Multi-User Fairness in Reconfigurable Intelligent Surface Assisted mmWave MIMO-NOMA System

Nouran Arafat
nouran.zaghlool@guc.edu.eg

Engy Aly Maher
engy.aly@guc.edu.eg

Ahmed El-Mahdy
ahmed.elmahdy@guc.edu.eg

Falko Dressler
dressler@ccs-labs.org

***Abstract*—Reconfigurable Intelligent Surface (RIS) enabled wireless communications with Non Orthogonal Multiple Access (NOMA) is a promising technology for the next generation mobile communications. Millimeter wave (mmWave) communication has high frequency ranges from 30 to 300 GHz and supports gigabit per second data rates. The path loss attenuation is very high at high frequencies compared to low frequencies. In this paper, we consider a downlink mmWave MIMO-NOMA cellular system aided by RIS where the base station is mounted with multiple antennas and multiple single antenna users. RIS phase shifts and users power are optimized to maximize energy efficiency such that the rate of each user exceeds a certain threshold. The optimization problem is a non-convex problem, which can be solved using Dinkelbach's algorithm with fractional programming. The maximization problem is converted to Quadratic Constraint Quadratic programming (QCQP), and the Lagrange augmented method is applied to get the optimum RIS phases. Numerical results show that all users satisfy the rate constraint under optimal power allocation.**

***Index Terms*—Energy Efficiency, RIS, MIMO, NOMA, Multi-user, Rate Constraint, User Fairness, Power Consumption, Sum Rate, passive RIS, mmWave.**

I. Introduction

As the number of users increases, the demand for unprecedented data and ambiguous wireless connections force to propose new technologies and devices to satisfy such demand. This led to the introduction of communication over high-frequency bands as millimeter wave (mmWave). MmWave communication is a frequency band ranging from 30-300 GHz, thus it supports gigabit-per-second data rates. In expense of suffering from high path loss when compared to low-frequency bands [1], [2]. It experiences a millimeter wavelength that enables the deployment of large number of antennas in small area [3]. Consequently, multiple input multiple output (MIMO) is integrated with mmWave communications. In MIMO system, many antennas are employed at BS to serve multiple users.

Non orthogonal multiple access (NOMA) is an emerging technology, that achieves high spectral and energy efficiencies [4]. In power domain NOMA, users share the same frequency-time resource. The spectral efficiency of the wireless system is enhanced by using superposition coding (SC) at the transmitter and successive interference cancellation (SIC) at the receiver.

979-8-3503-5614-4/24/$31.00 © 2024 IEEE

Reconfigurable intelligent surfaces (RISs) have attracted much attention in current research because it have several advantages. As it can easily integrated into the system and mounted on the building. RISs are two-dimensional meta-surfaces that direct the incident wave in the desired direction. RISs consist of low-cost passive elements with phase shifts to control the electromagnetic signal and reflect it in the intended position [5]. The phases are adjusted so the reflected signal is added constructively at the receiver location and destructively in other locations to mitigate the interference. Moreover, RIS does not require a power source like other technologies as relays and MIMO beamforming [6]. Thus, it can improve the performance of the wireless system in terms of spectral and energy efficiency. To benefit from the above technologies, We integrate RIS with multi-user MIMO-NOMA in mmWave band to improve the performance of the system.

Our main contributions can be summarized as follows:

1) User fairness is guaranteed by optimization of the power allocation to users for the MIMO-NOMA mmWave system and it is measured by Jain's Index.
2) Energy efficiency maximization problem is evaluated under satisfying the minimum rate constraint. The problem is converted to convex using Dinkelbach's algorithm and fractional programming.
3) A comparison between the proposed algorithm and the algorithm in [7] is performed where user fairness is not considered.

II. Literature Review

Different techniques of NOMA are discussed in detail in [8] and [9]. Cooperative NOMA which considers integrating MIMO and relay with NOMA is evaluated in [8]. Results show the superior performance of cooperative NOMA compared to orthogonal multiple access (OMA). In [10], max-min fairness problem among users is considered in mmWave NOMA cellular system. Results compare the proposed solution with the conventional mmWave-OMA. The proposed algorithm reaches an upper bound of the achievable rate.

In [11], the rate maximization problem in RIS mmWave is presented. Two schemes are proposed to reduce the channel

estimation overhead. The throughput maximization problem of the RIS-aided MIMO is introduced in [12]. Furthermore, the comparison between RIS and relay is evaluated based on the operator's perspective. The energy efficiency maximization problem in downlink MISO system aided (RIS) is presented in [13].

Integrating RIS with NOMA are introduced in [14], [15] and [16]. A partitioning technique for RIS is proposed in [14], in which the base station is equipped with a single antenna and the RIS elements are divided into sub-surface, where every user is served by one sub surface. In [15], NOMA-aided Intelligent Reflecting Surface (IRS) is considered under two users only. Also, IRS-OMA is compared with IRS-NOMA. Energy efficiency shows superior performance when integrating NOMA. Multiple IRS multi-user system is developed in [17] to maximize the weighted sum rate. Complex spherical and complex oblique manifold is applied to obtain the optimum beamforming at BS and RIS phases respectively. Results emphasize that the proposed algorithm outperforms the conventional methods. In [18], optimization for both active/passive beamforming and power allocation in RIS-NOMA mmWave is considered. Both [14], [15] do not consider mmWave band.

Fairness in wireless networks is explicated in [19], where a comparison between quantitative and qualitative fairness models is given. Quantitative fairness, i.e Jain's index usually expressed with real values. On the other hand, qualitative fairness such as max-min and proportional fairness, can't be represented by real numbers. Jain's fairness algorithm is very simple when compared to other fairness models.

In our previous work [7], the performance of energy efficiency and power consumption for MIMO-NOMA multi-user mmWave system is presented. Large Intelligent Surface (LIS) phases and gain, in addition, users power are optimized to maximize energy efficiency. This paper is an extension to our paper [7] after considering user rate constraint. This constraint changed the optimization methodology in which the optimal phase shift is obtained by applying Lagrange augmented method and the optimal power allocation is solved using numerical solver, i.e., CVX.

III. System Model

In the following, the system and channel models are presented. A downlink mmWave MIMO-NOMA system aided by RIS is considered, where the base station (BS) has N antennas, and RIS consists of L elements to serve K users with single antenna, as shown in Fig. 1. There are two links in which the BS transmits the data to the users through, a direct link and a reflected link from RIS. The channel between BS to m^{th} user in the c^{th} cluster is denoted by $\boldsymbol{g}_{c,m} \in \mathbb{C}^{N\times 1}$, the channel between BS and RIS is given by $\boldsymbol{H} \in \mathbb{C}^{N\times L}$, and the channel between RIS and m^{th} user in the c^{th} cluster is represented by $\boldsymbol{f}_{c,m} \in \mathbb{C}^{L\times 1}$. The precoding matrix of RIS is denoted by $\boldsymbol{\Theta}$=diag($\boldsymbol{\theta}$) $\in \mathbb{C}^{L\times L}$. The phase shift vector of RIS elements is given by $\boldsymbol{\theta} = [\theta_1, \theta_2, ..., \theta_L]^T \in \mathbb{C}^{L\times 1}$, where $\theta_l = e^{j\phi_l}$.

The mmWave channel model is given as [3]

$$\mathbf{g}_{c,m} = \sqrt{\frac{N}{PA_{c,m}}} \sum_{p_a=1}^{PA_{c,m}} \alpha_{c,m}^{(p_a)} \mathbf{a}(\varphi_{c,m}^{(p_a)}, \vartheta_{c,m}^{(p_a)}), \tag{1}$$

where $\varphi_{c,m}^{(p_a)}$ denotes the azimuth angle of departure and $\vartheta_{c,m}^{(p_a)}$ represents the elevation angle of arrival. The number of paths for the m^{th} user in the c^{th} cluster group is given by $PA_{c,m}$. The complex gain of the p_a^{th} path is denoted as $\alpha_{c,m}^{(p_a)}$. However, $\mathbf{a}(\varphi_{c,m}^{(p_a)}, \vartheta_{c,m}^{(p_a)})$ represents the steering vector $\in \mathbb{C}^{N\times 1}$. Assuming the antenna elements are distributed in a uniform planar array (UPA) with N_1 antenna in the x-axis and N_2 antenna in the y-axis where $N = N_1N_2$, so $\mathbf{a}(\varphi, \vartheta) = \mathbf{a}_{az}(\varphi) \otimes \mathbf{a}_{el}(\vartheta)$, where $\mathbf{a}_{az}(\varphi)$ = $\frac{1}{\sqrt{N_1}}[1,\ e^{j2\pi(\frac{d_1}{\lambda})sin(\varphi)}, .., e^{j2\pi(N_1-1)(\frac{d_1}{\lambda})sin(\varphi)}]$, and $\mathbf{a}_{el}(\vartheta)$ = $\frac{1}{\sqrt{N_2}}[1,\ e^{j2\pi(\frac{d_2}{\lambda})sin(\vartheta)}, .., e^{j2\pi(N_2-1)(\frac{d_2}{\lambda})sin(\vartheta)}]$, $\otimes$ is the Kronecker product. The signal wavelength is denoted by λ, d_1 and d_2 are the spacing between elements in the x-axis and y-axis respectively, with $d_1 = d_2 = \lambda/2$.

Using the clustering algorithm in our previous work [20], users are divided into C clusters to mitigate the interference among them. The indices of NOMA users in c^{th} cluster are collected in set S_c. Thus, the set $|S_c|$ length represents the number of users in c^{th} cluster. At least one user must exist in every cluster. $r_{c,m}$ represents the received signal for the m^{th} user in the c^{th} cluster. It is given by

$$\begin{aligned} r_{c,m} &= \boldsymbol{h}_{c,m}^H \sum_{i=1}^{C} \sum_{j=1}^{|S_i|} \mathbf{w}_i \sqrt{p_{i,j}} x_{i,j} + n_{c,m} \\ &= \underbrace{\boldsymbol{h}_{c,m}^H \mathbf{w}_c \sqrt{p_{c,m}} x_{c,m}}_{\text{desired signal}} \\ &\quad \underbrace{+ \boldsymbol{h}_{c,m}^H \mathbf{w}_c \Big(\sum_{j=1}^{m-1} \sqrt{p_{c,j}} x_{c,j} + \sum_{j=m+1}^{|S_c|} \sqrt{p_{c,j}} x_{c,j} \Big)}_{\text{intra cluster interferences}} \\ &\quad \underbrace{+ \boldsymbol{h}_{c,m}^H \sum_{i\neq c}^{C} \mathbf{w}_i \sum_{j=1}^{|S_i|} \sqrt{p_{i,j}} x_{i,j}}_{\text{inter cluster interferences}} + \underbrace{n_{c,m}}_{\text{noise at receiver}}, \end{aligned} \tag{2}$$

where $x_{c,m}$ is the transmitted signal with $E\{|x_{c,m}|^2\} = 1$ and $p_{c,m}$ is the transmitted power for m^{th} user and c^{th} cluster. The equivalent channel from BS to m^{th} user in the c^{th} cluster is $\boldsymbol{h}_{c,m} = \boldsymbol{g}_{c,m} + \boldsymbol{H}\boldsymbol{\Theta}\boldsymbol{f}_{c,m}$, where ($\boldsymbol{g}_{c,m}$) is the direct link between BS and user, however the second part ($\boldsymbol{H}\boldsymbol{\Theta}\boldsymbol{f}_{c,m}$) is the reflected link from RIS.

The received signal $r_{c,m}$ at m^{th} user in c^{th} cluster in (2) is divided into four terms: desired signal, interference from users in the same cluster, interference due to users in other clusters, and noise at the receiver, respectively. The BS precoding vector for the i^{th} cluster is given by $\mathbf{w}_i$, where the number of beams from base station equals the cluster size C. Digital zero-forcing (ZF) precoder is applied at the transmitter and it is

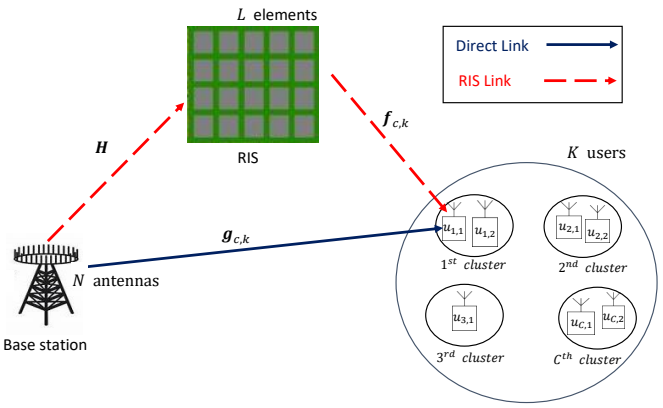

Fig. 1. RIS assisted downlink mmWave MIMO-NOMA system.

denoted by $\mathbf{w}_i \in \mathbb{C}^{N\times 1}$. The beamforming matrix $\hat{\mathbf{W}} \in \mathbb{C}^{N\times C}$ is given by

$$\hat{\mathbf{W}} = [\hat{\mathbf{w}}_1, \hat{\mathbf{w}}_2, ..., \hat{\mathbf{w}}_c] = \bar{\boldsymbol{H}}(\bar{\boldsymbol{H}}^H \bar{\boldsymbol{H}})^{-1} \tag{3}$$

where $\bar{\boldsymbol{H}} = [\boldsymbol{h}_{1,\underline{m}}, \boldsymbol{h}_{2,\underline{m}}, ..., \boldsymbol{h}_{C,\underline{m}}]$ and $\boldsymbol{h}_{c,\underline{m}}$ represents the $\underline{m}^{th}$ user with the highest channel gain in the c^{th} cluster. For each cluster, the normalized precoding vector is given by $\mathbf{w}_c = \frac{\hat{\mathbf{w}}_c}{||\hat{\mathbf{w}}_c||_2}$. The normalization ensures that the precoding is independent of the user power allocation, such that $||\mathbf{w}_c||^2 = 1$. The noise at each user is denoted by $n_{c,m}$ with complex Gaussian distribution $\mathcal{CN}(0, \sigma^2)$ with zero mean and variance σ^2. Assuming that users are arranged in descending order in each cluster based on channel gain. Thus, in (2), the intra-cluster interference term with summation greater than m is canceled due to successive interference cancellation (SIC). Then, the signal-to-interference and ratio (SINR) for m^{th} user in c^{th} cluster is given by

$$\begin{aligned}\gamma_{c,m} = (|\boldsymbol{h}_{c,m}^H \mathbf{w}_c|^2 \, p_{c,m}).(\, |\boldsymbol{h}_{c,m}^H \mathbf{w}_c|^2 \sum_{j=1}^{m-1} p_{c,j} \\ + \sum_{i\neq c}^{C} |\boldsymbol{h}_{c,m}^H \mathbf{w}_i|^2 \sum_{j=1}^{|S_i|} p_{i,j} + \sigma^2)^{-1}\end{aligned} \tag{4}$$

The rate of each user equals $R_{c,m} = log_2(1+\gamma_{c,m})$. The sum rate is given as

$$R_{\text{sum}} = \sum_{c=1}^{C} \sum_{m=1}^{|S_c|} R_{c,m}, \tag{5}$$

The definition of energy efficiency (EE) is the ratio of the sum rate to the total power consumption

$$EE = \frac{R_{\text{sum}}}{P_c^{\text{tot}}} \qquad (bps/Hz/W), \tag{6}$$

where P_c^{tot} is the total power dissipated due to the BS, RIS, and users

$$P_c^{\text{tot}} = \underbrace{\zeta P_t + P_b}_{\text{BS}} + \underbrace{LP_{PS}}_{\text{RIS}} + \underbrace{KP_u}_{\text{Users}}, \tag{7}$$

where ζ is the inverse energy conversion coefficients at the BS. In addition, the baseband power consumption is denoted by P_b. However, P_t is the transmitted power at BS which expressed as: $P_t = \sum_{c=1}^{C} \sum_{j=1}^{|S_c|} p_{c,j}$. The power dissipated per user is denoted by P_u. The RIS consumption power due to the phase shift hardware circuit is denoted by P_{PS}.

IV. Problem Formulation

In this section, the energy efficiency maximization is discussed. Our main concern is to optimize user powers ($\mathbf{p}$) and RIS phase shift ($\boldsymbol{\Theta}$) such that the maximum energy efficiency is achieved. Therefore, the optimization problem is given as follows

$$\max_{\mathbf{p},\boldsymbol{\Theta}} \quad \frac{R_{\text{sum}}}{P_c^{\text{tot}}} \tag{8a}$$

$$\text{s.t.} \quad C_1 : p_{c,m} \geq 0, \qquad \forall c \in C, m \in |S_c| \tag{8b}$$

$$C_2 : \zeta \sum_{c=1}^{C} \sum_{m=1}^{|S_c|} p_{c,m} \leq \tilde{P}_{\text{max}}^{\text{BS}}, \tag{8c}$$

$$C_3 : R_{c,m} \geq R_{\text{min}}, \qquad \forall c \in C, m \in |S_c| \tag{8d}$$

$$C_4 : |\theta_l| = 1, \qquad \forall l \in L \tag{8e}$$

where C_1 guarantees that the power assigned to each user must be positive, C_2 indicates that the total power consumption by the BS is less than the maximum power assigned to the BS ($P_{\text{max}}^{\text{BS}}$). However, C_3 is the user fairness constraint to guarantee the user data rate exceeds the minimum data rate required (R_{min}). The last constraint is for the unit modulus phase shift constraint.

Let $\tilde{P}_{\text{max}}^{\text{BS}} = P_{\text{max}}^{\text{BS}} - P_B$, and $\mathbf{p} = [p_{1,1}, p_{1,2}, p_{1,|S_1|}, ...p_{C,|S_C|}]^T$ is a vector of the power allocated for all users $\in \mathbb{R}_+^{K\times 1}$. The optimal solution is not affected by the base of the logarithmic function, thus natural logarithmic is considered throughout the rest of the paper.

The problem in (8) is non-convex because it is a non-linear function consisting of fractional and logarithmic functions. The non-convexity implies that the problem may has multiple local optima or discontinuities. To solve this non-convex problem, Dinkelbach's algorithm is applied to make the problem non-fractional so the Hessian matrix becomes non-negative. Dinkelbach's algorithm is an efficient method to solve fractional problems by decoupling the numerator and denominator of a single ratio [21]. Thus, Dinkelbach's algorithm transformed (8a) into parameterized problem as

$$f(\mathbf{p}, \boldsymbol{\Theta}) = R_{\text{sum}} - \eta P_c^{\text{tot}} \tag{9}$$

where $\eta = \frac{R_{\text{sum}}}{P_c^{\text{tot}}}$ is an auxiliary variable. The optimal energy efficiency η^{opt} satisfies this condition $f(\mathbf{p}^*, \boldsymbol{\Theta}^*) = 0$. The auxiliary variable η is updated iteratively until the condition is satisfied. Now, the maximization problem is given as

$$\begin{aligned}\max_{\mathbf{p},\boldsymbol{\Theta}} \quad & f(\mathbf{p}, \boldsymbol{\Theta}) \\ \text{s.t.} \quad & C_1, C_2, C_3, C_4\end{aligned} \tag{10}$$

In order to solve (10), which is still non-convex problem, We apply fractional programming to deal with the sum of logarithms of ratio [22]. After some mathematical manipulations, the original optimization problem (10) can be formulated as

$$\begin{aligned} \max_{\mathbf{p},\boldsymbol{\Theta},\boldsymbol{\mu},\boldsymbol{\nu}} \quad & f_1(\mathbf{p},\boldsymbol{\Theta},\boldsymbol{\mu},\boldsymbol{\nu}) \\ \text{s.t.} \quad & C_1, C_2, C_3, C_4 \end{aligned} \tag{11}$$

where $f_1(\mathbf{p},\boldsymbol{\Theta},\boldsymbol{\mu},\boldsymbol{\nu}) = \sum_{c=1}^{C}\sum_{m=1}^{|S_c|}[\ln(1+\mu_{c,m}) - \mu_{c,m}]$

$$+\sum_{c=1}^{C}\sum_{m=1}^{|S_c|} 2\sqrt{1+\mu_{c,m}}\mathfrak{Re}\{\nu_{c,m}^{*}A_{c,m}\}$$

$$-\sum_{c=1}^{C}\sum_{m=1}^{|S_c|}|\nu_{c,m}|^2 B_{c,m} - \eta P_c^{\text{tot}}$$

$\mu_{c,m}$ and $\nu_{c,m}$ are auxiliary variables, $A_{c,m}$ and $B_{c,m}$ are given as

$$A_{c,m} = \boldsymbol{h}_{c,m}^{H}\mathbf{w}_c\sqrt{p_{c,m}} \tag{12a}$$

$$B_{c,m} = \sum_{j=1}^{m}|\boldsymbol{h}_{c,m}^{H}\mathbf{w}_c\sqrt{p_{c,j}}|^2 + \sum_{i\neq c}^{C}\sum_{j=1}^{|S_i|}|\boldsymbol{h}_{c,m}^{H}\mathbf{w}_i\sqrt{p_{i,j}}|^2 + \sigma^2 \tag{12b}$$

The details of these mathematical manipulations can be found in our previous work [7]. The user fairness constraint C_3 in (8e) can be written as

$$C_3' : |A_{c,m}|^2 \geq E(B_{c,m} - |A_{c,m}|^2) \tag{13}$$

where $E = (2^{R_{\min}} - 1)$, $A_{c,m}$ and $B_{c,m}$ are given in (12).

Alternating joint optimization can be used to find the optimum solution [23]. The idea behind the joint optimization is to optimize each variable solely considering the remaining variables are fixed.

V. RIS Phases and Users Power Joint Optimization

We can now solve problem (11) using alternating optimization. Since our target is to attain the optimal RIS phases ($\boldsymbol{\Theta}^{opt}$), auxiliary variables ($\boldsymbol{\mu}^{opt}$ & $\boldsymbol{\nu}^{opt}$) and user power allocation ($\mathbf{p}^{opt}$). Thus, our main problem can be divided into sub-problems to decouple the variables.

A. RIS Phases Optimization

Since RIS phases matrix is defined as $\boldsymbol{\Theta} = \text{diag}(\boldsymbol{\theta})$, where $\boldsymbol{\theta} = [\theta_1, \theta_2, .., \theta_L]^T \in \mathbb{C}^{L\times 1}$. The optimal values for RIS phases is obtained by formulating problem (11) as follows

$$\begin{aligned} \max_{\boldsymbol{\theta}} \quad & f_2(\boldsymbol{\theta}) = \mathfrak{Re}\{\boldsymbol{\theta}^H\boldsymbol{b}\} - \boldsymbol{\theta}^H\boldsymbol{S}\boldsymbol{\theta} \\ \text{s.t.} \quad & C_3 : \boldsymbol{\theta}^H\boldsymbol{Q_{c,m}}\boldsymbol{\theta} + 2\mathfrak{Re}\{\boldsymbol{\theta}^H\boldsymbol{s_{c,m}}\} \geq E\sigma^2, \ \forall c, m \\ & C_4 : \boldsymbol{\theta}^H\boldsymbol{\theta} = L \end{aligned} \tag{14}$$

where

$$\boldsymbol{b} = \sum_{c=1}^{C}\sum_{m=1}^{|S_c|} 2\sqrt{p_{c,m}(1+\mu_{c,m})}\nu_{c,m}^{*}\text{diag}(\boldsymbol{f}_{c,m}^{H})\boldsymbol{\beta}_c - \sum_{c=1}^{C}\sum_{m=1}^{|S_c|}|\nu_{c,m}|^2\Bigg(\sum_{j=1}^{m}2\alpha_{c,m,c}^{*}\text{diag}(\boldsymbol{f}_{c,m}^{H})\boldsymbol{\beta}_c p_{c,j} + \sum_{i\neq c}^{C}\sum_{j=1}^{|S_i|}2\alpha_{c,m,i}^{*}\text{diag}(\boldsymbol{f}_{c,m}^{H})\boldsymbol{\beta}_i p_{i,j}\Bigg)$$

$$\boldsymbol{S} = \sum_{c=1}^{C}\sum_{m=1}^{|S_c|}|\nu_{c,m}|^2\Bigg(\sum_{j=1}^{m}p_{c,j}\text{diag}(\boldsymbol{\beta}_c^{*})\boldsymbol{f}_{c,m}\boldsymbol{f}_{c,m}^{H}\text{diag}(\boldsymbol{\beta}_c) + \sum_{i\neq c}^{C}\sum_{j=1}^{|S_i|}p_{i,j}\text{diag}(\boldsymbol{\beta}_i^{*})\boldsymbol{f}_{c,m}\boldsymbol{f}_{c,m}^{H}\text{diag}(\boldsymbol{\beta}_i)\Bigg)$$

$$\boldsymbol{Q_{c,m}} = (\boldsymbol{f}_{c,m}\odot\boldsymbol{\beta}_c^{*})(\boldsymbol{f}_{c,m}^{H}\odot\boldsymbol{\beta}_c^{T})\Bigg(p_{c,m} - E\sum_{j=1}^{m-1}p_{c,j}\Bigg) - E\sum_{i\neq c}^{C}\sum_{j=1}^{|S_i|}p_{i,j}(\boldsymbol{f}_{c,m}\odot\boldsymbol{\beta}_i^{*})(\boldsymbol{f}_{c,m}^{H}\odot\boldsymbol{\beta}_i^{T})$$

$$\boldsymbol{s_{c,m}} = \alpha_{c,m,c}\text{diag}(\boldsymbol{\beta}_c^{*})\boldsymbol{f}_{c,m}p_{c,m} - E\Bigg(\sum_{j=1}^{m-1}\alpha_{c,m,c}\text{diag}(\boldsymbol{\beta}_c^{*})\boldsymbol{f}_{c,m}p_{c,j} + \sum_{i\neq c}^{C}\sum_{j=1}^{|S_i|}\alpha_{c,m,i}\text{diag}(\boldsymbol{\beta}_i^{*})\boldsymbol{f}_{c,m}p_{i,j}\Bigg)$$

Since, the equivalent channel is given as $\boldsymbol{h}_{c,m} = \boldsymbol{g}_{c,m} + \boldsymbol{H}\boldsymbol{\Theta}\boldsymbol{f}_{c,m}$. Then, let $\boldsymbol{h}_{c,m}^{H}\mathbf{w}_i = \alpha_{c,m,i} + \boldsymbol{f}_{c,m}^{H}\text{diag}(\boldsymbol{\beta}_i)\boldsymbol{\theta}$, where $\alpha_{c,m,i} = \boldsymbol{g}_{c,m}^{H}\mathbf{w}_i$, $\alpha_{c,m,i} \in \mathbb{C}^{1\times 1}$ and $\boldsymbol{\beta}_i = \boldsymbol{H}^{H}\mathbf{w}_i$, $\boldsymbol{\beta}_i \in \mathbb{C}^{L\times 1}$.

The Augmented Lagrangian method can be applied to solve problem (14) as given in (16, top of next page), where $\boldsymbol{\lambda}, \boldsymbol{\rho}, \lambda_1$ and ρ_1 are Lagrange multipliers and penalty parameters for C_3 and C_4. Starting with initial values for $\boldsymbol{\theta}, \boldsymbol{\lambda}, \boldsymbol{\rho}, \lambda_1$ and ρ_1, then they are updated according to the sub-gradient method as follows

$$\lambda_{c,m}^{(t)} = \lambda_{c,m}^{(t-1)} - \delta_1\Bigg(\frac{\partial\mathfrak{L}(\boldsymbol{\theta},\boldsymbol{\lambda},\boldsymbol{\rho},\lambda_1,\rho_1)}{\partial\lambda_{c,m}}\Bigg) \tag{15a}$$

$$\lambda_1^{(t)} = \lambda_1^{(t-1)} - \delta_2\Bigg(\frac{\partial\mathfrak{L}(\boldsymbol{\theta},\boldsymbol{\lambda},\boldsymbol{\rho},\lambda_1,\rho_1)}{\partial\lambda_1}\Bigg) \tag{15b}$$

$$\boldsymbol{\theta}^{(t)} = \boldsymbol{\theta}^{(t-1)} - \Bigg(\frac{\partial\mathfrak{L}(\boldsymbol{\theta},\boldsymbol{\lambda},\boldsymbol{\rho},\lambda_1,\rho_1)}{\partial\boldsymbol{\theta}}\Bigg) \tag{15c}$$

$$\frac{\partial\mathfrak{L}(\boldsymbol{\theta},\boldsymbol{\lambda},\boldsymbol{\rho},\lambda_1,\rho_1)}{\partial\lambda_1} = (\boldsymbol{\theta}^H\boldsymbol{\theta} - L)$$

$$\frac{\partial\mathfrak{L}(\boldsymbol{\theta},\boldsymbol{\lambda},\boldsymbol{\rho},\lambda_1,\rho_1)}{\partial\lambda_{c,m}} = (\boldsymbol{\theta}^H\boldsymbol{Q_{c,m}}\boldsymbol{\theta} + 2\mathfrak{Re}\{\boldsymbol{\theta}^H\boldsymbol{s_{c,m}}\} - E\sigma^2)$$

where δ_1, δ_2 are the step size for constraint C_3 and C_4 respectively. t is the iteration number.

$$\mathfrak{L}(\boldsymbol{\theta},\boldsymbol{\lambda},\boldsymbol{\rho},\lambda_1,\rho_1) = f_2(\boldsymbol{\theta}) - \lambda_1(\boldsymbol{\theta}^H\boldsymbol{\theta} - L) + \frac{\rho_1}{2}(\boldsymbol{\theta}^H\boldsymbol{\theta} - L)^2 + \sum_{c=1}^{C}\sum_{m=1}^{|S_c|}\lambda_{c,m}(\boldsymbol{\theta}^H\boldsymbol{Q_{c,m}}\boldsymbol{\theta} + 2\mathfrak{Re}\{\boldsymbol{\theta}^H\boldsymbol{s_{c,m}}\} - E\sigma^2) + \sum_{c=1}^{C}\sum_{m=1}^{|S_c|}\frac{\rho_{c,m}}{2}(\boldsymbol{\theta}^H\boldsymbol{Q_{c,m}}\boldsymbol{\theta} + 2\mathfrak{Re}\{\boldsymbol{\theta}^H\boldsymbol{s_{c,m}}\} - E\sigma^2)^2 \tag{16}$$

$$f_3(\mathbf{p}) = \sum_{c=1}^{C}\sum_{m=1}^{|S_c|}\hat{A}_{c,m}\mathfrak{Re}\{\hat{B}_{c,m}\sqrt{p_{c,m}}\} - \sum_{c=1}^{C}\sum_{m=1}^{|S_c|}|\nu_{c,m}|^2\Bigg[\sum_{j=1}^{m}|D_{c,m,c}\sqrt{p_{c,j}}|^2 + \sum_{i\neq c}^{C}\sum_{j=1}^{|S_i|}|D_{c,m,i}\sqrt{p_{i,j}}|^2 + \sigma^2\Bigg] - \eta\zeta\Bigg(\sum_{i=1}^{C}\sum_{j=1}^{|S_i|}p_{i,j}\Bigg) - Q + X \tag{17}$$

B. Auxiliary Variables Optimization

The optimum auxiliary variables $(\boldsymbol{\mu}^{opt}, \boldsymbol{\nu}^{opt})$ can be obtained under fixing user powers and RIS phase shifters. By differentiating $f_1(\mathbf{p}, \boldsymbol{\Theta}, \boldsymbol{\mu}, \boldsymbol{\nu})$ given in (11) w.r.t $\mu_{c,m}$ and $\nu_{c,m}$. Thus, the optimal auxiliary variables are

$$\mu_{c,m}^{opt} = \frac{\mathfrak{D}_{c,m}}{2}(\mathfrak{D}_{c,m} + \sqrt{\mathfrak{D}_{c,m}^2 + 4}) \tag{18a}$$

$$\nu_{c,m}^{opt} = \frac{\sqrt{1+\mu_{c,m}}A_{c,m}}{B_{c,m}} \tag{18b}$$

where $\mathfrak{D}_{c,m} = \mathfrak{Re}\{\nu_{c,m}^* A_{c,m}\}$, $A_{c,m}$ and $B_{c,m}$ are given in (12).

C. Power Allocation Optimization

The optimal user powers are derived in this section. Thus, the following variables ($\boldsymbol{\mu}$, $\boldsymbol{\nu}$,$\boldsymbol{\Theta}$) are fixed. Hence, problem (11) is rewritten as

$$\max_{\mathbf{p}} \quad f_3(\mathbf{p}) \tag{19a}$$

$$\text{s.t.} \quad C_1 : p_{c,m} \geq 0, \quad \forall c \in C, m \in |S_c| \tag{19b}$$

$$C_2 : \sum_{c=1}^{C}\sum_{m=1}^{|S_c|}p_{c,m} \leq \bar{P}_{\max}^{\text{BS}}, \tag{19c}$$

$$C_3' : |A_{c,m}|^2 \geq E(B_{c,m} - |A_{c,m}|^2), \quad \forall c, m \tag{19d}$$

where $f_3(\mathbf{p})$ is given in (17, top of this page), and

$$\hat{A}_{c,m} = 2\sqrt{1+\mu_{c,m}}, \quad D_{c,m,i} = \boldsymbol{h}_{c,m}^H\mathbf{w}_i,$$
$$\hat{B}_{c,m} = \nu_{c,m}^* D_{c,m,c}, \qquad \bar{P}_{\max}^{\text{BS}} = \zeta^{-1}\tilde{P}_{\max}^{\text{BS}},$$
$$Q = \eta\Bigg[P_b + MP_u + LP_{PS}\Bigg],$$
$$X = \sum_{c=1}^{C}\sum_{m=1}^{|S_c|}\ln(1+\mu_{c,m}) - \mu_{c,m}.$$

This problem is a standard convex problem that can be solved using numerical solvers such as CVX [24].

D. User Fairness

The fairness is measured using Jain's fairness index [25] as

$$J = \frac{\Big(\sum_{c=1}^{C}\sum_{m=1}^{|S_c|}R_{c,m}\Big)^2}{K\Big(\sum_{c=1}^{C}\sum_{m=1}^{|S_c|}R_{c,m}^2\Big)} \tag{20}$$

where $\frac{1}{M} \leq J \leq 1$, the index has the following properties:

- If all users get the same data rate, i.e., the rate for each user equals R_{sum}/K. Then Jain's index reaches its maximum value $J = 1$, which indicates more fairness.
- If one of the users gets almost all the sum rate and the other users get almost zero rates (when all users have the same power allocation), then the index reaches its minimum value $J = \frac{1}{K}$.

VI. Performance Evaluation

We consider the mmWave channel model with three paths ($PA_{c,m} = 3$), one is line of sight $\alpha_{c,m}^{(1)}$ and the others are non-line-of-sight $\alpha_{c,m}^{(p_a)}$; $p_a \in \{2,3\}$. The paths follow complex Gaussian distribution with zero mean and variance for $\alpha_{c,m}^{(1)}$ equals 10^{-2}, and for the other two paths $\alpha_{c,m}^{(p_a)}$ equal to 10^{-3}. Regarding the azimuth angle of departure ($\varphi_{c,m}^{(p_a)}$) and the elevation angle of arrival ($\vartheta_{c,m}^{(p_a)}$), they follow uniform distribution $[-\pi, \pi]$. The base station deploys N antennas in a uniform linear array (ULA), however passive RIS employs a uniform planar array (UPA) with $L = L_x L_y$. Where L_x is the number of elements along the x-axis and L_y is the number of elements along the y-axis. The direct link between the base station and users is assumed to be blocked. Table I summarises the simulation parameters. For comparison purposes, same parameters are used as in [7]. In the simulations, without optimization means random RIS phase and equal distributed power among users are considered.

Fig. 2 depicts the fairness among users against the maximum BS power ($P_{\max}^{\text{BS}}$). From the figure, the algorithm in [7] has lower fairness when compared to the proposed algorithm,

TABLE I
SIMULATION VARAIBLES

Description	Values
BS position, (x_1, y_1, z_1)	(0m,-20m,0m)
RIS position, (x_2, y_2, z_2)	(100m,5m,0m)
BS antennas, (N)	8
Size of RIS, (L)	256 elements
Cluster size, (C)	4
K	6 users
ζ	1.1
User power dissipated, (P_u)	10mW
Phase shift power consumption, (P_{PS})	10mW
Noise variance at user, (σ^2)	10^{-8}mW
BS baseband power, (P_b)	6dB
BS maximum power, $(P_{\max}^{\text{BS}})$	[9dB, 18dB]

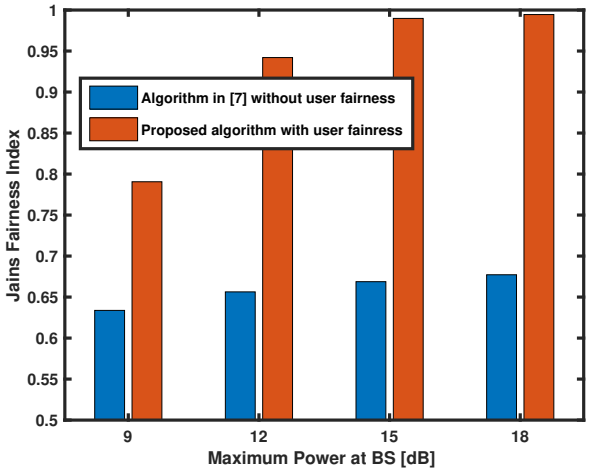

Fig. 2. **Jain's fairness index vs. power at BS** $(P_{\max}^{\text{BS}})$ **at** $R_{\min} = 0.8$**.**

which verifies that our proposed algorithm achieves high fairness among users.

Fig. 3 shows the energy efficiency versus the number of antennas at BS. As the number of antennas increases, the EE increases. The proposed optimization algorithm performs better than the conventional case with no optimization.

Fig. 4 shows energy efficiency versus the maximum power at BS $(P_{\max}^{\text{BS}})$. As the power at the BS increases, the EE decreases. The reason behind that, as the power at the base station increases, the total consumption power increases. Moreover, it can be observed that the EE obtained after optimization has better performance than without optimization. Because, the system's total power consumption decreases by optimization, which reflects on the EE and shows better performance. Also, the figure shows that the algorithm in [7] has better performance than the proposed algorithm with user fairness. The reason for this is that the algorithm in [7] sums all the rates of the users regardless of their values, but the algorithm with user fairness considers only the rates that exceed the threshold. In other words, in [7] there may be some users with rates $R_{c,m}$ higher than $R_{\min}$ and other users with lower rates than $R_{\min}$. Thus, the sum rate in [7] is high but not all users satisfy the

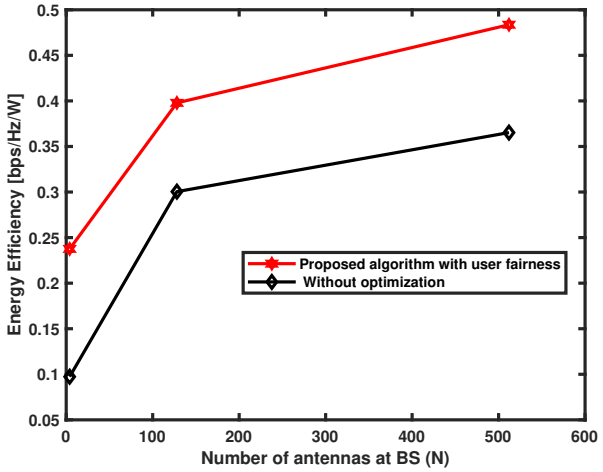

Fig. 3. **EE vs. BS antennas** (N) **with and without optimization at** $P_{\max}^{\text{BS}} = 15$**dB.**

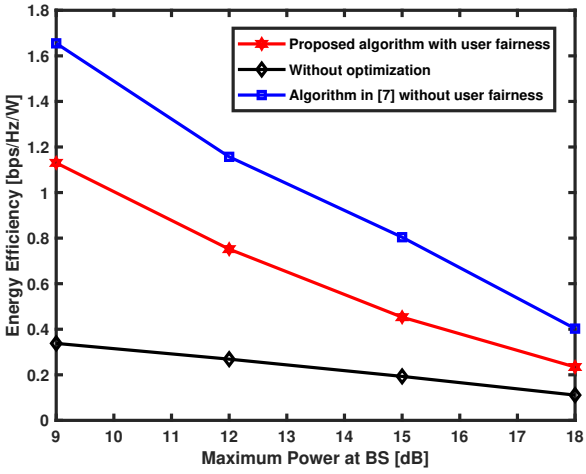

Fig. 4. **EE vs. power at BS** $(P_{\max}^{\text{BS}})$**.**

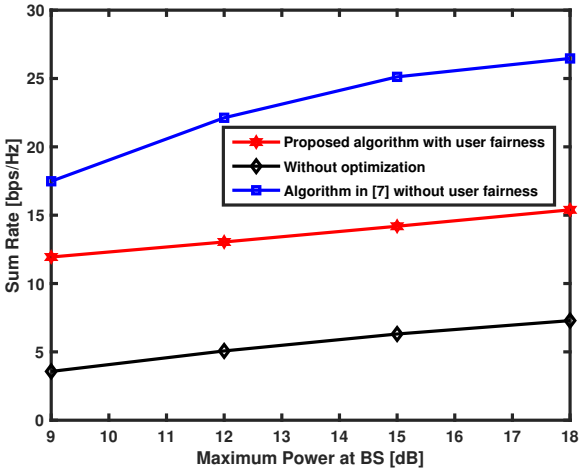

Fig. 5. **Sum rate vs. power at BS** $(P_{\max}^{\text{BS}})$**.**

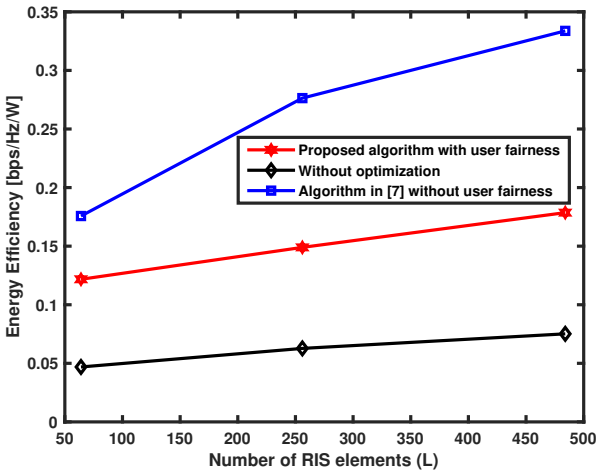

Fig. 6. **EE vs. size of RIS (L) elements.**

minimum required threshold. On the other hand, the algorithm proposed in this paper guarantees that all users satisfy the rate constraint.

The sum rate versus $P_{\max}^{\text{BS}}$ is given in Fig. 5. The sum rate shows clearly superior performance in case of the algorithm in [7]. As previously mentioned, this is because of the absence of user fairness constraints. Thus, the proposed algorithm has better user fairness at the expense of low sum rate and energy efficiency performance compared to algorithm [7]. This is the price of fairness. In addition, the figure shows that the sum rate increases as the BS power increases.

Fig. 6 illustrates the energy efficiency versus the number of RIS elements (L). When the number of elements at RIS increases, the energy efficiency increases. It's clear from the figure, that the performance when RIS phases and users power optimized outperforms the conventional case without any optimization.

VII. Conclusions

In this paper, a downlink mmWave MIMO-NOMA cellular system aided by RIS is considered. The users power and the RIS phases are optimized to maximize the energy efficiency considering user fairness for the rate. Dinkelbach's algorithm is applied to convert the maximization problem to convex. The results show that all the rates of the users satisfied the constraint. It is also shown that the performance of the proposed algorithm outperforms the non-optimized one in terms of the EE and sum rate. For future work, active RIS will be considered under user fairness constraint when optimizing NOMA users power in addition to phase and gain of RIS elements. Moreover, hybrid RIS can be evaluated where some elements are active and the rest are passive.

References

[1] I. A. Hemadeh, K. Satyanarayana, M. El-Hajjar, and L. Hanzo, "Millimeter-Wave Communications: Physical Channel Models, Design Considerations, Antenna Constructions, and Link-Budget," *IEEE Communications Surveys & Tutorials*, vol. 20, no. 2, pp. 870–913, 2017.

[2] M. Xiao, S. Mumtaz, Y. Huang, L. Dai, Y. Li, M. Matthaiou, G. K. Karagiannidis, E. Björnson, K. Yang, C.-L. I, and A. Ghosh, "Millimeter Wave Communications for Future Mobile Networks," *IEEE Journal on Selected Areas in Communications*, vol. 35, pp. 1909–1935, Sept. 2017.

[3] O. El Ayach, S. Rajagopal, S. Abu-Surra, Z. Pi, and R. W. Heath, "Spatially Sparse Precoding in Millimeter Wave MIMO Systems," *IEEE Transactions on Wireless Communications*, vol. 13, no. 3, pp. 1499–1513, 2014.

[4] 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 & Tutorials*, vol. 20, no. 3, pp. 2294–2323, 2018.

[5] R. Long, Y.-C. Liang, Y. Pei, and E. G. Larsson, "Active Reconfigurable Intelligent Surface-Aided Wireless Communications," *IEEE Transactions on Wireless Communications*, vol. 20, pp. 4962–4975, Aug. 2021.

[6] T. Sharma, A. Chehri, and P. Fortier, "Reconfigurable Intelligent Surfaces for 5G and beyond Wireless Communications: A Comprehensive Survey," *Energies*, vol. 14, p. 8219, Dec. 2021.

[7] N. Arafat, E. A. Maher, A. El-Mahdy, and F. Dressler, "Performance of Large Intelligent Surfaces in Multiuser Millimeter Wave MIMO-NOMA Systems," *IEEE Access*, vol. 11, pp. 142235–142247, 2023.

[8] S. R. Islam, N. Avazov, O. A. Dobre, and K.-S. Kwak, "Power-domain non-orthogonal multiple access (NOMA) in 5G systems: Potentials and challenges," *IEEE Communications Surveys & Tutorials*, vol. 19, no. 2, pp. 721–742, 2016.

[9] M. Vaezi, G. A. A. Baduge, Y. Liu, A. Arafa, F. Fang, and Z. Ding, "Interplay between NOMA and other emerging technologies: A survey," *IEEE Transactions on Cognitive Communications and Networking*, vol. 5, no. 4, pp. 900–919, 2019.

[10] Z. Xiao, L. Zhu, Z. Gao, D. O. Wu, and X.-G. Xia, "User fairness non-orthogonal multiple access (NOMA) for millimeter-wave communications with analog beamforming," *IEEE Transactions on Wireless Communications*, vol. 18, no. 7, pp. 3411–3423, 2019.

[11] D. Dampahalage, K. S. Manosha, N. Rajatheva, and M. Latva-aho, "Intelligent reflecting surface aided vehicular communications," in *2020 IEEE Globecom Workshops (GC Wkshps*, pp. 1–6, IEEE, 2020.

[12] Q. Gu, D. Wu, X. Su, J. Jin, Y. Yuan, and J. Wang, "Performance comparison between reconfigurable intelligent surface and relays: Theoretical methods and a perspective from operator," *arXiv preprint arXiv:2101.12091*, 2021.

[13] C. Huang, A. Zappone, G. C. Alexandropoulos, M. Debbah, and C. Yuen, "Reconfigurable intelligent surfaces for energy efficiency in wireless communication," *IEEE transactions on wireless communications*, vol. 18, no. 8, pp. 4157–4170, 2019.

[14] A. Khaleel and E. Basar, "A Novel NOMA Solution With RIS Partitioning," *IEEE Journal of Selected Topics in Signal Processing*, vol. 16, pp. 70–81, Jan. 2022.

[15] F. Fang, Y. Xu, Q.-V. Pham, and Z. Ding, "Energy-efficient design of IRS-NOMA networks," *IEEE Transactions on Vehicular Technology*, vol. 69, no. 11, pp. 14088–14092, 2020.

[16] F. Fang, B. Wu, S. Fu, Z. Ding, and X. Wang, "Energy-efficient design of STAR-RIS aided MIMO-NOMA networks," *IEEE Transactions on Communications*, vol. 71, no. 1, pp. 498–511, 2022.

[17] L. Zhang, Q. Wang, and H. Wang, "Multiple Intelligent Reflecting Surface aided Multi-user Weighted Sum-Rate Maximization using Manifold Optimization," in *2021 IEEE/CIC International Conference on Communications in China (ICCC)*, pp. 364–369, IEEE, 2021.

[18] J. Zuo, Y. Liu, E. Basar, and O. A. Dobre, "Intelligent Reflecting Surface Enhanced Millimeter-Wave NOMA Systems," *IEEE Communications Letters*, vol. 24, no. 11, pp. 2632–2636, 2020.

[19] H. SHI, R. V. Prasad, E. Onur, and I. Niemegeers, "Fairness in Wireless Networks: Issues, Measures and Challenges," *IEEE Communications Surveys & Tutorials*, vol. 16, no. 1, pp. 5–24, 2014.

[20] N. Arafat and A. El-Mahdy, "Performance of fully digital zero-forcing precoding in mmWave massive MIMO-NOMA with antenna reduction," in *2021 30th Wireless and Optical Communications Conference (WOCC)*, pp. 166–170, IEEE, 2021.

[21] K. Shen and W. Yu, "Fractional programming for communication systems—Part I: Power control and beamforming," *IEEE Transactions on Signal Processing*, vol. 66, no. 10, pp. 2616–2630, 2018.

[22] K. Shen and W. Yu, "Fractional programming for communication systems—Part II: Uplink scheduling via matching," *IEEE Transactions on Signal Processing*, vol. 66, no. 10, pp. 2631–2644, 2018.

[23] Y. Xu and W. Yin, "A block coordinate descent method for regularized multiconvex optimization with applications to nonnegative tensor fac-

torization and completion," *SIAM Journal on imaging sciences*, vol. 6, no. 3, pp. 1758–1789, 2013.

[24] M. Grant and S. Boyd, "CVX: Matlab software for disciplined convex programming, version 2.1," 2014.

[25] Z. Wei, J. Guo, D. W. K. Ng, and J. Yuan, "Fairness comparison of uplink noma and oma," in *2017 IEEE 85th Vehicular Technology Conference (VTC Spring)*, pp. 1–6, 2017.