IEEE Journal on Selected Areas in Communications*, vol. 35, no. 12, pp. 2681–2695, Dec 2017.

[14] Q. Sun, S. Han, C. L. I, and Z. Pan, "Energy Efficiency Optimization for Fading MIMO Non-Orthogonal Multiple Access Systems," in *2015 IEEE International Conference on Communications (ICC)*, June 2015, pp. 2668–2673.

[15] F. Fang, H. Zhang, J. Cheng, and V. C. M. Leung, "Energy-Efficient Resource Allocation for Downlink Non-Orthogonal Multiple Access Network," *IEEE Transactions on Communications*, vol. 64, no. 9, pp. 3722–3732, Sept 2016.

[16] Y. Zhang, H. M. Wang, T. X. Zheng, and Q. Yang, "Energy-Efficient Transmission Design in Non-Orthogonal Multiple Access," *IEEE Transactions on Vehicular Technology*, vol. 66, no. 3, pp. 2852–2857, March 2017.

[17] M. Zeng, A. Yadav, O. A. Dobre, and H. V. Poor, "Energy-Efficient Power Allocation for MIMO-NOMA With Multiple Users in a Cluster," *IEEE Access*, vol. 6, pp. 5170–5181, 2018.

[18] M. S. Ali, H. Tabassum, and E. Hossain, "Dynamic User Clustering and Power Allocation for Uplink and Downlink Non-Orthogonal Multiple Access (NOMA) Systems," *IEEE Access*, vol. 4, pp. 6325–6343, Aug 2016.